Contour Approximation of Spatial Data, with Applications

Adi Ben-Israel

September 19, 2019

Outline

Abstract

Contour Approximation

Given a set S of points in $\mathbb{R}^n$, a **contour approximation** of S is a function that captures most points of S in its lower level sets.

A concrete application is the **home-range** of an animal population, or the territory occupied by it, shown in 1980 by Dixon and Chapman to involve the harmonic mean of distances, a result since then confirmed for many species. The harmonic mean of distances, or resistances, also features in **inverse distance weighted interpolation**, **clustering**, **parallel circuits** and **multi–facility location**. This lecture gives an axiomatic framework, and a probabilistic optimization model that unifies the above results, a model applied successfully to clustering and classification.

Joint work with Tsvetan Asamov and Cem Iyigun.

Abstract

Contour Approximation

Given a set S of points in $\mathbb{R}^n$, a **contour approximation** of S is a function that captures most points of S in its lower level sets.

A concrete application is the **home-range** of an animal population, or the territory occupied by it, shown in 1980 by Dixon and Chapman to involve the harmonic mean of distances, a result since then confirmed for many species. The harmonic mean of distances, or resistances, also features in **inverse distance weighted interpolation**, **clustering**, **parallel circuits** and **multi–facility location**. This lecture gives an axiomatic framework, and a probabilistic optimization model that unifies the above results, a model applied successfully to clustering and classification.

Joint work with Tsvetan Asamov and Cem Iyigun.

Abstract

Contour Approximation

Given a set S of points in $\mathbb{R}^n$, a **contour approximation** of S is a function that captures most points of S in its lower level sets.

A concrete application is the **home-range** of an animal population, or the territory occupied by it, shown in 1980 by Dixon and Chapman to involve the harmonic mean of distances, a result since then confirmed for many species. The harmonic mean of distances, or resistances, also features in **inverse distance weighted interpolation**, **clustering**, **parallel circuits** and **multi–facility location**. This lecture gives an axiomatic framework, and a probabilistic optimization model that unifies the above results, a model applied successfully to clustering and classification.

Joint work with Tsvetan Asamov and Cem Iyigun.

Abstract

Contour Approximation

Given a set S of points in $\mathbb{R}^n$, a **contour approximation** of S is a function that captures most points of S in its lower level sets.

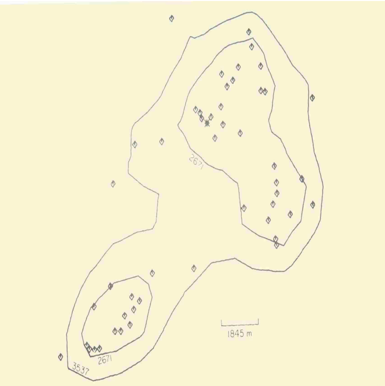
A concrete application is the **home-range** of an animal population, or the territory occupied by it, shown in 1980 by Dixon and Chapman to involve the harmonic mean of distances, a result since then confirmed for many species. The harmonic mean of distances, or resistances, also features in **inverse distance weighted interpolation**, **clustering**, **parallel circuits** and **multi–facility location**. This lecture gives an axiomatic framework, and a probabilistic optimization model that unifies the above results, a model applied successfully to clustering and classification.

Joint work with Tsvetan Asamov and Cem Iyigun.

Outline

Example: Home range of black bear (ID)

A **home range** is the area in which an animal lives and moves on a periodic basis.



M.D. Samuel, D.J. Pierce and E.O. Garton, Identifying areas of concentrated use within the home range, *J. Animal Ecology* **54**(1985), 711–719

The Home Range, Dixon & Chapman, [7]

A new method of calculating centers and areas of animal activity is presented based on the **harmonic mean** of an areal distribution. The **center of activity** is located in the area of greatest activity; in fact more than one "center" may exist.

...

The calculation of **home range** allows for heterogeneity of any habitat and is illustrated with data collected near Corvallis, Oregon, on the brush rabbit (*Sylvilagus bachmani*)

K.R. Dixon and J.A. Chapman, Harmonic mean measure of animal activity areas, *Ecology* **61**(1980), 1040–1044

More black bears, [6, p. 76]

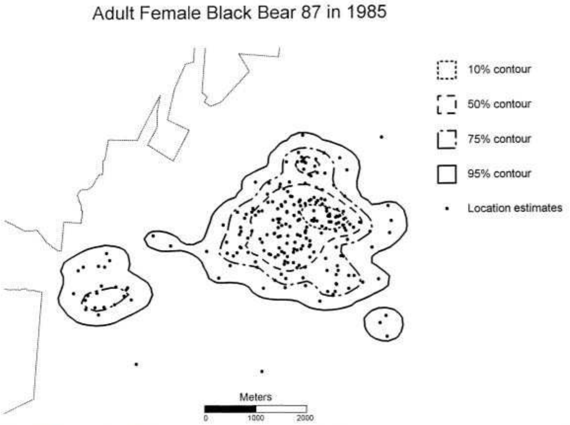


Figure 3.2 Location estimates (circles) and contours for the probability density function for adult female black bear 87 studied in 1985. The lightly dotted black line marks the study area border.

The bears are confused, [6, p. 88]

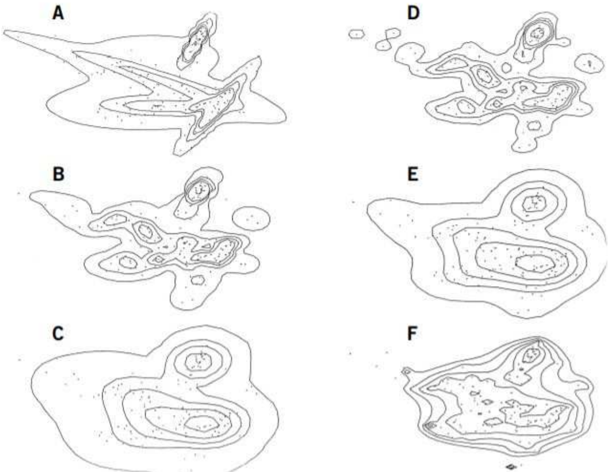
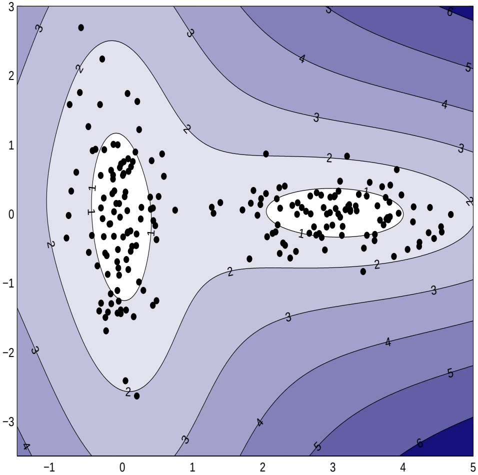


Figure 3.4 A complex, simulated home range. (A) True density contours. (B) Fixed kernel density estimate with cross-validated band width choice. (C) Adaptive kernel density estimate with cross-validated band width choice. (D) Fixed kernel density estimate with ad hoc band width choice. (E) Adaptive kernel density estimate with ad hoc band width choice. (F) Harmonic mean estimate. Modified from Powell et al. (1997).

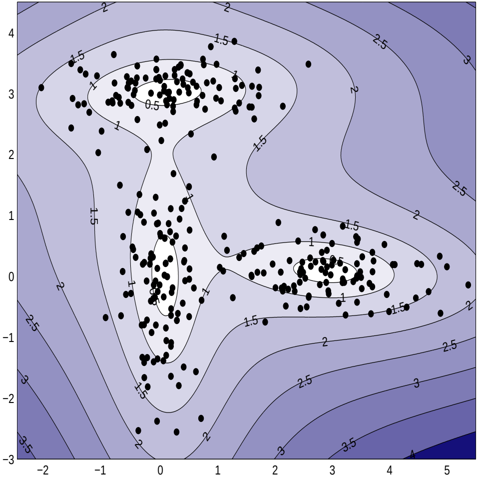
Example: 2 centers



Two clusters with different “geometries”.

10 numbers suffice to represent the data.

Example: 3 centers



15 numbers suffice to represent the data.

Outline

The harmonic mean

The **harmonic mean** of n positive numbers $x_1, \cdots, x_n$ is

$$H(x_1, \cdots, x_n) = \frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \quad (1)$$

The AGH inequality,

$$A(x_1, \cdots, x_n) \geq G(x_1, \cdots, x_n) \geq H(x_1, \cdots, x_n)$$

with equality iff $x_1 = \cdots = x_n$

If $\{x_1, \cdots, x_n\}$ have weights $\{w_1, \cdots, w_n\}$, their **weighted harmonic mean** is

$$H(x_1, \cdots, x_n; w_1, \cdots, w_n) = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}} \quad (2)$$

(1) is a special case of (2) for all weights $w_i = 1$.

The harmonic mean

The **harmonic mean** of n positive numbers $x_1, \cdots, x_n$ is

$$H(x_1, \cdots, x_n) = \frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \tag{1}$$

The AGH inequality,

$$A(x_1, \cdots, x_n) \geq G(x_1, \cdots, x_n) \geq H(x_1, \cdots, x_n)$$

with equality iff $x_1 = \cdots = x_n$

If $\{x_1, \cdots, x_n\}$ have weights $\{w_1, \cdots, w_n\}$, their **weighted harmonic mean** is

$$H(x_1, \cdots, x_n; w_1, \cdots, w_n) = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}} \tag{2}$$

(1) is a special case of (2) for all weights $w_i = 1$.

The harmonic mean

The **harmonic mean** of n positive numbers $x_1, \cdots, x_n$ is

$$H(x_1, \cdots, x_n) = \frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}} \tag{1}$$

The AGH inequality,

$$A(x_1, \cdots, x_n) \geq G(x_1, \cdots, x_n) \geq H(x_1, \cdots, x_n)$$

with equality iff $x_1 = \cdots = x_n$

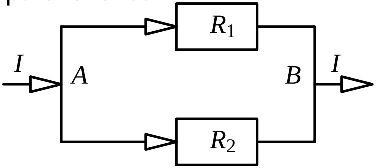
If $\{x_1, \cdots, x_n\}$ have weights $\{w_1, \cdots, w_n\}$, their **weighted harmonic mean** is

$$H(x_1, \cdots, x_n; w_1, \cdots, w_n) = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{x_i}} \tag{2}$$

(1) is a special case of (2) for all weights $w_i = 1$.

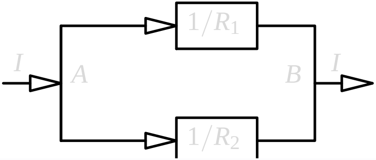
Example: Parallel circuit

Two resistances, R_1 and R_2, are connected in parallel. What is the resistance R of the parallel circuit?



The **conductance** $C = 1/R$ of the parallel circuit is

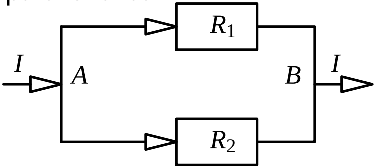
$$C = C_1 + C_2 = 1/R_1 + 1/R_2,$$



$$\therefore \; R = \frac{1}{1/R_1 + 1/R_2} = \tfrac{1}{2} H(R_1, R_2).$$

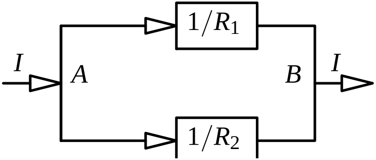
Example: Parallel circuit

Two resistances, R_1 and R_2, are connected in parallel. What is the resistance R of the parallel circuit?



The **conductance** $C = 1/R$ of the parallel circuit is

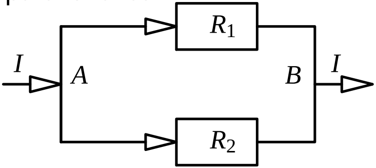
$$C = C_1 + C_2 = 1/R_1 + 1/R_2,$$



$$\therefore\ R = \frac{1}{1/R_1 + 1/R_2} = \tfrac{1}{2} H(R_1, R_2).$$

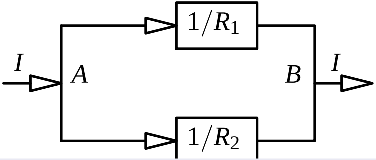
Example: Parallel circuit

Two resistances, R_1 and R_2, are connected in parallel. What is the resistance R of the parallel circuit?



The **conductance** $C = 1/R$ of the parallel circuit is

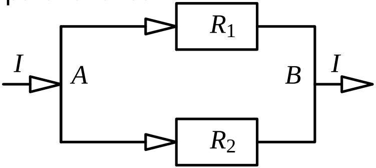
$$C = C_1 + C_2 = 1/R_1 + 1/R_2,$$



$$\therefore\ R = \frac{1}{1/R_1 + 1/R_2} = \frac{1}{2} H(R_1, R_2).$$

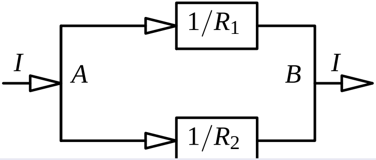
Example: Parallel circuit

Two resistances, R_1 and R_2, are connected in parallel. What is the resistance R of the parallel circuit?



The **conductance** $C = 1/R$ of the parallel circuit is

$$C = C_1 + C_2 = 1/R_1 + 1/R_2,$$



$$\therefore\ R = \frac{1}{1/R_1 + 1/R_2} = \tfrac{1}{2} H(R_1, R_2).$$

Example: Working together

There are n workers, and one job.
Worker i, working alone, can do the job in D_i days, $i \in \overline{1,n}$.

Question: In how many days will the job be done by the n workers working together?

Answer: The job will be done in

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{D_i}} \quad \text{days,}$$

which is

$$\frac{1}{n} H(D_1, \cdots, D_n) \quad \text{days.}$$

If $D_i > 0$, $i \in \overline{1,n}$, then

$$\frac{1}{n} H(D_1, \cdots, D_n) \leq \min\{D_1, \cdots, D_n\} \leq H(D_1, \cdots, D_n) \tag{3}$$

Example: Working together

There are n workers, and one job.
Worker i, working alone, can do the job in D_i days, $i \in \overline{1,n}$.

Question: In how many days will the job be done by the n workers working together?

Answer: The job will be done in

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{D_i}} \quad \text{days,}$$

which is

$$\frac{1}{n} H(D_1, \cdots, D_n) \quad \text{days.}$$

If $D_i > 0$, $i \in \overline{1,n}$, then

$$\frac{1}{n} H(D_1, \cdots, D_n) \leq \min\{D_1, \cdots, D_n\} \leq H(D_1, \cdots, D_n) \tag{3}$$

Example: Working together

There are n workers, and one job.
Worker i, working alone, can do the job in D_i days, $i \in \overline{1,n}$.

Question: In how many days will the job be done by the n workers working together?

Answer: The job will be done in

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{D_i}} \quad \text{days,}$$

which is

$$\frac{1}{n} H(D_1, \cdots, D_n) \quad \text{days.}$$

If $D_i > 0$, $i \in \overline{1,n}$, then

$$\frac{1}{n} H(D_1, \cdots, D_n) \leq \min\{D_1, \cdots, D_n\} \leq H(D_1, \cdots, D_n) \tag{3}$$

Example: Working together

There are n workers, and one job.
Worker i, working alone, can do the job in D_i days, $i \in \overline{1,n}$.

Question: In how many days will the job be done by the n workers working together?

Answer: The job will be done in

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{D_i}} \quad \text{days,}$$

which is

$$\frac{1}{n} H(D_1, \cdots, D_n) \quad \text{days.}$$

If $D_i > 0$, $i \in \overline{1,n}$, then

$$\frac{1}{n} H(D_1, \cdots, D_n) \leq \min\{D_1, \cdots, D_n\} \leq H(D_1, \cdots, D_n) \tag{3}$$

Example: Working together

There are n workers, and one job.
Worker i, working alone, can do the job in D_i days, $i \in \overline{1,n}$.

Question: In how many days will the job be done by the n workers working together?

