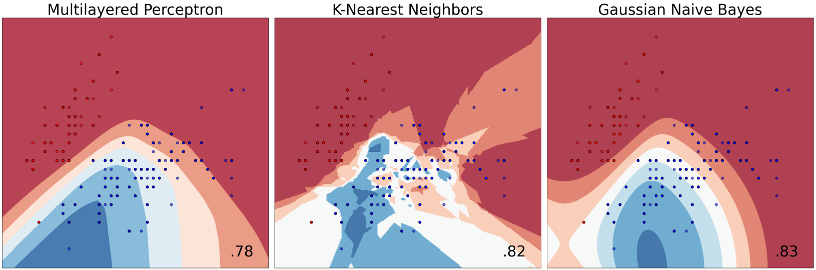
Fundamentals of Machine Learning

Billy Braithwaite

IT Center for Science Ltd.

October 28, 2022

About the course



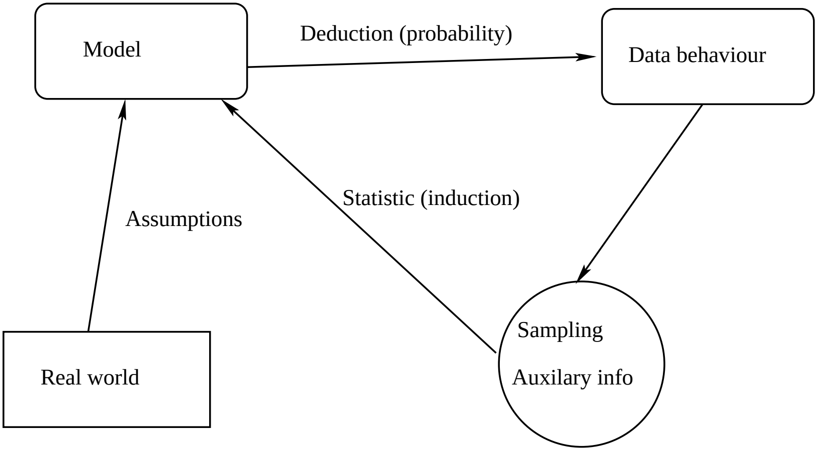
The core message of the course

Georges Matheron

”Illegitimate use of scientific concepts beyond the limits within which they have an operative meaning is nothing else but a surreptitious passage into metaphysics”

Course agenda

Recap of Day 1: Difficulties of interpretation



Recap of Day 1: Statistical inference as an optimization problem

$$\begin{bmatrix} x_{0,0} & x_{0,1} & x_{0,2} & \cdots & x_{0,n-1} \\ x_{0,0} & x_{0,1} & x_{0,2} & \cdots & x_{0,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m-1,0} & x_{m-1,1} & x_{m-1,2} & \cdots & x_{m-1,n-1} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \\ w_{n-1} \end{bmatrix}^T = \begin{bmatrix} y_0 \\ y_1 \\ \\ y_{m-1} \end{bmatrix}$$

$$X^{m \times n} = \begin{cases} m > n & \text{(overdetermined)} \\ n \gg m & \text{(underdetermined)} \end{cases}$$

$$\hat{x} \leftarrow \operatorname*{arg\,min}_{\vec{x} \in X \subset \mathbb{F}^n} f(\vec{x}), \ \text{s.t.} \ A\vec{x} = \vec{b}, \ A \in \mathbb{F}^{n \times n}, \vec{b} \in \mathbb{F}^n$$

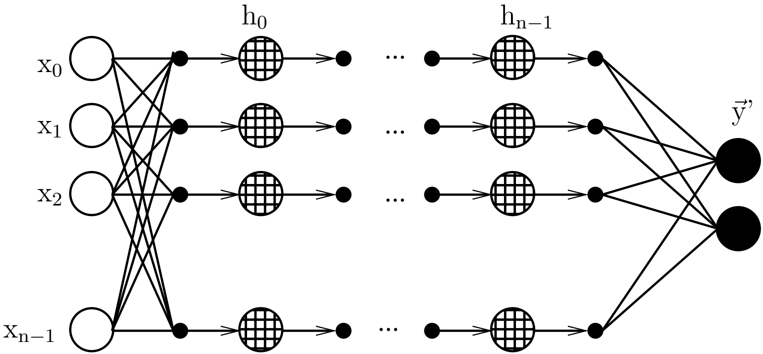
Recap of Day 1: Statistical inference from empirical data