Answer: The job will be done in

$$\frac{1}{\sum_{i=1}^{n} \frac{1}{D_i}} \quad \text{days,}$$

which is

$$\frac{1}{n} H(D_1, \cdots, D_n) \quad \text{days.}$$

If $D_i > 0$, $i \in \overline{1,n}$, then

$$\frac{1}{n} H(D_1, \cdots, D_n) \leq \min\{D_1, \cdots, D_n\} \leq H(D_1, \cdots, D_n) \tag{3}$$

Detour: A harmonic mean for matrices

For $a, b > 0$,

$$\tfrac{1}{2} H(a,b) = \frac{ab}{a+b}$$

The **parallel sum** of matrices $A, B \in \mathbb{C}^{m\times n}$ is, [2],

$$A : B \;:=\; A\,(A+B)^{\dagger}\,B \tag{4}$$

Let L be a **subspace** of $\mathbb{C}^n$,
P_L the **orthogonal projection** on L, i.e.

$$P_L = P_L^2 = P_L^*, \; R(P_L) = L.$$

Theorem (Anderson & Duffin, [2])

Let L, M be subspaces of $\mathbb{C}^n$,
P_L, P_M the corresponding orthogonal projections.
Then

$$P_{L\cap M} = 2\,P_L\,(P_L + P_M)^{\dagger}\,P_M \tag{5}$$

Detour: A harmonic mean for matrices

For $a, b > 0$,

$$\tfrac{1}{2}H(a,b) = \frac{a\,b}{a+b}$$

The **parallel sum** of matrices $A, B \in \mathbb{C}^{m\times n}$ is, [2],

$$A\!:\!B \;:=\; A\,(A+B)^{\dagger}\,B \tag{4}$$

Let L be a **subspace** of $\mathbb{C}^n$,
P_L the **orthogonal projection** on L, i.e.

$$P_L = P_L^2 = P_L^*,\ R(P_L) = L.$$

Theorem (Anderson & Duffin, [2])

Let L, M be subspaces of $\mathbb{C}^n$,
P_L, P_M the corresponding orthogonal projections.
Then

$$P_{L\bigcap M} = 2\,P_L\,(P_L + P_M)^{\dagger}P_M \tag{5}$$

An extremal principle

Theorem (Morley, [14])

If $A, B \in \mathbb{C}^{n\times n}$ are PSD, then for any $\mathbf{x} \in \mathbb{C}^n$,

$$\langle \mathbf{x}, A : B\mathbf{x}\rangle = \inf\{\langle \mathbf{y}, A\mathbf{y}\rangle + \langle \mathbf{z}, B\mathbf{z}\rangle \; : \; \mathbf{y} + \mathbf{z} = \mathbf{x}\} \tag{6}$$

Example. If resistances R_1, R_2 are connected in parallel, the resulting resistance is

$$R_1 : R_2 = \frac{1}{1/R_1 + 1/R_2} = \frac{R_1 R_2}{R_1 + R_2}$$

A current I through $R_1 : R_2$ splits into currents I_1, I_2 so as to minimize the **dissipated power**

$$(R_1 : R_2) I^2 = \min\{R_1 I_1^2 + R_2 I_2^2 \; : \; I_1 + I_2 = I\}$$

An extremal principle

Theorem (Morley, [14])

If $A, B \in \mathbb{C}^{n\times n}$ are PSD, then for any $\mathbf{x} \in \mathbb{C}^n$,

$$\langle \mathbf{x}, A:B\mathbf{x}\rangle = \inf\{\langle \mathbf{y}, A\mathbf{y}\rangle + \langle \mathbf{z}, B\mathbf{z}\rangle \,:\, \mathbf{y}+\mathbf{z} = \mathbf{x}\} \tag{6}$$

Example. If resistances R_1, R_2 are connected in parallel, the resulting resistance is

$$R_1 : R_2 = \frac{1}{1/R_1 + 1/R_2} = \frac{R_1 R_2}{R_1 + R_2}$$

A current I through $R_1 : R_2$ splits into currents I_1, I_2 so as to minimize the **dissipated power**

$$(R_1 : R_2) I^2 = \min\{R_1 I_1^2 + R_2 I_2^2 \,:\, I_1 + I_2 = I\}$$

An extremal principle

Theorem (Morley, [14])

If $A, B \in \mathbb{C}^{n\times n}$ are PSD, then for any $\mathbf{x} \in \mathbb{C}^n$,

$$\langle \mathbf{x}, A:B\mathbf{x}\rangle = \inf\{\langle \mathbf{y}, A\mathbf{y}\rangle + \langle \mathbf{z}, B\mathbf{z}\rangle \;:\; \mathbf{y}+\mathbf{z}=\mathbf{x}\} \tag{6}$$

Example. If resistances R_1, R_2 are connected in parallel, the resulting resistance is

$$R_1 : R_2 = \frac{1}{1/R_1 + 1/R_2} = \frac{R_1 R_2}{R_1 + R_2}$$

A current I through $R_1 : R_2$ splits into currents I_1, I_2 so as to minimize the **dissipated power**

$$(R_1 : R_2) I^2 = \min\left\{R_1 I_1^2 + R_2 I_2^2 \;:\; I_1 + I_2 = I\right\}$$

An extremal principle

Theorem (Morley, [14])

If $A, B \in \mathbb{C}^{n\times n}$ are PSD, then for any $\mathbf{x} \in \mathbb{C}^n$,

$$\langle \mathbf{x}, A:B\mathbf{x}\rangle = \inf\{\langle \mathbf{y}, A\mathbf{y}\rangle + \langle \mathbf{z}, B\mathbf{z}\rangle \;:\; \mathbf{y}+\mathbf{z} = \mathbf{x}\} \tag{6}$$

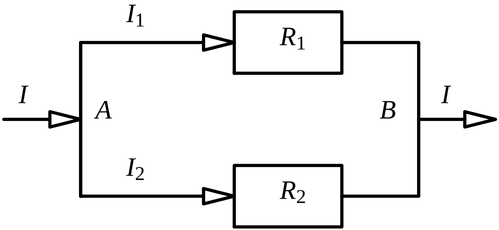
Example. If resistances R_1, R_2 are connected in parallel, the resulting resistance is

$$R_1 : R_2 = \frac{1}{1/R_1 + 1/R_2} = \frac{R_1 R_2}{R_1 + R_2}$$

A current I through $R_1 : R_2$ splits into currents I_1, I_2 so as to minimize the **dissipated power**

$$(R_1 : R_2)\, I^2 = \min\left\{R_1 I_1^2 + R_2 I_2^2 \;:\; I_1 + I_2 = I\right\}$$

Maxwell's variational principle, [9]



The problem

$$\min\{I_1^2 R_1 + I_2^2 R_2 : I_1 + I_2 = I\},$$

has the **Lagrangian**,

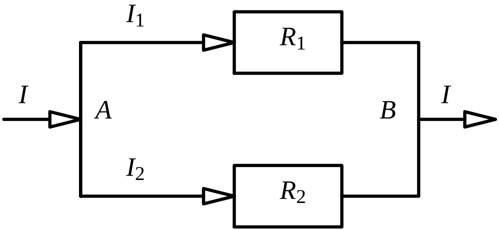
$$L(I_1, I_2, \lambda) = I_1^2 R_1 + I_2^2 R_2 - \lambda(I_1 + I_2 - I).$$

Differentiating L w.r.t. I_1, I_2 results in **Ohm's law**

$$I_1 R_1 = I_2 R_2,$$

the voltage drop from A to B.

Maxwell's variational principle, [9]



The problem

$$\min\{I_1^2R_1 + I_2^2R_2 : I_1 + I_2 = I\},$$

has the **Lagrangian**,

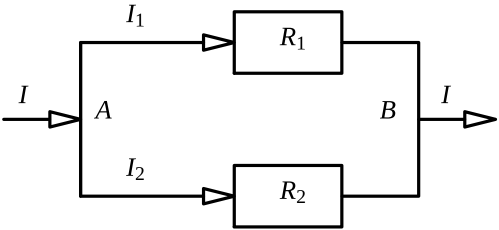
$$L(I_1, I_2, \lambda) = I_1^2R_1 + I_2^2R_2 - \lambda(I_1 + I_2 - I).$$

Differentiating L w.r.t. I_1, I_2 results in **Ohm's law**

$$I_1R_1 = I_2R_2,$$

the voltage drop from A to B.

Maxwell's variational principle, [9]



The problem

$$\min\{I_1^2 R_1 + I_2^2 R_2 : I_1 + I_2 = I\},$$

has the **Lagrangian**,

$$L(I_1, I_2, \lambda) = I_1^2 R_1 + I_2^2 R_2 - \lambda(I_1 + I_2 - I).$$

Differentiating L w.r.t. I_1, I_2 results in **Ohm's law**

$$I_1 R_1 = I_2 R_2,$$

the voltage drop from A to B.

Outline

Inverse distance weighted (IDW) interpolation

Shepard (1968)

A function $u : \mathbb{R}^n \to \mathbb{R}$ is evaluated at K given points $\{\mathbf{x}_k : k \in \overline{1,K}\}$ in $\mathbb{R}^n$, giving the values $\{u_k : k \in \overline{1,K}\}$, respectively. It is required to estimate u at a point $\mathbf{x} \in \operatorname{conv}\{\mathbf{x}_k : k \in \overline{1,K}\}$.

Shepard [15] estimated $u(\mathbf{x})$ as a convex combination,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \lambda_k(\mathbf{x})\, u_k \tag{7}$$

with weights $\lambda_k(\mathbf{x})$ inversely proportional to distances $d(\mathbf{x}, \mathbf{x}_k)$

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x},\mathbf{x}_k)}}{\sum\limits_{j=1}^{K} \dfrac{1}{d(\mathbf{x},\mathbf{x}_j)}} \right) u_k \tag{8}$$

Inverse distance weighted (IDW) interpolation

Shepard (1968)

A function $u : \mathbb{R}^n \to \mathbb{R}$ is evaluated at K given points $\{\mathbf{x}_k : k \in \overline{1,K}\}$ in $\mathbb{R}^n$, giving the values $\{u_k : k \in \overline{1,K}\}$, respectively. It is required to estimate u at a point $\mathbf{x} \in \operatorname{conv}\{\mathbf{x}_k : k \in \overline{1,K}\}$.

Shepard [15] estimated $u(\mathbf{x})$ as a convex combination,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \lambda_k(\mathbf{x})\, u_k \tag{7}$$

with weights $\lambda_k(\mathbf{x})$ inversely proportional to distances $d(\mathbf{x}, \mathbf{x}_k)$

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x}, \mathbf{x}_k)}}{\sum\limits_{j=1}^{K} \dfrac{1}{d(\mathbf{x}, \mathbf{x}_j)}} \right) u_k \tag{8}$$

Inverse distance weighted (IDW) interpolation

Shepard (1968)

A function $u : \mathbb{R}^n \to \mathbb{R}$ is evaluated at K given points $\{\mathbf{x}_k : k \in \overline{1,K}\}$ in $\mathbb{R}^n$, giving the values $\{u_k : k \in \overline{1,K}\}$, respectively. It is required to estimate u at a point $\mathbf{x} \in \mathrm{conv}\{\mathbf{x}_k : k \in \overline{1,K}\}$.

Shepard [15] estimated $u(\mathbf{x})$ as a convex combination,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \lambda_k(\mathbf{x})\, u_k \tag{7}$$

with weights $\lambda_k(\mathbf{x})$ inversely proportional to distances $d(\mathbf{x}, \mathbf{x}_k)$

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x}, \mathbf{x}_k)}}{\sum\limits_{j=1}^{K} \dfrac{1}{d(\mathbf{x}, \mathbf{x}_j)}} \right) u_k \tag{8}$$

Inverse distance weighted (IDW) interpolation

Shepard (1968)

A function $u : \mathbb{R}^n \to \mathbb{R}$ is evaluated at K given points $\{\mathbf{x}_k : k \in \overline{1,K}\}$ in $\mathbb{R}^n$, giving the values $\{u_k : k \in \overline{1,K}\}$, respectively. It is required to estimate u at a point $\mathbf{x} \in \operatorname{conv}\{\mathbf{x}_k : k \in \overline{1,K}\}$.

Shepard [15] estimated $u(\mathbf{x})$ as a convex combination,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \lambda_k(\mathbf{x})\, u_k \tag{7}$$

with weights $\lambda_k(\mathbf{x})$ inversely proportional to distances $d(\mathbf{x}, \mathbf{x}_k)$

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x}, \mathbf{x}_k)}}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x}, \mathbf{x}_j)}} \right) u_k \tag{8}$$

IDW interpolation (cont'd)

The **IDW interpolation** at $\mathbf{x}$ of values u_k given at $\mathbf{x}_k$, $k \in \overline{1,K}$,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x},\mathbf{x}_k)}}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x},\mathbf{x}_j)}} \right) u_k \tag{8}$$

Interpolating the K distances $d(\mathbf{x},\mathbf{x}_k)$, i.e. taking $u_k = d(\mathbf{x},\mathbf{x}_k)$ in (8), gives

$$\frac{K}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x},\mathbf{x}_j)}} \tag{9}$$

the **harmonic mean** of the distances $\{d(\mathbf{x},\mathbf{x}_k) : k \in \overline{1,K}\}$, a measure of how far is $\mathbf{x}$ from the points $\{\mathbf{x}_k\}$.

IDW interpolation (cont'd)

The **IDW interpolation** at $\mathbf{x}$ of values u_k given at $\mathbf{x}_k$, $k \in \overline{1,K}$,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x}, \mathbf{x}_k)}}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x}, \mathbf{x}_j)}} \right) u_k \tag{8}$$

Interpolating the K distances $d(\mathbf{x}, \mathbf{x}_k)$, i.e. taking $u_k = d(\mathbf{x}, \mathbf{x}_k)$ in (8), gives

$$\frac{K}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x}, \mathbf{x}_j)}} \tag{9}$$

the **harmonic mean** of the distances $\{d(\mathbf{x}, \mathbf{x}_k) : k \in \overline{1,K}\}$, a measure of how far is $\mathbf{x}$ from the points $\{\mathbf{x}_k\}$.

IDW interpolation (cont'd)

The **IDW interpolation** at $\mathbf{x}$ of values u_k given at $\mathbf{x}_k$, $k \in \overline{1,K}$,

$$u(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{\dfrac{1}{d(\mathbf{x},\mathbf{x}_k)}}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x},\mathbf{x}_j)}} \right) u_k \tag{8}$$

Interpolating the K distances $d(\mathbf{x},\mathbf{x}_k)$, i.e. taking $u_k = d(\mathbf{x},\mathbf{x}_k)$ in (8), gives

$$\frac{K}{\sum_{j=1}^{K} \dfrac{1}{d(\mathbf{x},\mathbf{x}_j)}} \tag{9}$$

the **harmonic mean** of the distances $\{d(\mathbf{x},\mathbf{x}_k) : k \in \overline{1,K}\}$, a measure of how far is $\mathbf{x}$ from the points $\{\mathbf{x}_k\}$.

Outline

Clusters and distances

Consider a data set $\mathscr{D} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$. We assume that $\mathscr{D}$ is partitioned into K clusters $\mathscr{C}_k$, $1 < K < N$.

With each cluster $\mathscr{C}_k$, we associate:
• a **distance function** d_k, for example the **Mahalanobis distance**

$$d_k(\mathbf{x},\mathbf{y}) = \langle \mathbf{x}-\mathbf{y}, \Sigma_k^{-1}(\mathbf{x}-\mathbf{y})\rangle \tag{10}$$

where Σ_k is the **covariance–matrix** of $\mathscr{C}_k$, and
• a **center** $\mathbf{c}_k$ minimizing the sum of distances to all points in the cluster

$$\mathbf{c}_k := \arg\min_{\mathbf{c}} \sum_{\mathbf{x}\in\mathscr{C}_k} d_k(\mathbf{x},\mathbf{c}) \tag{11}$$

The distance between a point $\mathbf{x}$ and the cluster $\mathscr{C}_k$ is defined as

$$d(\mathbf{x},\mathscr{C}_k) := d_k(\mathbf{x},\mathbf{c}_k) \tag{12}$$

The distance between clusters is not defined, and not needed.

Clusters and distances

Consider a data set $\mathscr{D} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$. We assume that $\mathscr{D}$ is partitioned into K clusters $\mathscr{C}_k$, $1 < K < N$.

With each cluster $\mathscr{C}_k$, we associate:
• a **distance function** d_k, for example the **Mahalanobis distance**

$$d_k(\mathbf{x},\mathbf{y}) = \langle \mathbf{x}-\mathbf{y}, \Sigma_k^{-1}(\mathbf{x}-\mathbf{y})\rangle \tag{10}$$

where Σ_k is the **covariance–matrix** of $\mathscr{C}_k$, and
• a **center** $\mathbf{c}_k$ minimizing the sum of distances to all points in the cluster

$$\mathbf{c}_k := \arg\min_{\mathbf{c}} \sum_{\mathbf{x}\in\mathscr{C}_k} d_k(\mathbf{x},\mathbf{c}) \tag{11}$$

The distance between a point $\mathbf{x}$ and the cluster $\mathscr{C}_k$ is defined as

$$d(\mathbf{x},\mathscr{C}_k) := d_k(\mathbf{x},\mathbf{c}_k) \tag{12}$$

The distance between clusters is not defined, and not needed.

Clusters and distances

Consider a data set $\mathscr{D} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$. We assume that $\mathscr{D}$ is partitioned into K clusters $\mathscr{C}_k$, $1 < K < N$.

With each cluster $\mathscr{C}_k$, we associate:

- a **distance function** d_k, for example the **Mahalanobis distance**

$$d_k(\mathbf{x},\mathbf{y}) = \langle \mathbf{x}-\mathbf{y}, \Sigma_k^{-1}(\mathbf{x}-\mathbf{y})\rangle \tag{10}$$

where Σ_k is the **covariance–matrix** of $\mathscr{C}_k$, and

- a **center** $\mathbf{c}_k$ minimizing the sum of distances to all points in the cluster

$$\mathbf{c}_k := \arg\min_{\mathbf{c}} \sum_{\mathbf{x}\in\mathscr{C}_k} d_k(\mathbf{x},\mathbf{c}) \tag{11}$$

The distance between a point $\mathbf{x}$ and the cluster $\mathscr{C}_k$ is defined as

$$d(\mathbf{x},\mathscr{C}_k) := d_k(\mathbf{x},\mathbf{c}_k) \tag{12}$$

The distance between clusters is not defined, and not needed.

Clusters and distances

Consider a data set $\mathscr{D} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$. We assume that $\mathscr{D}$ is partitioned into K clusters $\mathscr{C}_k$, $1 < K < N$.

With each cluster $\mathscr{C}_k$, we associate:
• a **distance function** d_k, for example the **Mahalanobis distance**

$$d_k(\mathbf{x},\mathbf{y}) = \langle \mathbf{x}-\mathbf{y}, \Sigma_k^{-1}(\mathbf{x}-\mathbf{y})\rangle \tag{10}$$

where Σ_k is the **covariance–matrix** of $\mathscr{C}_k$, and
• a **center** $\mathbf{c}_k$ minimizing the sum of distances to all points in the cluster

$$\mathbf{c}_k := \arg\min_{\mathbf{c}} \sum_{\mathbf{x}\in\mathscr{C}_k} d_k(\mathbf{x},\mathbf{c}) \tag{11}$$

The distance between a point $\mathbf{x}$ and the cluster $\mathscr{C}_k$ is defined as

$$d(\mathbf{x},\mathscr{C}_k) := d_k(\mathbf{x},\mathbf{c}_k) \tag{12}$$

The distance between clusters is not defined, and not needed.

Clusters and distances

Consider a data set $\mathscr{D} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$. We assume that $\mathscr{D}$ is partitioned into K clusters $\mathscr{C}_k$, $1 < K < N$.

With each cluster $\mathscr{C}_k$, we associate:
• a **distance function** d_k, for example the **Mahalanobis distance**

$$d_k(\mathbf{x},\mathbf{y}) = \langle \mathbf{x}-\mathbf{y}, \Sigma_k^{-1}(\mathbf{x}-\mathbf{y})\rangle \tag{10}$$

where Σ_k is the **covariance–matrix** of $\mathscr{C}_k$, and
• a **center** $\mathbf{c}_k$ minimizing the sum of distances to all points in the cluster

$$\mathbf{c}_k := \arg\min_{\mathbf{c}} \sum_{\mathbf{x}\in\mathscr{C}_k} d_k(\mathbf{x},\mathbf{c}) \tag{11}$$

The distance between a point $\mathbf{x}$ and the cluster $\mathscr{C}_k$ is defined as

$$d(\mathbf{x},\mathscr{C}_k) := d_k(\mathbf{x},\mathbf{c}_k) \tag{12}$$

The distance between clusters is not defined, and not needed.

The quasi–linear mean, [1, Section 5.3.2]

Let $I = [a,b]$, $f : I \to I$ continuous, strictly monotonic.
Let $w_1, w_2 \geq 0$; $w_1 + w_2 > 0$.

The **quasi–linear mean** of $x_1, x_2 \in I$ is

$$F(x_1,x_2;w_1,w_2) := f\left(\frac{w_1 f^{-1}(x_1) + w_2 f^{-1}(x_2)}{w_1 + w_2}\right). \tag{13}$$

For $f(t) = t$, (13) gives the weighted arithmetic mean,

$$F(x_1,x_2;w_1,w_2) = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$

For $f(t) = t^{-1}$, $0 \notin I$, (13) gives

$$F(x_1,x_2;w_1,w_2) = \frac{w_1 + w_2}{\frac{w_1}{x_1} + \frac{w_2}{x_2}}$$

the weighted harmonic mean of x_1, x_2.

The quasi–linear mean, [1, Section 5.3.2]

Let $I = [a,b]$, $f : I \to I$ continuous, strictly monotonic.
Let $w_1, w_2 \geq 0$; $w_1 + w_2 > 0$.

The **quasi–linear mean** of $x_1, x_2 \in I$ is

$$F(x_1,x_2;w_1,w_2) := f\left(\frac{w_1 f^{-1}(x_1) + w_2 f^{-1}(x_2)}{w_1 + w_2}\right). \tag{13}$$

For $f(t) = t$, (13) gives the weighted arithmetic mean,

$$F(x_1,x_2;w_1,w_2) = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$

For $f(t) = t^{-1}$, $0 \notin I$, (13) gives

$$F(x_1,x_2;w_1,w_2) = \frac{w_1 + w_2}{\frac{w_1}{x_1} + \frac{w_2}{x_2}}$$

the weighted harmonic mean of x_1, x_2.

The quasi–linear mean, [1, Section 5.3.2]

Let $I = [a,b]$, $f : I \to I$ continuous, strictly monotonic.
Let $w_1, w_2 \geq 0$; $w_1 + w_2 > 0$.

The **quasi–linear mean** of $x_1, x_2 \in I$ is

$$F(x_1, x_2; w_1, w_2) := f\left(\frac{w_1 f^{-1}(x_1) + w_2 f^{-1}(x_2)}{w_1 + w_2}\right). \tag{13}$$

For $f(t) = t$, (13) gives the weighted arithmetic mean,

$$F(x_1, x_2; w_1, w_2) = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$

For $f(t) = t^{-1}$, $0 \notin I$, (13) gives

$$F(x_1, x_2; w_1, w_2) = \frac{w_1 + w_2}{\frac{w_1}{x_1} + \frac{w_2}{x_2}}$$

the weighted harmonic mean of x_1, x_2.

Contour approximation, [3], [10]

Definition

Let $D(\mathbf{x})$ be a function of the K distances $d_k(\mathbf{x}, \mathbf{c}_k)$ and cluster sizes q_k. Then $D(\cdot)$ is a **contour approximation** of $\mathscr{D}$ if

$$D(\mathbf{x}) \leq d_k(\mathbf{x}, \mathbf{c}_k), \ \forall \ \mathbf{x} \in \mathbb{R}^n, \ k \in \overline{1,K} \tag{14}$$

Let $F : \mathbb{R}^{2k} \to \mathbb{R}_+$ be a quasi–linear mean of $d_1(\mathbf{x}, \mathbf{c}_1), \cdots, d_K(\mathbf{x}, \mathbf{c}_K)$ and

$$D(\mathbf{x}) = \frac{F(d_1(\mathbf{x}, \mathbf{c}_1), \cdots, d_K(\mathbf{x}, \mathbf{c}_K); q_1, \cdots, q_K)}{q_1 + \cdots + q_k} \tag{15}$$

where q_k is the **size** of the cluster $\mathscr{C}_k$. Arav [3] listed desirable properties for F and proved

$$F(d_1, \cdots, d_K; q_1, \cdots, q_K) = \frac{q_1 + \cdots + q_k}{q_1/d_1 + \cdots + q_k/d_k} \tag{16}$$

a **weighted harmonic mean** of the distances.

Contour approximation, [3], [10]

Definition

Let $D(\mathbf{x})$ be a function of the K distances $d_k(\mathbf{x}, \mathbf{c}_k)$ and cluster sizes q_k. Then $D(\cdot)$ is a **contour approximation** of $\mathscr{D}$ if

$$D(\mathbf{x}) \leq d_k(\mathbf{x}, \mathbf{c}_k), \ \forall\, \mathbf{x} \in \mathbb{R}^n, \ k \in \overline{1,K} \tag{14}$$

Let $F : \mathbb{R}^{2k} \to \mathbb{R}_+$ be a quasi–linear mean of $d_1(\mathbf{x}, \mathbf{c}_1), \cdots, d_K(\mathbf{x}, \mathbf{c}_K)$ and

$$D(\mathbf{x}) = \frac{F(d_1(\mathbf{x}, \mathbf{c}_1), \cdots, d_K(\mathbf{x}, \mathbf{c}_K); q_1, \cdots, q_K)}{q_1 + \cdots + q_k} \tag{15}$$

where q_k is the **size** of the cluster $\mathscr{C}_k$. Arav [3] listed desirable properties for F and proved

$$F(d_1, \cdots, d_K; q_1, \cdots, q_K) = \frac{q_1 + \cdots + q_k}{q_1/d_1 + \cdots + q_k/d_k} \tag{16}$$

a **weighted harmonic mean** of the distances.

Contour approximation, [3], [10]

Definition

Let $D(\mathbf{x})$ be a function of the K distances $d_k(\mathbf{x}, \mathbf{c}_k)$ and cluster sizes q_k. Then $D(\cdot)$ is a **contour approximation** of $\mathscr{D}$ if

$$D(\mathbf{x}) \leq d_k(\mathbf{x}, \mathbf{c}_k), \ \forall \ \mathbf{x} \in \mathbb{R}^n, \ k \in \overline{1,K} \tag{14}$$

Let $F : \mathbb{R}^{2k} \to \mathbb{R}_+$ be a quasi–linear mean of $d_1(\mathbf{x}, \mathbf{c}_1), \cdots, d_K(\mathbf{x}, \mathbf{c}_K)$ and

$$D(\mathbf{x}) = \frac{F(d_1(\mathbf{x}, \mathbf{c}_1), \cdots, d_K(\mathbf{x}, \mathbf{c}_K); q_1, \cdots, q_K)}{q_1 + \cdots + q_k} \tag{15}$$

where q_k is the **size** of the cluster $\mathscr{C}_k$. Arav [3] listed desirable properties for F and proved

$$F(d_1, \cdots, d_K; q_1, \cdots, q_K) = \frac{q_1 + \cdots + q_k}{q_1/d_1 + \cdots + q_k/d_k} \tag{16}$$

a **weighted harmonic mean** of the distances.

Outline

Probabilities and distances

Given K facilities (stores, gyms, hospitals, etc.) to go to, denote by

$(\mathbf{x} \to k)$ the event that a person at $\mathbf{x}$ goes to the k_{th} facility,

and let

$$p_k(\mathbf{x}) = \text{probability of } (\mathbf{x} \to k),\ k \in \overline{1,K}$$

Assume that at any point $\mathbf{x}$ these probabilities depend on

$d_k(\mathbf{x})$ = the **distance** of $\mathbf{x}$ from the k_{th} facility,

as follows:

a facility is more likely to be chosen the closer it is.

Meaning that for any point $\mathbf{x}$, and any two facilities $i, j \in \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}). \tag{a}$$

Probabilities and distances

Given K facilities (stores, gyms, hospitals, etc.) to go to, denote by

$(\mathbf{x} \to k)$ the event that a person at $\mathbf{x}$ goes to the k_{th} facility,

and let

$$p_k(\mathbf{x}) = \text{probability of } (\mathbf{x} \to k),\ k \in \overline{1,K}$$

Assume that at any point $\mathbf{x}$ these probabilities depend on

$d_k(\mathbf{x}) =$ the **distance** of $\mathbf{x}$ from the k_{th} facility,

as follows:

a facility is more likely to be chosen the closer it is.

Meaning that for any point $\mathbf{x}$, and any two facilities $i, j \in \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}). \quad \text{(a)}$$

Probabilities and distances

Given K facilities (stores, gyms, hospitals, etc.) to go to, denote by

$(\mathbf{x} \to k)$ the event that a person at $\mathbf{x}$ goes to the k_{th} facility,

and let

$$p_k(\mathbf{x}) = \text{probability of } (\mathbf{x} \to k),\ k \in \overline{1,K}$$

Assume that at any point $\mathbf{x}$ these probabilities depend on

$d_k(\mathbf{x}) =$ the **distance** of $\mathbf{x}$ from the k_{th} facility,

as follows:

a facility is more likely to be chosen the closer it is.

Meaning that for any point $\mathbf{x}$, and any two facilities $i, j \in \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}). \qquad \text{(a)}$$

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x})$,

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i, j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\, \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S)$ = **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x}),$

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i, j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\, \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S)$ = **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x}),$

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i, j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\, \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S) =$ **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x}),$

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i, j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\,\mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\,p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S)$ = **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x}),$

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i, j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\, \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S) =$ **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x}),$

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i, j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\, \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S) =$ **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

Desirable properties of $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x}))$

Let $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^K$ map distances $\mathbf{d}(\mathbf{x}) = (d_k(\mathbf{x}))$ into probabilities $\mathbf{p}(\mathbf{x}) = (p_k(\mathbf{x}))$,

$$\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{d}(\mathbf{x})) \tag{17}$$

For any $\mathbf{x}$, $i,j \in \overline{1,K}$, permutation matrix Q, set $S \subset \overline{1,K}$,

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

$$\mathbf{f}(\lambda\, \mathbf{d}(\mathbf{x})) = \mathbf{f}(\mathbf{d}(\mathbf{x})), \text{ for any } \lambda > 0 \tag{b}$$

$$Q\,\mathbf{p}(\mathbf{x}) = \mathbf{f}(Q\,\mathbf{d}(\mathbf{x})), \tag{c}$$

$$\mathbf{f} \text{ is continuous,} \tag{d}$$

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}), \ \forall\, k \in S, \tag{e}$$

$$\text{where} \quad p_S(\mathbf{x}) := \sum_{s \in S} p_s(\mathbf{x}),$$

and $p_k(\mathbf{x}|S) =$ **conditional probability** of $(\mathbf{x} \to k)$ given $(\mathbf{x} \to S)$.

The choice axiom of Luce

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}),\ \forall\, k \in S \subset \overline{1,K} \tag{e}$$

is the **choice axiom** of Luce [12], shown equivalent to

$$p_k(\mathbf{x}|S) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in S} v_s(\mathbf{x})}, \tag{18}$$

where $v_k(\mathbf{x})$ is a **scale function**, in particular,

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in\overline{1,K}} v_s(\mathbf{x})},\ \forall\, k \in \overline{1,K}. \tag{19}$$

Therefore, for all $k \in \overline{1,K}$,

$$p_k(\mathbf{x})\, v_k^{-1}(\mathbf{x}) = D(\mathbf{x}), \tag{20}$$

a function of $\mathbf{x}$, **independent** of k.

The choice axiom of Luce

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\,p_S(\mathbf{x}),\ \forall\ k \in S \subset \overline{1,K} \tag{e}$$

is the **choice axiom** of Luce [12], shown equivalent to

$$p_k(\mathbf{x}|S) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in S} v_s(\mathbf{x})}, \tag{18}$$

where $v_k(\mathbf{x})$ is a **scale function**, in particular,

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in\overline{1,K}} v_s(\mathbf{x})},\ \forall\ k \in \overline{1,K}. \tag{19}$$

Therefore, for all $k \in \overline{1,K}$,

$$p_k(\mathbf{x})\,v_k^{-1}(\mathbf{x}) = D(\mathbf{x}), \tag{20}$$

a function of $\mathbf{x}$, **independent** of k.

The choice axiom of Luce

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\, p_S(\mathbf{x}),\ \forall\ k \in S \subset \overline{1,K} \tag{e}$$

is the **choice axiom** of Luce [12], shown equivalent to

$$p_k(\mathbf{x}|S) = \frac{v_k(\mathbf{x})}{\sum_{s \in S} v_s(\mathbf{x})}, \tag{18}$$

where $v_k(\mathbf{x})$ is a **scale function**, in particular,

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum_{s \in \overline{1,K}} v_s(\mathbf{x})}\ ,\ \forall\ k \in \overline{1,K}. \tag{19}$$

Therefore, for all $k \in \overline{1,K}$,

$$p_k(\mathbf{x})\, v_k^{-1}(\mathbf{x}) = D(\mathbf{x}), \tag{20}$$

a function of $\mathbf{x}$, **independent** of k.