Given a traning set $\mathrm{X_D}$, evaluate

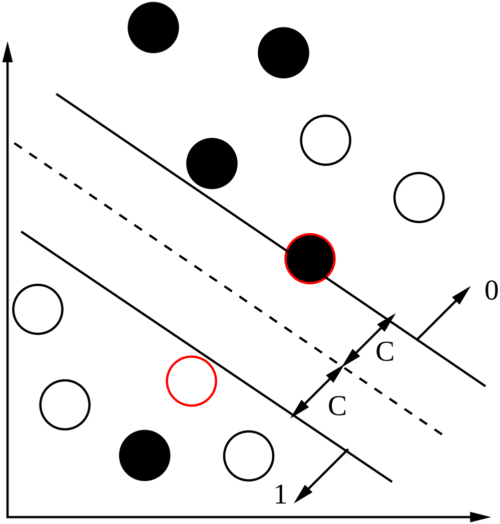
$$\int \mathcal{L}(\mathrm{f}(\vec{\mathrm{X}}_\mathrm{D}, \alpha^*), \omega)\ \mathrm{dF}(\vec{\mathrm{X}}_\mathrm{D}),\ \ \alpha^* \in \Lambda$$

$$\frac{1}{\text{\#training samples}} \sum_{\mathrm{i}=0}^{\text{\#training samples - 1}} \mathcal{L}(\mathrm{f_i}(\vec{\mathrm{X}}_\mathrm{D}, \alpha^*), \omega_\mathrm{i}),\ \ \alpha^* \in \Lambda$$

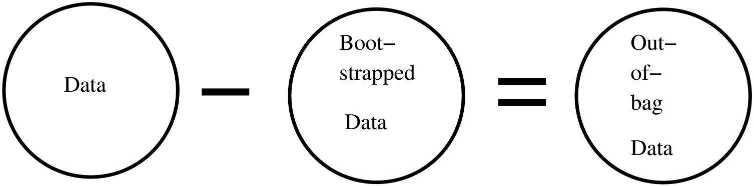
Recap of Day 1: Artificial Neural Networks



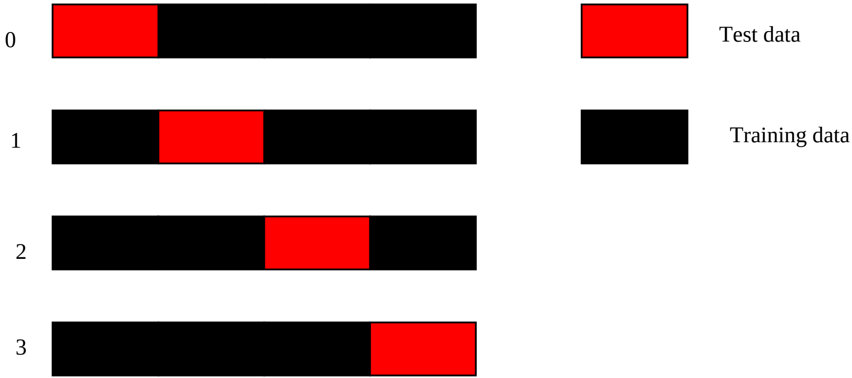
Recap of Day 1: Max margins



Recap of Day 1: Bagging



Recap of Day 1: Cross-validation



Unsupervised Learning

Estimation:

$$\prod_{i=0}^{n-1} f(x_i \mid \theta) \stackrel{\text{def}}{=} L(\vec{x} \mid \theta), \ \vec{x} \in \mathbb{F}^n$$

Unsupervised Learning

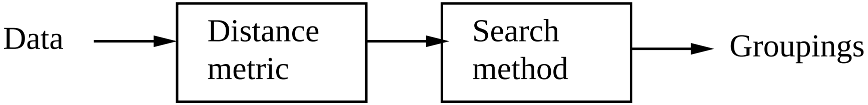
Estimation:

$$\prod_{i=0}^{n-1} f(x_i \mid \theta) \stackrel{\text{def}}{=} L(\vec{x} \mid \theta), \; \vec{x} \in \mathbb{F}^n$$

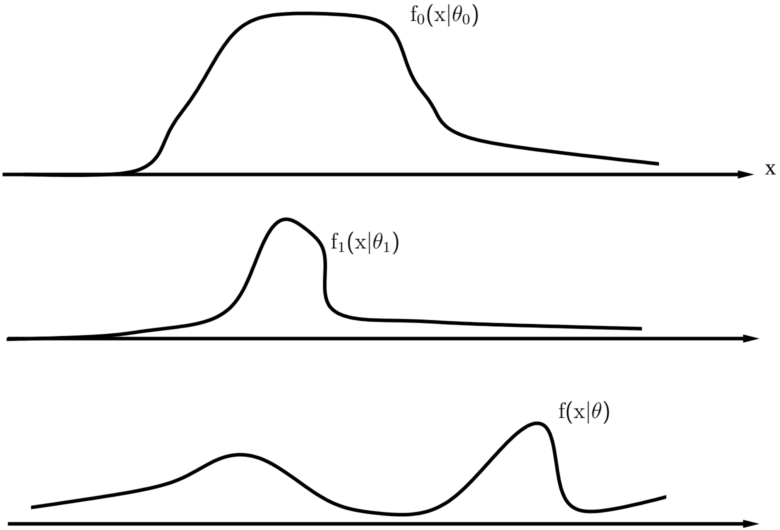
Classification:

$$f_k(x_i \mid \theta_k), \; k = 0, \ldots, C$$

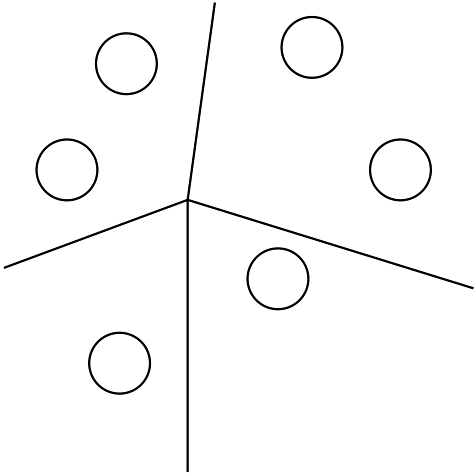
Unsupervised Learning



Mixture of densities

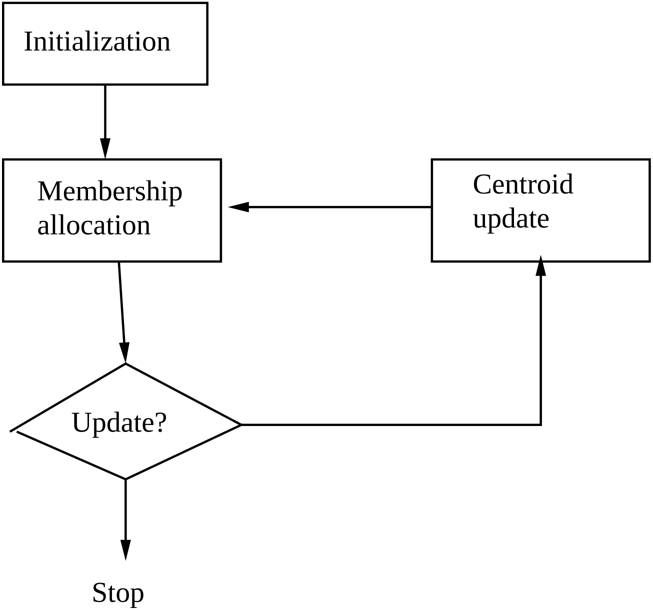


Hard clustering

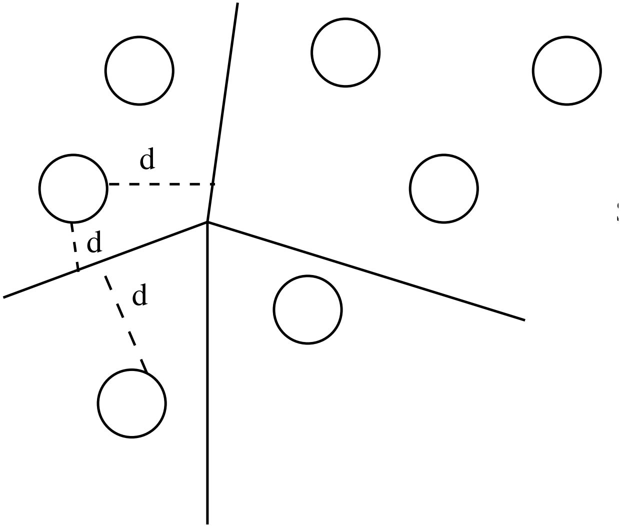


$J() = ||x\text{-}\mu||$

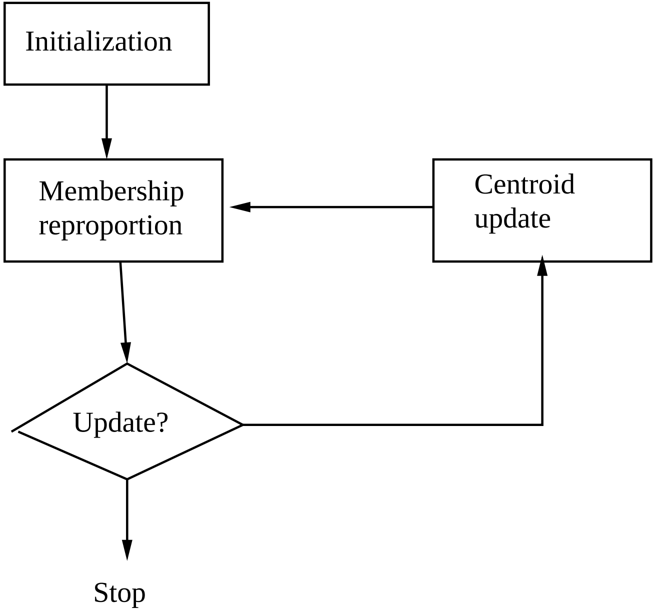
K-means steps



Cluster "quality" estimation



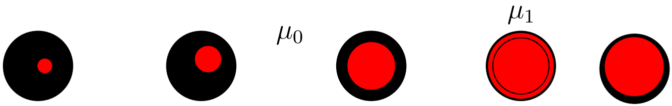