The choice axiom of Luce

$$p_k(\mathbf{x}) = p_k(\mathbf{x}|S)\,p_S(\mathbf{x}),\ \forall\ k \in S \subset \overline{1,K} \tag{e}$$

is the **choice axiom** of Luce [12], shown equivalent to

$$p_k(\mathbf{x}|S) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in S} v_s(\mathbf{x})}, \tag{18}$$

where $v_k(\mathbf{x})$ is a **scale function**, in particular,

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in\overline{1,K}} v_s(\mathbf{x})}\ ,\ \forall\ k \in \overline{1,K}. \tag{19}$$

Therefore, for all $k \in \overline{1,K}$,

$$p_k(\mathbf{x})\,v_k^{-1}(\mathbf{x}) = D(\mathbf{x}), \tag{20}$$

a function of $\mathbf{x}$, **independent** of k.

$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k$

From

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \Longrightarrow\ p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

and

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in\overline{1,K}} v_s(\mathbf{x})}\ . \tag{19}$$

it follows that $v_k(\mathbf{x})$ is a **decreasing function** of $d_k(\mathbf{x})$, in particular,

$$v_k(\mathbf{x}) = \frac{1}{d_k(\mathbf{x})}$$

and

$$p_k(\mathbf{x})\, v_k^{-1}(\mathbf{x}) = D(\mathbf{x}) \tag{20}$$

becomes

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}) \tag{A}$$

independent of k. This is our working principle.

$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k$

From

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \Longrightarrow p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

and

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum\limits_{s\in\overline{1,K}} v_s(\mathbf{x})}\,. \tag{19}$$

it follows that $v_k(\mathbf{x})$ is a **decreasing function** of $d_k(\mathbf{x})$, in particular,

$$v_k(\mathbf{x}) = \frac{1}{d_k(\mathbf{x})}$$

and

$$p_k(\mathbf{x})\, v_k^{-1}(\mathbf{x}) = D(\mathbf{x}) \tag{20}$$

becomes

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}) \tag{A}$$

independent of k. This is our working principle.

$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\ k$

From

$$d_i(\mathbf{x}) < d_j(\mathbf{x}) \implies p_i(\mathbf{x}) > p_j(\mathbf{x}), \tag{a}$$

and

$$p_k(\mathbf{x}) = \frac{v_k(\mathbf{x})}{\sum_{s \in \overline{1,K}} v_s(\mathbf{x})} \,. \tag{19}$$

it follows that $v_k(\mathbf{x})$ is a **decreasing function** of $d_k(\mathbf{x})$, in particular,

$$v_k(\mathbf{x}) = \frac{1}{d_k(\mathbf{x})}$$

and

$$p_k(\mathbf{x})\, v_k^{-1}(\mathbf{x}) = D(\mathbf{x}) \tag{20}$$

becomes

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}) \tag{A}$$

independent of k. This is our working principle.

Probabilities

From

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k \in \overline{1,K}, \tag{A}$$

and the fact that probabilities add to 1, we get

$$p_k(\mathbf{x}) = \frac{1/d_k(\mathbf{x})}{\sum_{i=1}^{K} 1/d_i(\mathbf{x})},\ k \in \overline{1,K}. \tag{21}$$

In particular, for $K = 2$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},\ p_2(\mathbf{x}) = \frac{d_1(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

Probabilities

From

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k \in \overline{1,K}, \tag{A}$$

and the fact that probabilities add to 1, we get

$$p_k(\mathbf{x}) = \frac{1/d_k(\mathbf{x})}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})},\ k \in \overline{1,K}. \tag{21}$$

In particular, for $K = 2$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},\ p_2(\mathbf{x}) = \frac{d_1(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

Probabilities

From

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k \in \overline{1,K}, \tag{A}$$

and the fact that probabilities add to 1, we get

$$p_k(\mathbf{x}) = \frac{1/d_k(\mathbf{x})}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})},\ k \in \overline{1,K}. \tag{21}$$

In particular, for $K = 2$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},\ p_2(\mathbf{x}) = \frac{d_1(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

Probabilities

From

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\ k \in \overline{1,K}, \tag{A}$$

and the fact that probabilities add to 1, we get

$$p_k(\mathbf{x}) = \frac{1/d_k(\mathbf{x})}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})},\ k \in \overline{1,K}. \tag{21}$$

In particular, for $K = 2$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},\ p_2(\mathbf{x}) = \frac{d_1(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

Probabilities

From

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\ k \in \overline{1,K}, \tag{A}$$

and the fact that probabilities add to 1, we get

$$p_k(\mathbf{x}) = \frac{1/d_k(\mathbf{x})}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})},\ k \in \overline{1,K}. \tag{21}$$

In particular, for $K = 2$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},\ p_2(\mathbf{x}) = \frac{d_1(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$p_1(\mathbf{x}) = \frac{d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

The joint distance function (JDF)

The function $D(\cdot)$ in

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k \in \overline{1,K}, \tag{A}$$

is called the **Joint Distance Function (JDF)**.

From and (A) and (21) we get

$$D(\mathbf{x}) = \frac{1}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})} = \frac{1}{K} H(d_1(\mathbf{x}), \cdots, d_K(\mathbf{x})). \tag{22}$$

In particular, for $K = 2$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

The joint distance function (JDF)

The function $D(\cdot)$ in

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\ k \in \overline{1,K}, \tag{A}$$

is called the **Joint Distance Function (JDF)**.
From and (A) and (21) we get

$$D(\mathbf{x}) = \frac{1}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})} = \frac{1}{K} H(d_1(\mathbf{x}), \cdots, d_K(\mathbf{x})). \tag{22}$$

In particular, for $K = 2$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

The joint distance function (JDF)

The function $D(\cdot)$ in

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}), \ \forall\, k \in \overline{1,K}, \tag{A}$$

is called the **Joint Distance Function (JDF)**.
From and (A) and (21) we get

$$D(\mathbf{x}) = \frac{1}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})} = \frac{1}{K} H(d_1(\mathbf{x}), \cdots, d_K(\mathbf{x})). \tag{22}$$

In particular, for $K = 2$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})}, \text{ etc.}$$

The joint distance function (JDF)

The function $D(\cdot)$ in

$$p_k(\mathbf{x})\,d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k \in \overline{1,K}, \tag{A}$$

is called the **Joint Distance Function (JDF)**.
From and (A) and (21) we get

$$D(\mathbf{x}) = \frac{1}{\sum_{i=1}^{K} 1/d_i(\mathbf{x})} = \frac{1}{K} H(d_1(\mathbf{x}),\cdots,d_K(\mathbf{x})). \tag{22}$$

In particular, for $K = 2$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\,d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\,d_2(\mathbf{x})\,d_3(\mathbf{x})}{d_1(\mathbf{x})\,d_2(\mathbf{x}) + d_1(\mathbf{x})\,d_3(\mathbf{x}) + d_2(\mathbf{x})\,d_3(\mathbf{x})}, \text{ etc.}$$

The joint distance function (JDF)

The function $D(\cdot)$ in

$$p_k(\mathbf{x})\, d_k(\mathbf{x}) = D(\mathbf{x}),\ \forall\, k \in \overline{1{,}K}, \tag{A}$$

is called the **Joint Distance Function (JDF)**.
From and (A) and (21) we get

$$D(\mathbf{x}) = \frac{1}{\sum\limits_{i=1}^{K} 1/d_i(\mathbf{x})} = \frac{1}{K} H(d_1(\mathbf{x}), \cdots, d_K(\mathbf{x})). \tag{22}$$

In particular, for $K = 2$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})},$$

and for $K = 3$,

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})\, d_3(\mathbf{x})}{d_1(\mathbf{x})\, d_2(\mathbf{x}) + d_1(\mathbf{x})\, d_3(\mathbf{x}) + d_2(\mathbf{x})\, d_3(\mathbf{x})},\ \text{etc.}$$

Outline

An extremal principle for probabilities

Let $K = 2$, and let $d_1(\mathbf{x}), d_2(\mathbf{x})$ be given. The principle

$$p_1(\mathbf{x})d_1(\mathbf{x}) = p_2(\mathbf{x})d_2(\mathbf{x}) \, , \tag{A}$$

is an **optimality condition** for the problem

$$\begin{aligned} \min \quad & d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 \\ \text{s.t.} \quad & p_1 + p_2 = 1 \\ & p_1, p_2 \geq 0 \end{aligned} \tag{P}$$

as shown by differentiating the **Lagrangian**,

$$L(p_1, p_2, \lambda) = d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 - \lambda (p_1 + p_2 - 1)$$

with respect to p_1, p_2,

$$\frac{\partial L}{\partial p_1} = 2 p_1 d_1(\mathbf{x}) - \lambda = 0 \, , \text{ etc.}$$

An extremal principle for probabilities

Let $K = 2$, and let $d_1(\mathbf{x}), d_2(\mathbf{x})$ be given. The principle

$$p_1(\mathbf{x})d_1(\mathbf{x}) = p_2(\mathbf{x})d_2(\mathbf{x}) \ , \tag{A}$$

is an **optimality condition** for the problem

$$\begin{aligned} \min \quad & d_1(\mathbf{x})\,p_1^2 + d_2(\mathbf{x})\,p_2^2 \\ \text{s.t.} \quad & p_1 + p_2 = 1 \\ & p_1, p_2 \geq 0 \end{aligned} \tag{P}$$

as shown by differentiating the **Lagrangian**,

$$L(p_1, p_2, \lambda) = d_1(\mathbf{x})\,p_1^2 + d_2(\mathbf{x})\,p_2^2 - \lambda(p_1 + p_2 - 1)$$

with respect to p_1, p_2,

$$\frac{\partial L}{\partial p_1} = 2p_1\,d_1(\mathbf{x}) - \lambda = 0 \ , \text{ etc.}$$

An extremal principle for probabilities

Let $K = 2$, and let $d_1(\mathbf{x}), d_2(\mathbf{x})$ be given. The principle

$$p_1(\mathbf{x})d_1(\mathbf{x}) = p_2(\mathbf{x})d_2(\mathbf{x}) \ , \qquad \text{(A)}$$

is an **optimality condition** for the problem

$$\begin{aligned} \min \quad & d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 \\ \text{s.t.} \quad & p_1 + p_2 = 1 \\ & p_1, p_2 \geq 0 \end{aligned} \qquad \text{(P)}$$

as shown by differentiating the **Lagrangian**,

$$L(p_1, p_2, \lambda) = d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 - \lambda\,(p_1 + p_2 - 1)$$

with respect to p_1, p_2,

$$\frac{\partial L}{\partial p_1} = 2\, p_1\, d_1(\mathbf{x}) - \lambda = 0 \ , \ \text{etc.}$$

Did he say p^2 ? Yes

The problem

$$\min\{p_1^2 d_1 + p_2^2 d_2 : p_1 + p_2 = 1\ ,\ p_1, p_2 \geq 0\}$$

is a **smoothed** version of

$$\min\{p_1 d_1 + p_2 d_2 : p_1 + p_2 = 1\ ,\ p_1, p_2 \geq 0\} \Longrightarrow \min\{d_1, d_2\}\ ,$$

see Teboulle, [16].

Other schemes include **entropic smoothing**

$$\min\{p_1 d_1 + p_2 d_2 + p_1 \log p_1 + p_2 \log p_2 : p_1 + p_2 = 1\ ,\ p_1, p_2 \geq 0\}$$

resulting in the principle

$$p_k(\mathbf{x})\, e^{d_k(\mathbf{x})} = E(\mathbf{x}),\ k \in \overline{1,K},$$

where $E(\mathbf{x})$ does not depend on k.

Did he say p^2 ? Yes

The problem

$$\min\{p_1^2 d_1 + p_2^2 d_2 : p_1 + p_2 = 1\ ,\ p_1, p_2 \geq 0\}$$

is a **smoothed** version of

$$\min\{p_1 d_1 + p_2 d_2 : p_1 + p_2 = 1\ ,\ p_1, p_2 \geq 0\} \Longrightarrow \min\{d_1, d_2\}\ ,$$

see Teboulle, [16].

Other schemes include **entropic smoothing**

$$\min\{p_1 d_1 + p_2 d_2 + p_1 \log p_1 + p_2 \log p_2 : p_1 + p_2 = 1\ ,\ p_1, p_2 \geq 0\}$$

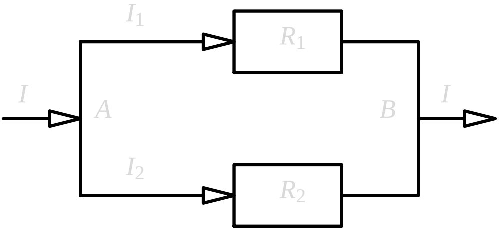
resulting in the principle

$$p_k(\mathbf{x})\, e^{d_k(\mathbf{x})} = E(\mathbf{x}),\ k \in \overline{1,K},$$

where $E(\mathbf{x})$ does not depend on k.

An electrical analogy: Parallel circuit

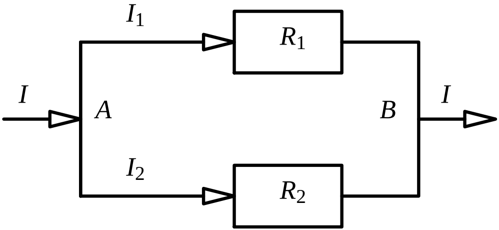
$$\min\{p_1^2 d_1 + p_2^2 d_2 : p_1 + p_2 = 1, p_1, p_2 \geq 0\}$$
$$p_1 d_1 = p_2 d_2$$



$$\min\{I_1^2 R_1 + I_2^2 R_2 : I_1 + I_2 = I\}$$
$$I_1 R_1 = I_2 R_2 = \text{voltage drop from } A \text{ to } B$$

An electrical analogy: Parallel circuit

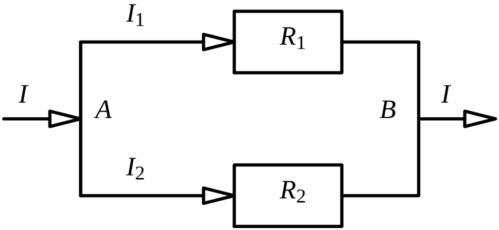
$$\min\{p_1^2 d_1 + p_2^2 d_2 : p_1 + p_2 = 1, p_1, p_2 \geq 0\}$$
$$p_1 d_1 = p_2 d_2$$



$$\min\{I_1^2 R_1 + I_2^2 R_2 : I_1 + I_2 = I\}$$
$$I_1 R_1 = I_2 R_2 = \text{voltage drop from } A \text{ to } B$$

An electrical analogy: Parallel circuit

$$\min\{p_1^2 d_1 + p_2^2 d_2 : p_1 + p_2 = 1, p_1, p_2 \geq 0\}$$
$$p_1 d_1 = p_2 d_2$$



$$\min\{I_1^2 R_1 + I_2^2 R_2 : I_1 + I_2 = I\}$$
$$I_1 R_1 = I_2 R_2 = \text{voltage drop from } A \text{ to } B$$

The JDF as optimal value

Using

$$p_k \, d_k(\mathbf{x}) = D(\mathbf{x}) \tag{A}$$

rewrite

$$\begin{aligned} \min \quad & d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 \\ \text{s.t.} \quad & p_1 + p_2 = 1 \\ & p_1, p_2 \geq 0 \end{aligned} \tag{P}$$

as

$$\min \quad D(\mathbf{x})(p_1 + p_2), \text{ etc.}$$

Therefore the **optimal value** of (P) is the JDF

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})}$$

The JDF as optimal value

Using

$$p_k\, d_k(\mathbf{x}) = D(\mathbf{x}) \tag{A}$$

rewrite

$$\begin{aligned} \min \quad & d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 \\ \text{s.t.} \quad & p_1 + p_2 = 1 \\ & p_1, p_2 \geq 0 \end{aligned} \tag{P}$$

as

$$\min \quad D(\mathbf{x})(p_1 + p_2), \text{ etc.}$$

Therefore the **optimal value** of (P) is the JDF

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})}$$

The JDF as optimal value

Using

$$p_k \, d_k(\mathbf{x}) = D(\mathbf{x}) \tag{A}$$

rewrite

$$\begin{aligned} \min \quad & d_1(\mathbf{x})\, p_1^2 + d_2(\mathbf{x})\, p_2^2 \\ \text{s.t.} \quad & p_1 + p_2 = 1 \\ & p_1, p_2 \geq 0 \end{aligned} \tag{P}$$

as

$$\min \quad D(\mathbf{x})(p_1 + p_2), \text{ etc.}$$

Therefore the **optimal value** of (P) is the JDF

$$D(\mathbf{x}) = \frac{d_1(\mathbf{x})\, d_2(\mathbf{x})}{d_1(\mathbf{x}) + d_2(\mathbf{x})}$$

Outline

The Fermat–Weber Location Problem

Given $\mathbf{X} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$ and weights $w_i > 0$, find $\mathbf{c}$ minimizing

$$f(\mathbf{c}) = \sum_{i=1}^{N} w_i \, \|\mathbf{x}_i - \mathbf{c}\|, \quad \|\cdot\| \text{ Euclidean.}$$

The **gradient** (for $\mathbf{c} \notin \mathbf{X}$)

$$\nabla f(\mathbf{c}) = -\sum_{i=1}^{N} w_i \frac{\mathbf{x}_i - \mathbf{c}}{\|\mathbf{x}_i - \mathbf{c}\|}$$

the **resultant** of forces w_i, with direction from $\mathbf{c}$ to $\mathbf{x}_i$.

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \,, \; \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\displaystyle\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

expressing $\mathbf{c}$ as a convex combination of $\{\mathbf{x}_i : i \in \overline{1,N}\}$.

The Fermat–Weber Location Problem

Given $\mathbf{X} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$ and weights $w_i > 0$, find $\mathbf{c}$ minimizing

$$f(\mathbf{c}) = \sum_{i=1}^{N} w_i \|\mathbf{x}_i - \mathbf{c}\|, \quad \|\cdot\| \text{ Euclidean.}$$

The **gradient** (for $\mathbf{c} \notin \mathbf{X}$)

$$\nabla f(\mathbf{c}) = -\sum_{i=1}^{N} w_i \frac{\mathbf{x}_i - \mathbf{c}}{\|\mathbf{x}_i - \mathbf{c}\|}$$

the **resultant** of forces w_i, with direction from $\mathbf{c}$ to $\mathbf{x}_i$.

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \,, \ \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\displaystyle\sum_{j=1}^{N} \frac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

expressing $\mathbf{c}$ as a convex combination of $\{\mathbf{x}_i : i \in \overline{1,N}\}$.

The Fermat–Weber Location Problem

Given $\mathbf{X} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$ and weights $w_i > 0$, find $\mathbf{c}$ minimizing

$$f(\mathbf{c}) = \sum_{i=1}^{N} w_i \|\mathbf{x}_i - \mathbf{c}\|, \quad \|\cdot\| \text{ Euclidean.}$$

The **gradient** (for $\mathbf{c} \notin \mathbf{X}$)

$$\nabla f(\mathbf{c}) = -\sum_{i=1}^{N} w_i \frac{\mathbf{x}_i - \mathbf{c}}{\|\mathbf{x}_i - \mathbf{c}\|}$$

the **resultant** of forces w_i, with direction from $\mathbf{c}$ to $\mathbf{x}_i$.

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \,,\ \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

expressing $\mathbf{c}$ as a convex combination of $\{\mathbf{x}_i : i \in \overline{1,N}\}$.

The Fermat–Weber Location Problem

Given $\mathbf{X} = \{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$ and weights $w_i > 0$, find $\mathbf{c}$ minimizing

$$f(\mathbf{c}) = \sum_{i=1}^{N} w_i \, \|\mathbf{x}_i - \mathbf{c}\|, \quad \|\cdot\| \text{ Euclidean.}$$

The **gradient** (for $\mathbf{c} \notin \mathbf{X}$)

$$\nabla f(\mathbf{c}) = -\sum_{i=1}^{N} w_i \, \frac{\mathbf{x}_i - \mathbf{c}}{\|\mathbf{x}_i - \mathbf{c}\|}$$

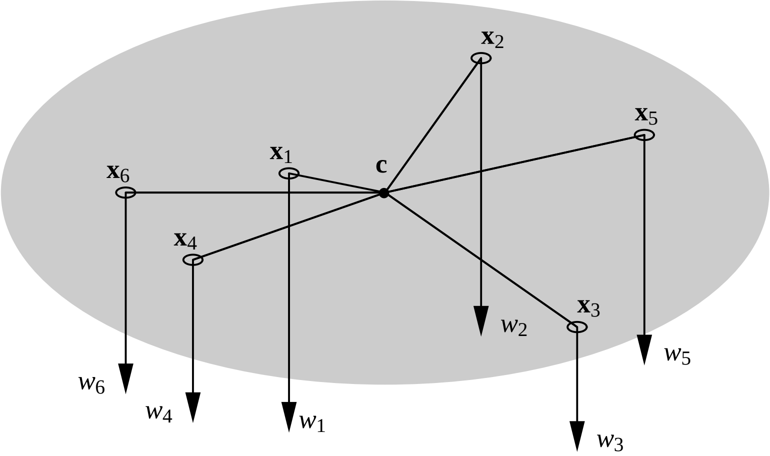
the **resultant** of forces w_i, with direction from $\mathbf{c}$ to $\mathbf{x}_i$.

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \, \mathbf{x}_i \; , \; \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\displaystyle\sum_{j=1}^{N} \frac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

expressing $\mathbf{c}$ as a convex combination of $\{\mathbf{x}_i : i \in \overline{1,N}\}$.

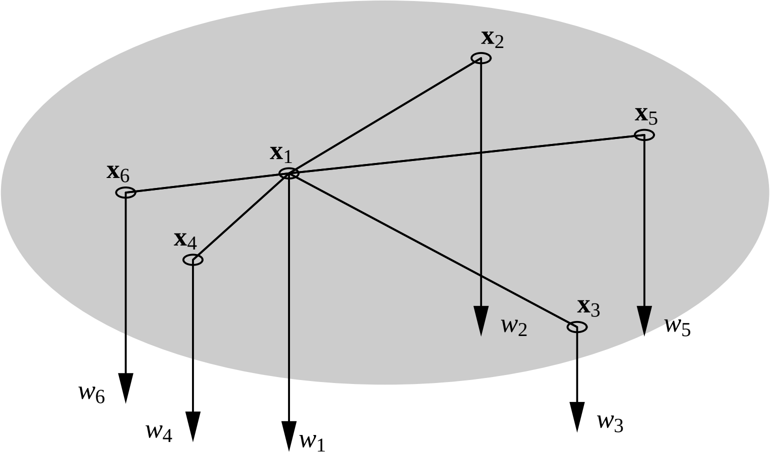
A mechanical solution

The point $\mathbf{c}$ is tied to the weights w_i through holes in $\mathbf{x}_i$.
$\mathbf{c}$ is free to move, will come to rest at the optimal center.

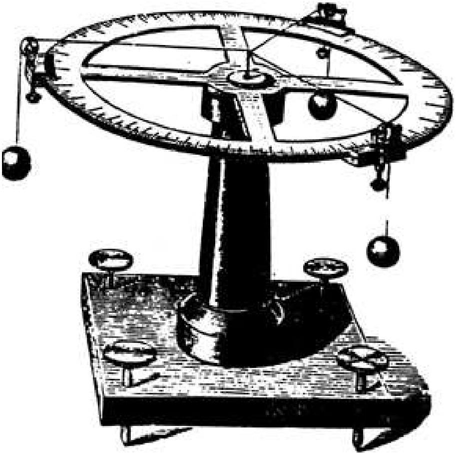


What if $\mathbf{c}$ falls in one of the holes?

The point $\mathbf{c}$ may settle at one of the holes, say $\mathbf{c} = \mathbf{x}_1$, if the weight w_1 is greater than the resultant of the other weights.



The Varignon frame



Pierre Varignon (1654–1722)

The Weiszfeld Method, [17]

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \,,\; \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\displaystyle\sum_{j=1}^{N} \frac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

The **Weiszfeld method** is the iterations

$$\mathbf{c}_+ = T(\mathbf{c}) := \begin{cases} \displaystyle\sum_{i=1}^{N} \left(\frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}} \right) \mathbf{x}_i & , \mathbf{c} \notin \mathbf{X}; \\ \mathbf{c} & , \mathbf{c} \in \mathbf{X}. \end{cases}$$

Endre Weiszfeld, Andrew Vazsonyi (1916–2003)

The Weiszfeld Method, [17]

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \, , \; \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

The **Weiszfeld method** is the iterations

$$\mathbf{c}_+ = T(\mathbf{c}) := \begin{cases} \sum_{i=1}^{N} \left(\dfrac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}} \right) \mathbf{x}_i & , \mathbf{c} \notin \mathbf{X}; \\ \mathbf{c} & , \mathbf{c} \in \mathbf{X}. \end{cases}$$

Endre Weiszfeld, Andrew Vazsonyi (1916–2003)

The Weiszfeld Method, [17]

$$\nabla f(\mathbf{c}) = 0 \quad \Longrightarrow \quad \mathbf{c} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \,, \; \lambda_i = \frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\displaystyle\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}}$$

The **Weiszfeld method** is the iterations

$$\mathbf{c}_+ = T(\mathbf{c}) := \begin{cases} \displaystyle\sum_{i=1}^{N} \left(\frac{\dfrac{w_i}{\|\mathbf{x}_i - \mathbf{c}\|}}{\sum_{j=1}^{N} \dfrac{w_j}{\|\mathbf{x}_j - \mathbf{c}\|}} \right) \mathbf{x}_i & , \mathbf{c} \notin \mathbf{X}; \\ \mathbf{c} & , \mathbf{c} \in \mathbf{X}. \end{cases}$$

Endre Weiszfeld, Andrew Vazsonyi (1916–2003)

A multi–facility location problem

Given $\{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$, weights $w_i > 0$, integer K, $1 \le K < N$,

Locate facilities $\{\mathbf{c}_k : k \in \overline{1,K}\}$ so as to

$$\text{minimize } f(\mathbf{c}_1, \ldots, \mathbf{c}_K) = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, \|\mathbf{x}_i - \mathbf{c}_k\| \qquad (\text{L.}K)$$

where $\mathscr{C}_k = \{\mathbf{x} : \mathbf{x} \text{ assigned to } \mathbf{c}_k\}$.

The Fermat–Weber problem is the special case (L.1).

For $K > 1$, the problem (L.K) is NP hard, [13].

A multi–facility location problem

Given $\{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$, weights $w_i > 0$, integer K, $1 \leq K < N$,

Locate facilities $\{\mathbf{c}_k : k \in \overline{1,K}\}$ so as to

$$\text{minimize } f(\mathbf{c}_1, \ldots, \mathbf{c}_K) = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, \|\mathbf{x}_i - \mathbf{c}_k\| \qquad \text{(L.}K\text{)}$$

where $\mathscr{C}_k = \{\mathbf{x} : \mathbf{x} \text{ assigned to } \mathbf{c}_k\}$.

The Fermat–Weber problem is the special case (L.1).

For $K > 1$, the problem (L.K) is NP hard, [13].

A multi–facility location problem

Given $\{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$, weights $w_i > 0$, integer K, $1 \leq K < N$,

Locate facilities $\{\mathbf{c}_k : k \in \overline{1,K}\}$ so as to

$$\text{minimize } f(\mathbf{c}_1, \ldots, \mathbf{c}_K) = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \left\| \mathbf{x}_i - \mathbf{c}_k \right\| \tag{L.K}$$

where $\mathscr{C}_k = \{\mathbf{x} : \mathbf{x} \text{ assigned to } \mathbf{c}_k\}$.

The Fermat–Weber problem is the special case (L.1).

For $K > 1$, the problem (L.K) is NP hard, [13].

A multi–facility location problem

Given $\{\mathbf{x}_i : i \in \overline{1,N}\} \subset \mathbb{R}^n$, weights $w_i > 0$, integer K, $1 \le K < N$,

Locate facilities $\{\mathbf{c}_k : k \in \overline{1,K}\}$ so as to

$$\text{minimize } f(\mathbf{c}_1, \ldots, \mathbf{c}_K) = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, \|\mathbf{x}_i - \mathbf{c}_k\| \qquad \text{(L.}K\text{)}$$

where $\mathscr{C}_k = \{\mathbf{x} : \mathbf{x} \text{ assigned to } \mathbf{c}_k\}$.

The Fermat–Weber problem is the special case (L.1).

For $K > 1$, the problem (L.K) is NP hard, [13].

Probabilistic assignments, [5], [11]

The multi–facility location problem

$$\min \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, d(\mathbf{x}_i, \mathbf{c}_k) \tag{L.K}$$

with $d(\mathbf{x}_i, \mathbf{c}_k) = \|\mathbf{x}_i - \mathbf{c}_k\|$, is approximated by

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d(\mathbf{x}_i, \mathbf{c}_k) \tag{P.K}$$

where $\{p_k(\mathbf{x}_i)\}$ are the **cluster membership probabilities**,

$$p_k(\mathbf{x}_i) = \text{Prob}\,\{\mathbf{x}_i \in \mathscr{C}_k\}, \; k \in \overline{1,K}, \, i \in \overline{1,N}$$

The problem (P.K) has two sets of variables,

- **centers** $\{\mathbf{c}_k\}$, as before, and
- **probabilities** $\{p_k(\mathbf{x}_i)\}$, corresponding to the **assignments**.

Probabilistic assignments, [5], [11]

The multi–facility location problem

$$\min \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, d(\mathbf{x}_i, \mathbf{c}_k) \qquad \text{(L.}K\text{)}$$

with $d(\mathbf{x}_i, \mathbf{c}_k) = \|\mathbf{x}_i - \mathbf{c}_k\|$, is approximated by

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d(\mathbf{x}_i, \mathbf{c}_k) \qquad \text{(P.}K\text{)}$$

where $\{p_k(\mathbf{x}_i)\}$ are the **cluster membership probabilities**,

$$p_k(\mathbf{x}_i) = \text{Prob}\,\{\mathbf{x}_i \in \mathscr{C}_k\}, \; k \in \overline{1,K}, \; i \in \overline{1,N}$$

The problem (P.K) has two sets of variables,

- **centers** $\{\mathbf{c}_k\}$, as before, and
- **probabilities** $\{p_k(\mathbf{x}_i)\}$, corresponding to the **assignments**.

Probabilistic assignments, [5], [11]

The multi–facility location problem

$$\min \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, d(\mathbf{x}_i, \mathbf{c}_k) \tag{L.K}$$

with $d(\mathbf{x}_i, \mathbf{c}_k) = \|\mathbf{x}_i - \mathbf{c}_k\|$, is approximated by

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d(\mathbf{x}_i, \mathbf{c}_k) \tag{P.K}$$

where $\{p_k(\mathbf{x}_i)\}$ are the **cluster membership probabilities**,

$$p_k(\mathbf{x}_i) = \text{Prob}\,\{\mathbf{x}_i \in \mathscr{C}_k\}, \; k \in \overline{1,K}, \; i \in \overline{1,N}$$

The problem (P.K) has two sets of variables,

- **centers** $\{\mathbf{c}_k\}$, as before, and
- **probabilities** $\{p_k(\mathbf{x}_i)\}$, corresponding to the **assignments**.

Probabilistic assignments, [5], [11]

The multi–facility location problem

$$\min \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, d(\mathbf{x}_i, \mathbf{c}_k) \tag{L.K}$$

with $d(\mathbf{x}_i, \mathbf{c}_k) = \|\mathbf{x}_i - \mathbf{c}_k\|$, is approximated by

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d(\mathbf{x}_i, \mathbf{c}_k) \tag{P.K}$$

where $\{p_k(\mathbf{x}_i)\}$ are the **cluster membership probabilities**,

$$p_k(\mathbf{x}_i) = \text{Prob}\,\{\mathbf{x}_i \in \mathscr{C}_k\}, \; k \in \overline{1,K}, \, i \in \overline{1,N}$$

The problem (P.K) has two sets of variables,

centers $\{\mathbf{c}_k\}$, as before, and

probabilities $\{p_k(\mathbf{x}_i)\}$, corresponding to the **assignments**.

Probabilistic assignments, [5], [11]

The multi–facility location problem

$$\min \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathscr{C}_k} w_i \, d(\mathbf{x}_i, \mathbf{c}_k) \qquad \text{(L.}K\text{)}$$

with $d(\mathbf{x}_i, \mathbf{c}_k) = \|\mathbf{x}_i - \mathbf{c}_k\|$, is approximated by

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d(\mathbf{x}_i, \mathbf{c}_k) \qquad \text{(P.}K\text{)}$$

where $\{p_k(\mathbf{x}_i)\}$ are the **cluster membership probabilities**,

$$p_k(\mathbf{x}_i) = \text{Prob}\,\{\mathbf{x}_i \in \mathscr{C}_k\}, \; k \in \overline{1,K}, \, i \in \overline{1,N}$$

The problem (P.K) has two sets of variables,

centers $\{\mathbf{c}_k\}$, as before, and

probabilities $\{p_k(\mathbf{x}_i)\}$, corresponding to the **assignments**.

The approximate problem (P.K)

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i\, p_k(\mathbf{x}_i)\, \|\mathbf{x}_i - \mathbf{c}_k\| \tag{P.K}$$

$$\text{s.t.} \sum_{k=1}^{K} p_k(\mathbf{x}_i) = 1,\ i \in \overline{1,N},$$

$$p_k(\mathbf{x}_i) \geq 0,\ k \in \overline{1,K},\ i \in \overline{1,N},$$

with variables $\{\mathbf{c}_k\}$ and $\{p_k(\mathbf{x}_i)\}$.

Fix one set of variables, and minimize (P.K) with respect to the second set, then fix the second set, etc. We thus alternate between

(1) the **probabilities problem**, with given centers, and

(2) the **centers problem**, (P.K) with given probabilities.

The approximate problem (P.K)

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, \|\mathbf{x}_i - \mathbf{c}_k\| \tag{P.K}$$

$$\text{s.t.} \sum_{k=1}^{K} p_k(\mathbf{x}_i) = 1, \; i \in \overline{1,N},$$

$$p_k(\mathbf{x}_i) \geq 0, \; k \in \overline{1,K}, \; i \in \overline{1,N},$$

with variables $\{\mathbf{c}_k\}$ and $\{p_k(\mathbf{x}_i)\}$.

Fix one set of variables, and minimize (P.K) with respect to the second set, then fix the second set, etc. We thus alternate between

(1) the **probabilities problem**, with given centers, and

(2) the **centers problem**, (P.K) with given probabilities.

The approximate problem (P.K)

$$\min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, \|\mathbf{x}_i - \mathbf{c}_k\| \tag{P.K}$$

$$\text{s.t.} \sum_{k=1}^{K} p_k(\mathbf{x}_i) = 1, \; i \in \overline{1,N},$$

$$p_k(\mathbf{x}_i) \geq 0, \; k \in \overline{1,K}, \; i \in \overline{1,N},$$

with variables $\{\mathbf{c}_k\}$ and $\{p_k(\mathbf{x}_i)\}$.

Fix one set of variables, and minimize (P.K) with respect to the second set, then fix the second set, etc. We thus alternate between

(1) the **probabilities problem**, with given centers, and

(2) the **centers problem**, (P.K) with given probabilities.

The approximate problem (P.K)

$$\begin{aligned} \min \sum_{k=1}^{K} \sum_{i=1}^{N} w_i\, p_k(\mathbf{x}_i)\, \|\mathbf{x}_i - \mathbf{c}_k\| \qquad \text{(P.}K\text{)} \\ \text{s.t.} \sum_{k=1}^{K} p_k(\mathbf{x}_i) = 1,\ i \in \overline{1,N}, \\ p_k(\mathbf{x}_i) \geq 0,\ k \in \overline{1,K},\ i \in \overline{1,N}, \end{aligned}$$

with variables $\{\mathbf{c}_k\}$ and $\{p_k(\mathbf{x}_i)\}$.

Fix one set of variables, and minimize (P.K) with respect to the second set, then fix the second set, etc. We thus alternate between

(1) the **probabilities problem**, with given centers, and

(2) the **centers problem**, (P.K) with given probabilities.

The probabilities problem

$$p_k(\mathbf{x}_i) = \text{Prob}\{\mathbf{x}_i \in \mathscr{C}_k\},\ i \in \overline{1,N},\ k \in \overline{1,K}.$$

Given

data points $\{\mathbf{x}_i :\ i \in \overline{1,N}\}$,
weights $\{w_i :\ i \in \overline{1,N}\}$,
centers $\{\mathbf{c}_k :\ k \in \overline{1,K}\}$,
distances $\{d_k(\mathbf{x}_i) = d(\mathbf{x}_i, \mathbf{c}_k) :\ i \in \overline{1,N},\ k \in \overline{1,K}\}$.

For $\mathbf{x}_i,\ i \in \overline{1,N}$,

$$p_k(\mathbf{x}_i) = \frac{1/d_k(\mathbf{x}_i)}{\sum\limits_{j=1}^{K} 1/d_j(\mathbf{x}_i)},\ k \in \overline{1,K},$$

independent of the weights w_i.

The probabilities problem

$$p_k(\mathbf{x}_i) = \text{Prob}\{\mathbf{x}_i \in \mathscr{C}_k\},\ i \in \overline{1,N},\ k \in \overline{1,K}.$$

Given

data points $\{\mathbf{x}_i : i \in \overline{1,N}\}$,
weights $\{w_i : i \in \overline{1,N}\}$,
centers $\{\mathbf{c}_k : k \in \overline{1,K}\}$,
distances $\{d_k(\mathbf{x}_i) = d(\mathbf{x}_i, \mathbf{c}_k) : i \in \overline{1,N},\ k \in \overline{1,K}\}$.

For $\mathbf{x}_i,\ i \in \overline{1,N}$,

$$p_k(\mathbf{x}_i) = \frac{1/d_k(\mathbf{x}_i)}{\sum\limits_{j=1}^{K} 1/d_j(\mathbf{x}_i)},\ k \in \overline{1,K},$$

independent of the weights w_i.

The centers problem

$$\{\mathbf{c}_k : \; k \in \overline{1,K}\}$$

Given

data points $\{\mathbf{x}_i : \; i \in \overline{1,N}\}$,
weights $\{w_i : \; i \in \overline{1,N}\}$,
distances $\{d_k(\mathbf{x}_i) = d(\mathbf{x}_i, \mathbf{c}_k) : \; i \in \overline{1,N}, \; k \in \overline{1,K}\}$,
probabilities $\{p_k(\mathbf{x}_i) : \; i \in \overline{1,N}, \; k \in \overline{1,K}\}$

The K centers $\mathbf{c}_k$ are computed separately

$$\mathbf{c}_k = \arg\min_{\mathbf{c}} \left\{ \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d_k(\mathbf{x}_i, \mathbf{c}) \right\}, \; k \in \overline{1,K}. \tag{23}$$

Note: Each $\mathbf{c}_k$ is the center of all N points $\mathbf{x}_i$, with "weights" $w_i \, p_k(\mathbf{x}_i)$.

The centers problem

$$\{\mathbf{c}_k : k \in \overline{1,K}\}$$

Given

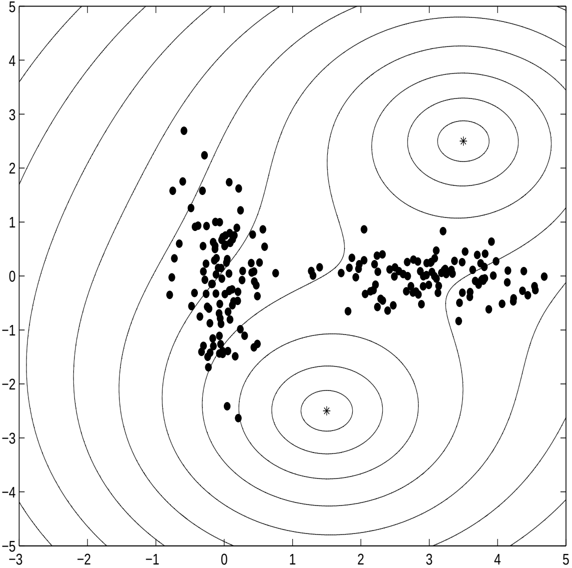
data points $\{\mathbf{x}_i : i \in \overline{1,N}\}$,
weights $\{w_i : i \in \overline{1,N}\}$,
distances $\{d_k(\mathbf{x}_i) = d(\mathbf{x}_i, \mathbf{c}_k) : i \in \overline{1,N}, k \in \overline{1,K}\}$,
probabilities $\{p_k(\mathbf{x}_i) : i \in \overline{1,N}, k \in \overline{1,K}\}$

The K centers $\mathbf{c}_k$ are computed separately

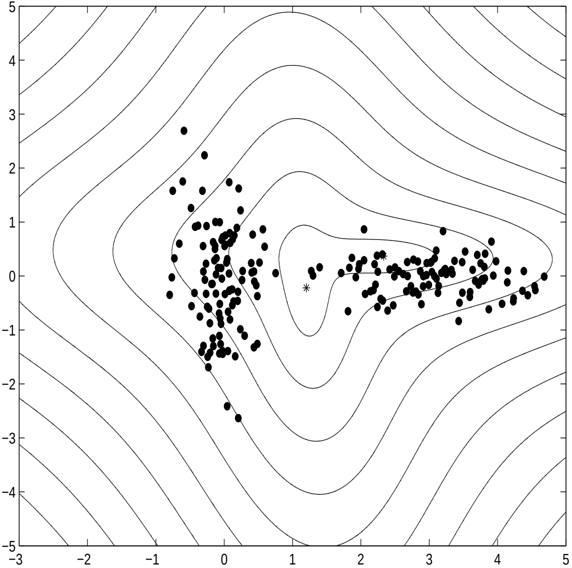
$$\mathbf{c}_k = \arg\min_{\mathbf{c}} \left\{ \sum_{i=1}^{N} w_i \, p_k(\mathbf{x}_i) \, d_k(\mathbf{x}_i, \mathbf{c}) \right\}, \; k \in \overline{1,K}. \tag{23}$$

Note: Each $\mathbf{c}_k$ is the center of all N points $\mathbf{x}_i$, with "weights" $w_i \, p_k(\mathbf{x}_i)$.

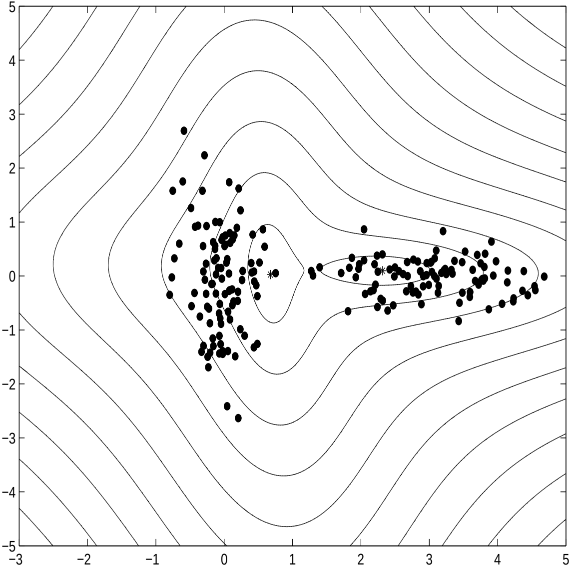
Example: $N = 200$, $K = 2$, iteration 0



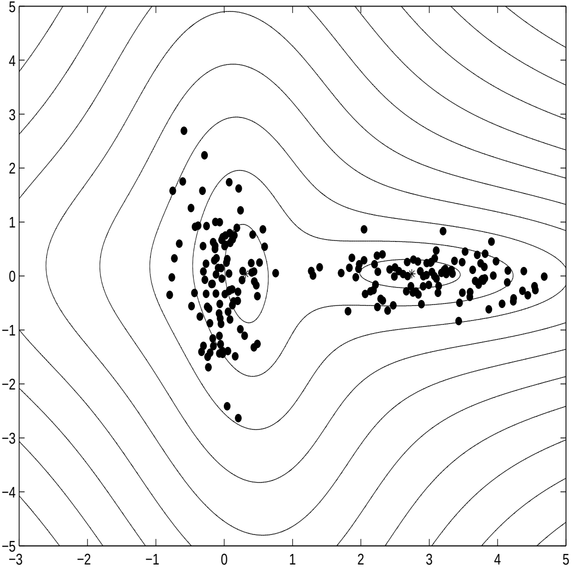
Example: $N = 200$, $K = 2$, iteration 1



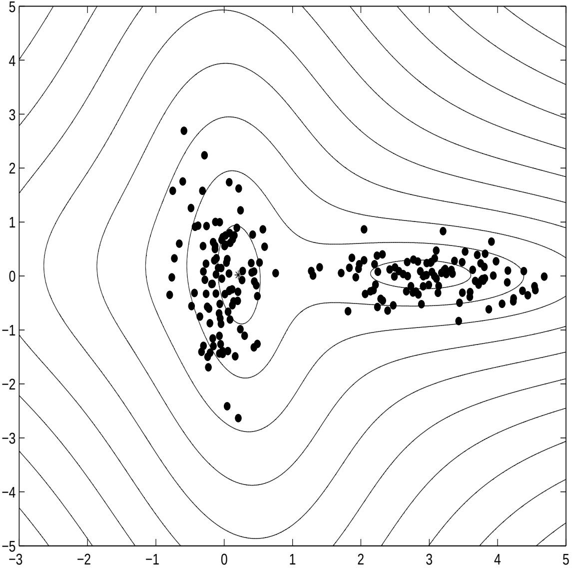
Example: $N = 200$, $K = 2$, iteration 2



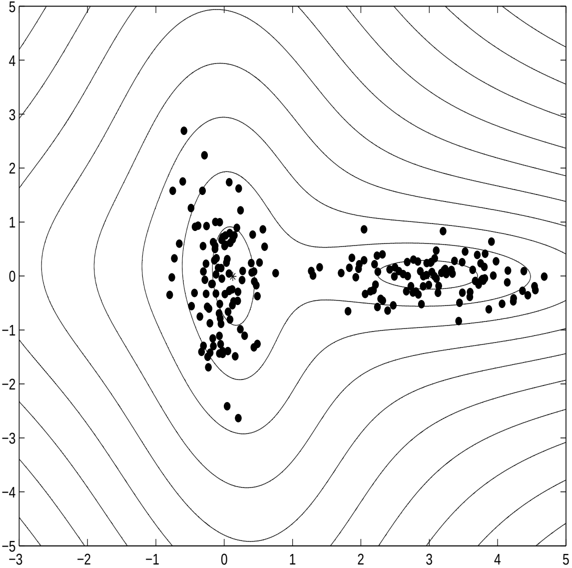
Example: $N = 200$, $K = 2$, iteration 3



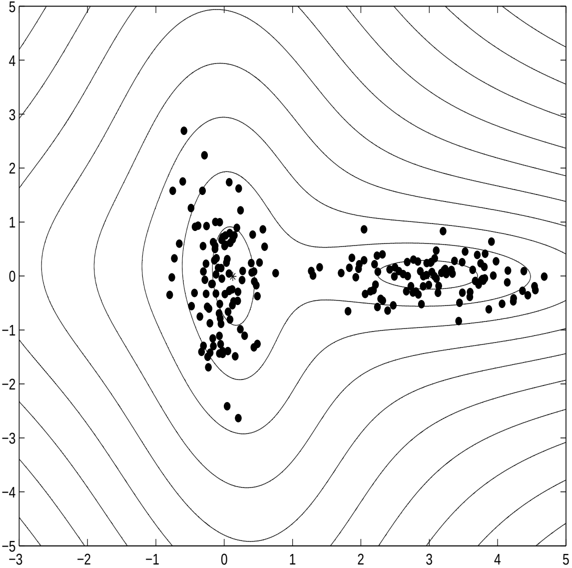
Example: $N = 200$, $K = 2$, iteration 4



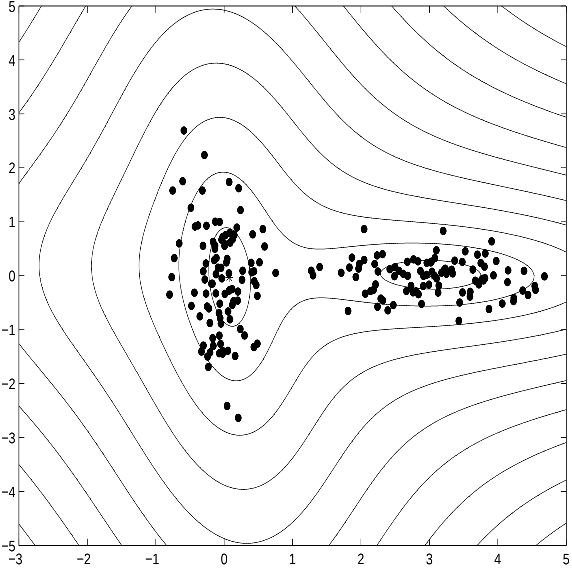
Example: $N = 200$, $K = 2$, iteration 5



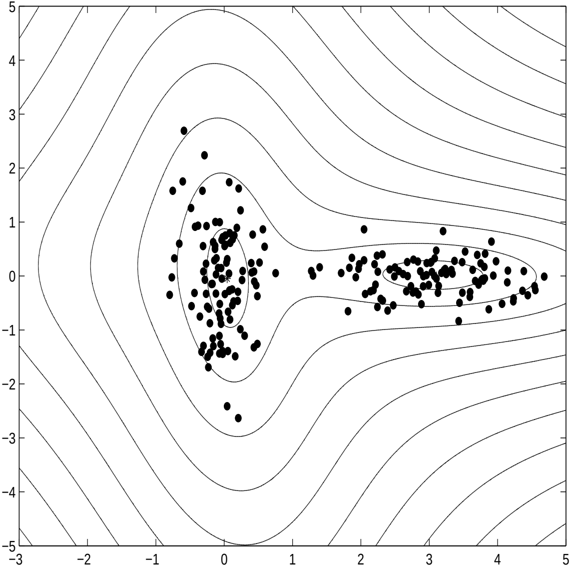
Example: $N = 200$, $K = 2$, iteration 6



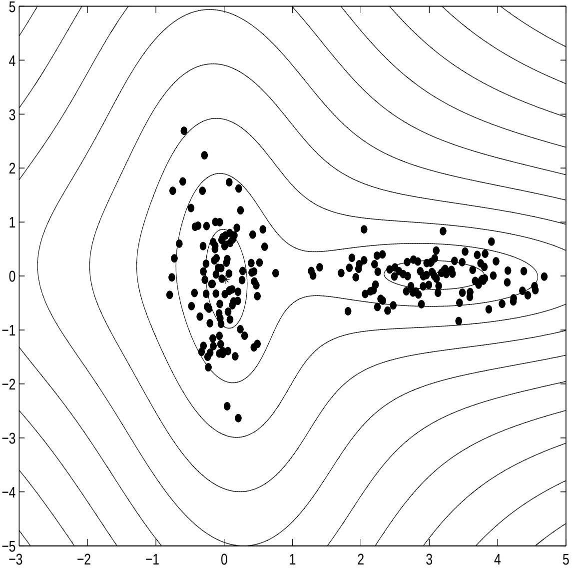
Example: $N = 200$, $K = 2$, iteration 7



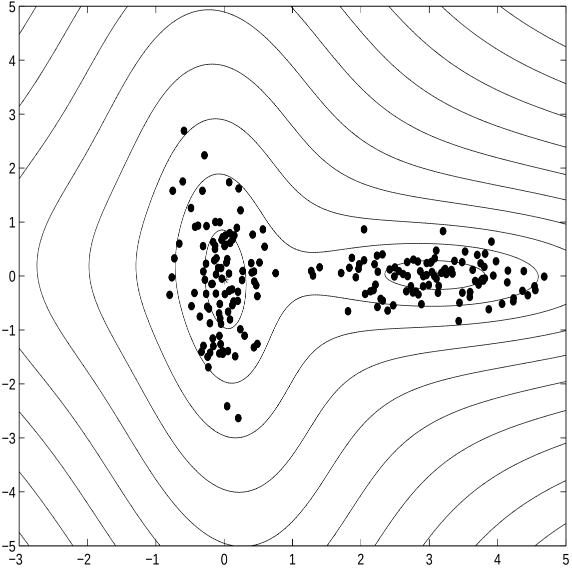
Example: $N = 200$, $K = 2$, iteration 8



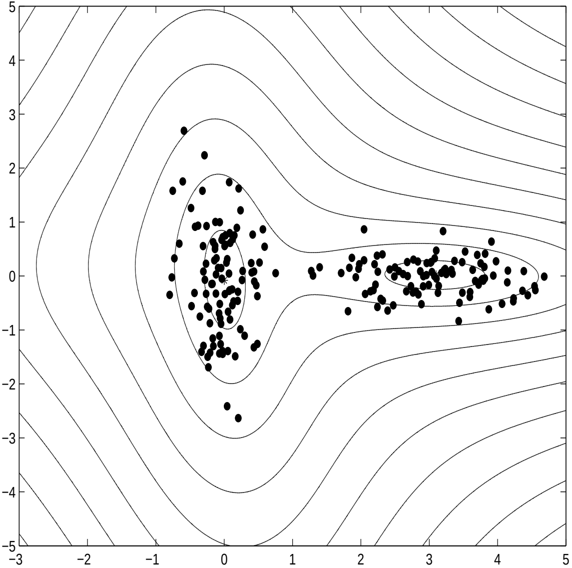
Example: $N = 200$, $K = 2$, iteration 9



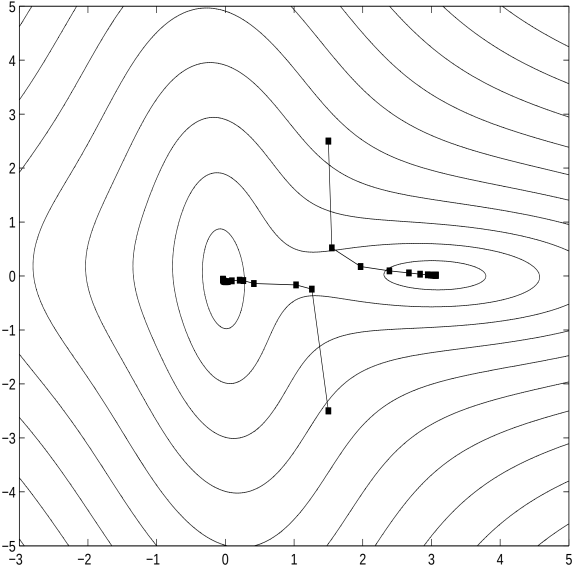
Example: $N = 200$, $K = 2$, iteration 10



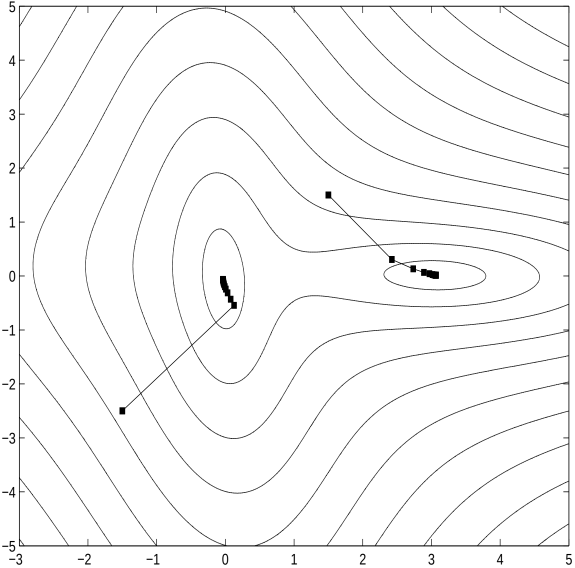
Example: $N = 200$, $K = 2$, iteration 11



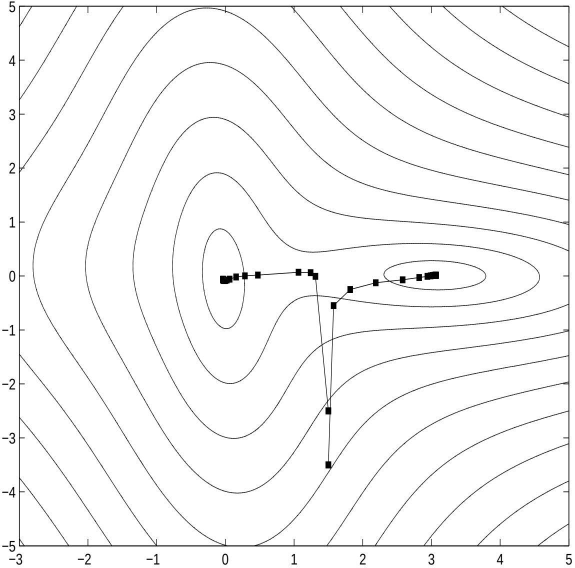
Example: $N = 200$, $K = 2$. Movements of centers



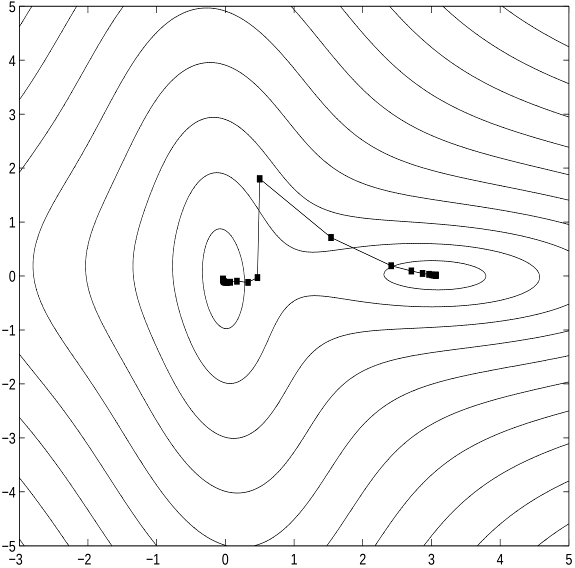
Example: $N = 200$, $K = 2$. Movements of centers



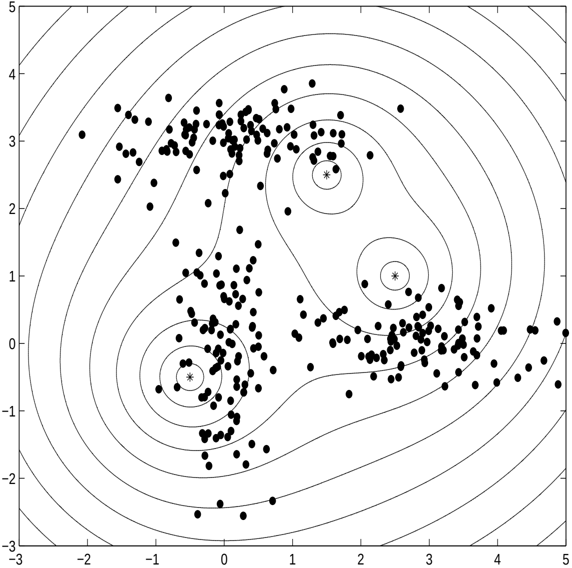
Example: $N = 200$, $K = 2$. Movements of centers



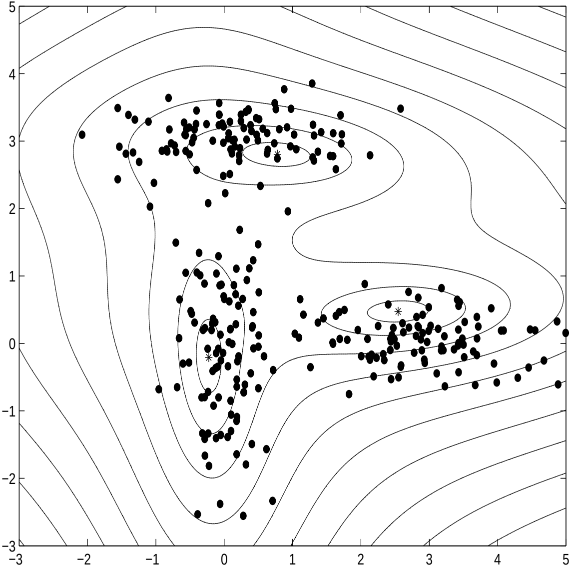
Example: $N = 200$, $K = 2$. Movements of centers



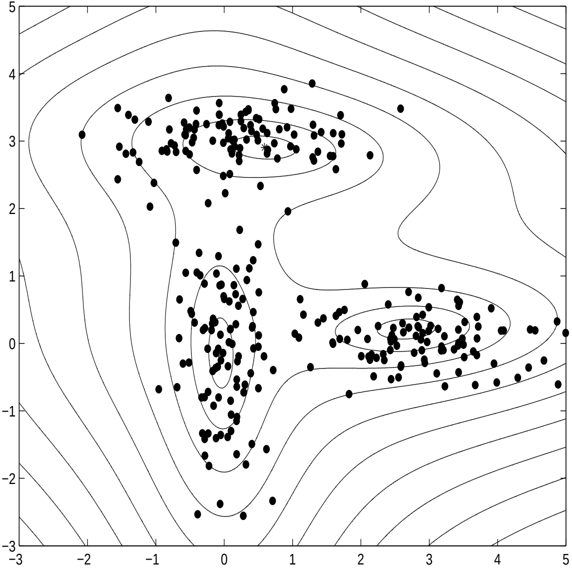
Example: $N = 300$, $K = 3$, iteration 0



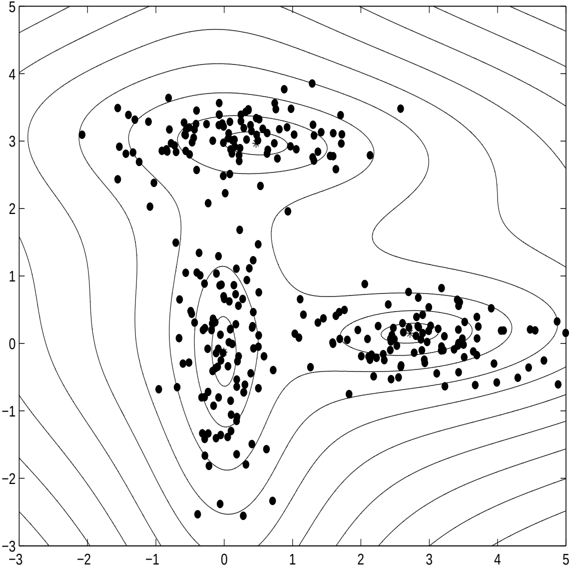
Example: $N = 300$, $K = 3$, iteration 1



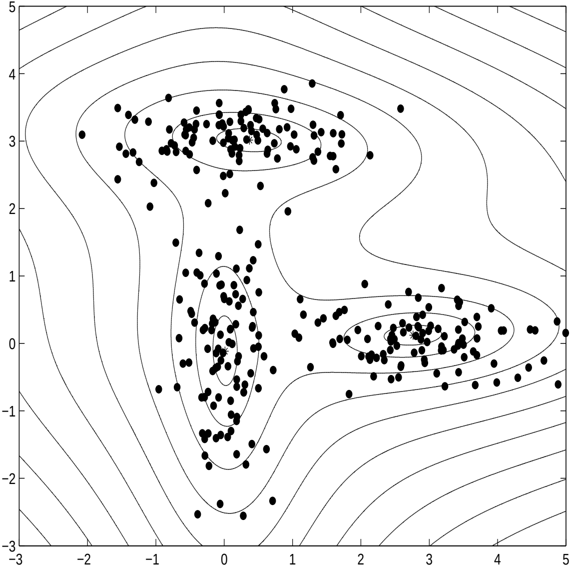
Example: $N = 300$, $K = 3$, iteration 2



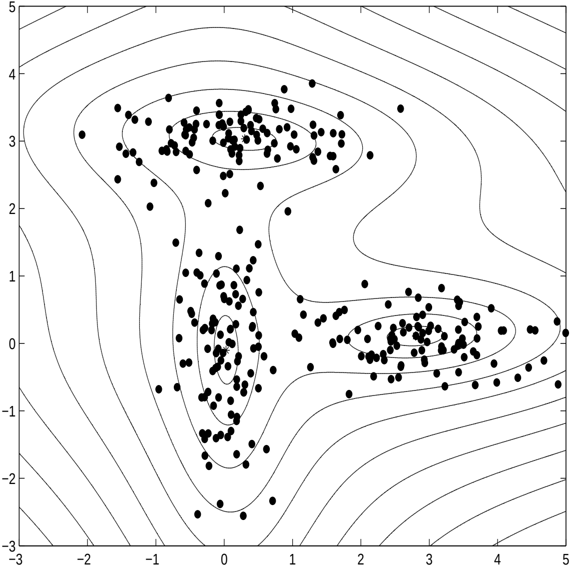
Example: $N = 300$, $K = 3$, iteration 3



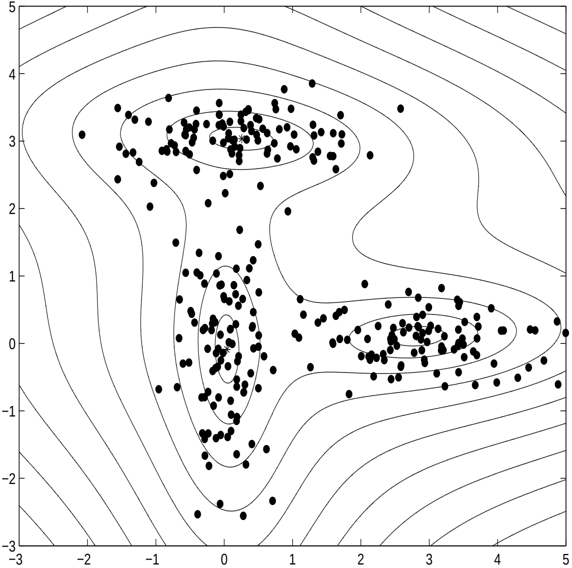
Example: $N = 300$, $K = 3$, iteration 4



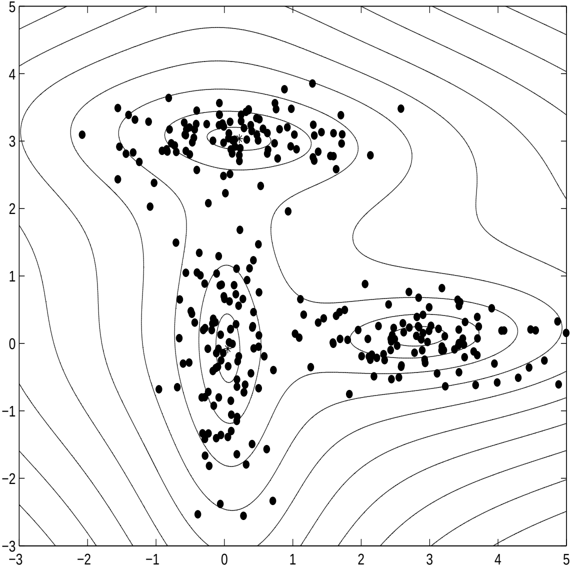
Example: $N = 300$, $K = 3$, iteration 5



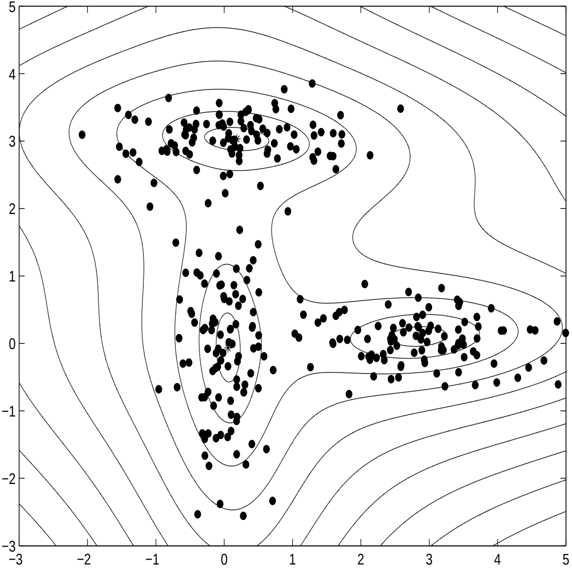
Example: $N = 300$, $K = 3$, iteration 6



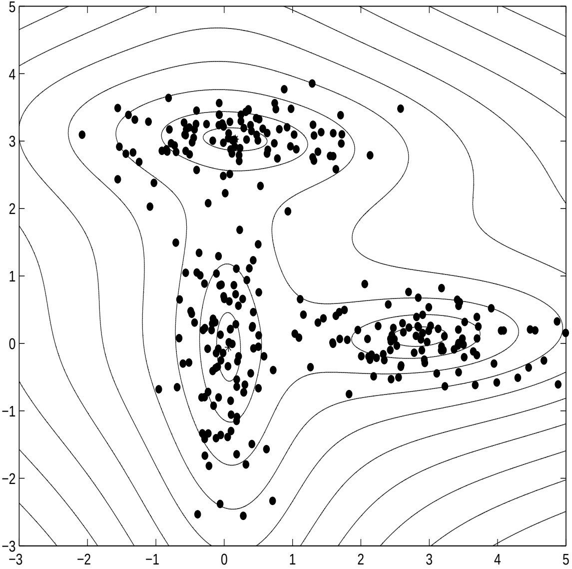
Example: $N = 300$, $K = 3$, iteration 7



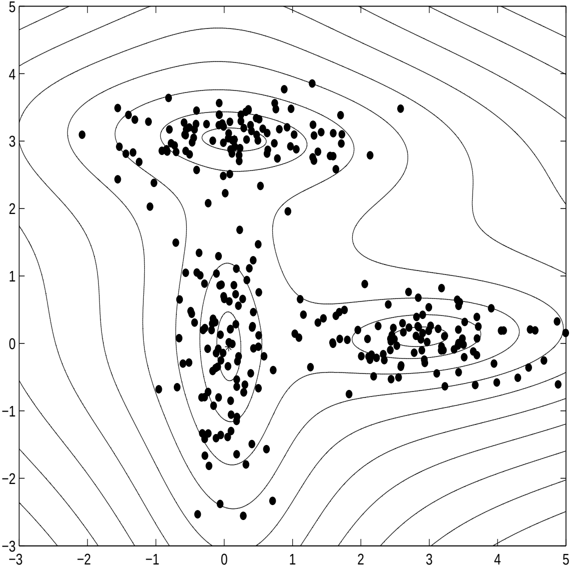
Example: $N = 300$, $K = 3$, iteration 8



Example: $N = 300$, $K = 3$, iteration 9

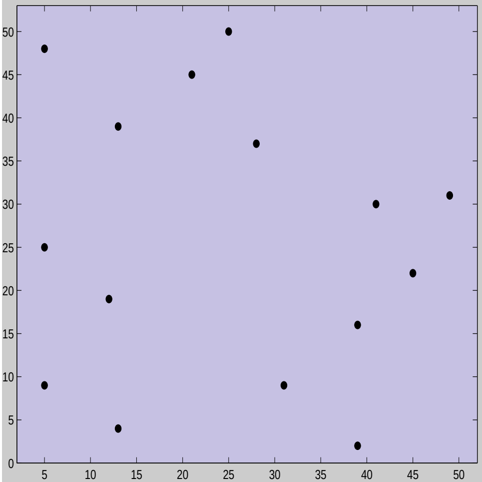


Example: $N = 300$, $K = 3$, iteration 10

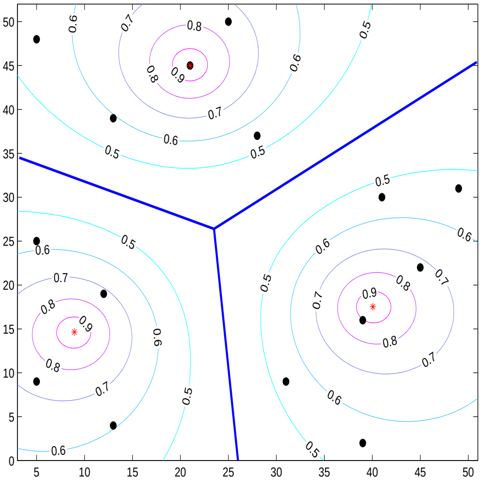


Outline

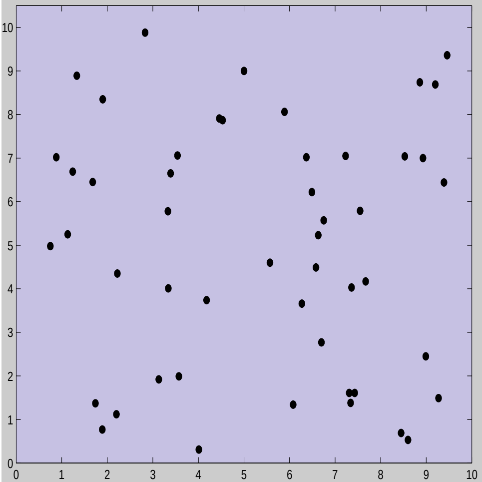
$N = 15, K = 3$ (Cooper, 1964)



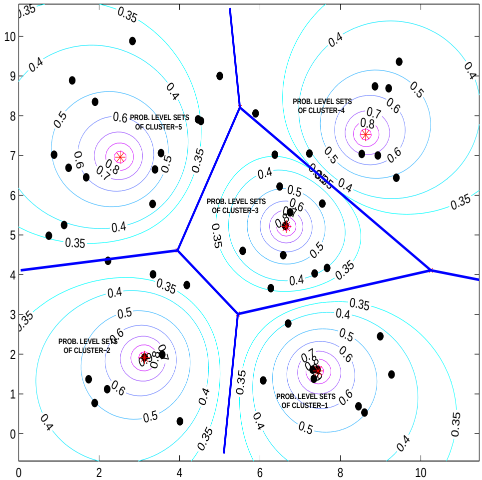
$N = 15, K = 3$ (Cooper, 1964)



$N = 50,\ K = 5$ (Eilon, 1971)



$N = 50,\ K = 5$ (Eilon, 1971)



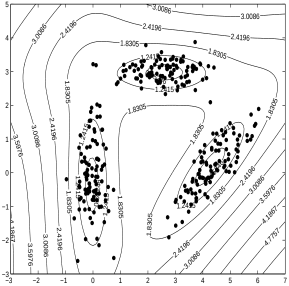
Outline

The JDF of the dataset

The JDF of the dataset $\mathbf{X}$ is the sum of the JDF's of all N data points $\mathbf{x} \in \mathbf{X}$,

$$D(\mathbf{X}) = \sum_{\mathbf{x} \in \mathbf{X}} D(\mathbf{x}) \tag{24}$$

a function of the **cluster centers** $\mathbf{c}_k$, and **distance functions** d_k

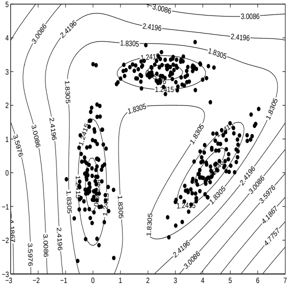


The JDF of the dataset

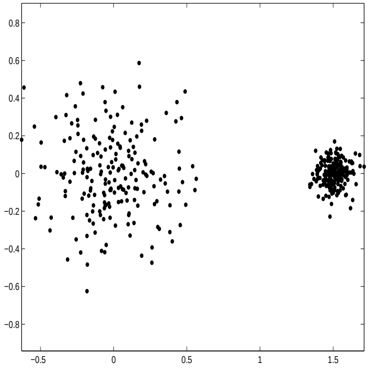
The JDF of the dataset $\mathbf{X}$ is the sum of the JDF's of all N data points $\mathbf{x} \in \mathbf{X}$,

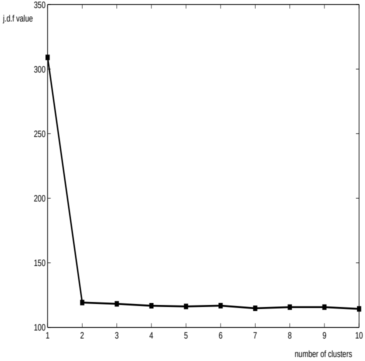
$$D(\mathbf{X}) = \sum_{\mathbf{x} \in \mathbf{X}} D(\mathbf{x}) \tag{24}$$

a function of the **cluster centers** $\mathbf{c}_k$, and **distance functions** d_k

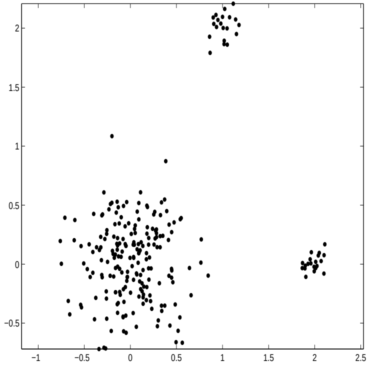


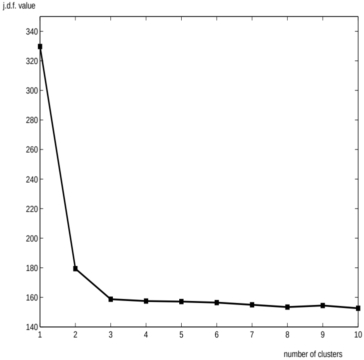
Example: 2 clusters



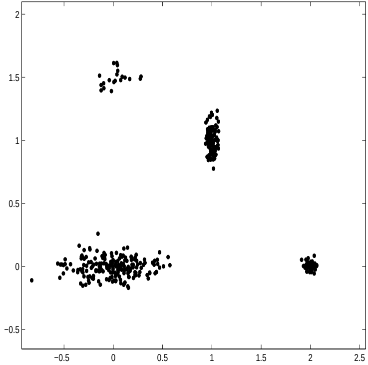


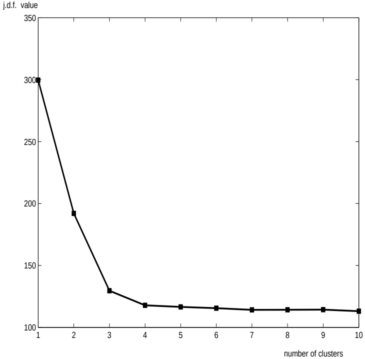
Example: 3 clusters



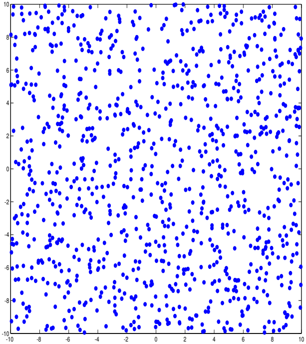


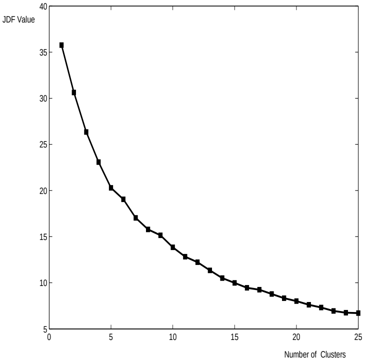
Example: 4 clusters



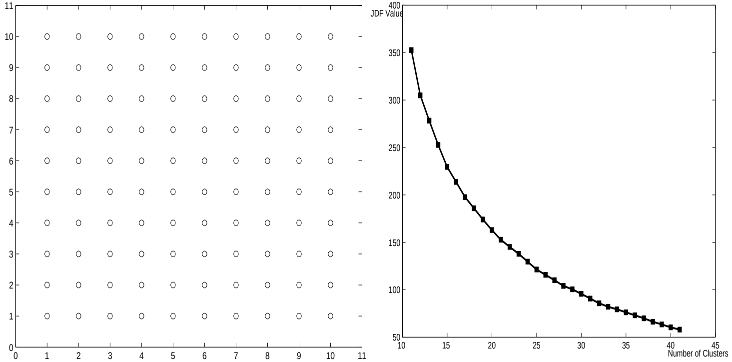


What if there is no structure?





What is the correct number of clusters?



Outline

References

[1] J. Aczél, *Lectures on Functional Equations and their Applications*, Academic Press, 1966

[2] W.N. Anderson Jr. and R.J. Duffin, Series and parallel addition of matrices, *J. Math. Anal. Appl.* **26**(1969), 576–594.

[3] M. Arav, Contour approximation of data, *J. of Math. Inequalities* **2**(2008), 161–167

[4] A. Ben–Israel, Probabililistic clustering papers [**Link**]

[5] A. Ben–Isreal and C. Iyigun, Probabilistic distance clustering, *J. Classification* **25**(2008), 5–26

[6] L. Boitani and T.K. Fuller, *Research Techniques in Animal Ecology: Controversies and Consequences*, Columbia University Press, 2000, xxxii + 442 pp.

[7] K.R. Dixon and J.A. Chapman, Harmonic mean measure of animal activity areas, *Ecology* **61**(1980), 1040–1044

[8] R.J. Duffin, *Network Models*, Mathematical Aspects of Electrical Network Analysis, SIAM–AMS Proceedings, Vol III, 65–91, American Mathematical Society, 1971.

[9] N.R. Sree Harsha, A review of the variational methods for solving DC circuits, *European J. Physics* **40**(2019)

[10] C. Iyigun and A. Ben–Israel, Contour approximation of data: A duality theory, *Lin. Algeb. and Appl.* **430**(2009), 2771–2780

[11] C. Iyigun and A. Ben–Israel, A generalized Weiszfeld method for the multi–facility location problem, *O.R. Letters* **38**(2010), 207–214

[12] R.D. Luce, *Individual Choice Behavior*, Wiley, New York, 1959

[13] N. Megiddo and K.J. Supowit, On the complexity of some common geometric location problems, *SIAM Journal on Computing* **13**(1984), 182–196

[14] T.D. Morley, Parallel summation, Maxwell's principle and the infimum of projections, *J. Math. Anal. Appl.* **70**(1979), 33–41

[15] D. Shepard, A two-dimensional interpolation function for irregularly-spaced data, *Proceedings of the 1968 ACM National Conference*, pp. 517–524

[16] M. Teboulle, A unified continuous optimization framework for center-based clustering methods, *J. of Machine Learning* **8**(2007), 65–102

[17] E. Weiszfeld, Sur le point par lequel la somme des distances de n points donnés est minimum, *Tohoku Math. J.* **43**(1937) 355-386