Expectation-maximization steps



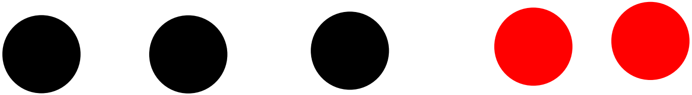
Expectation-Maximization

Expectation-Maximization

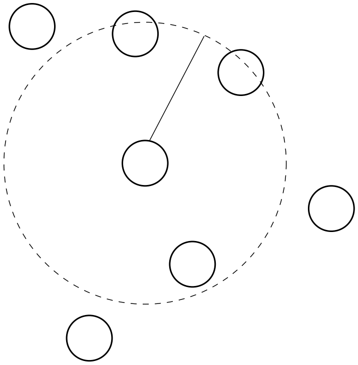
Expectation-Maximization

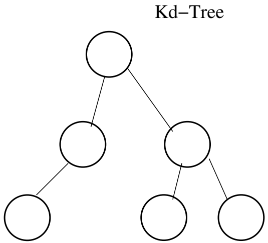


Expectation-Maximization

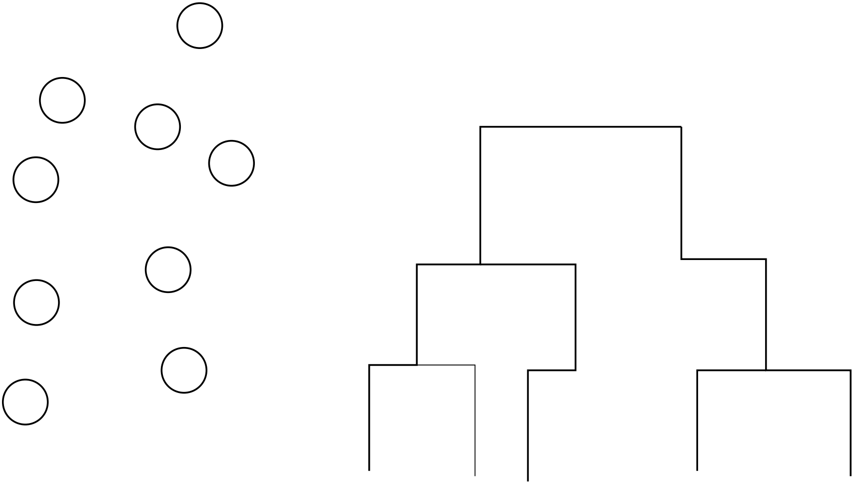


K-Nearest Neighbors

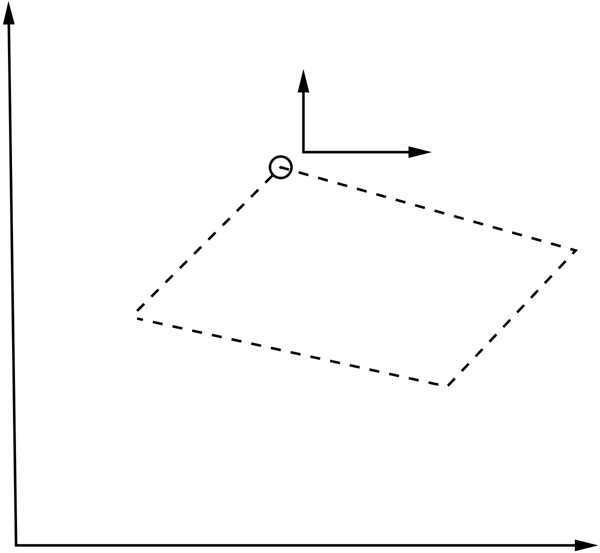




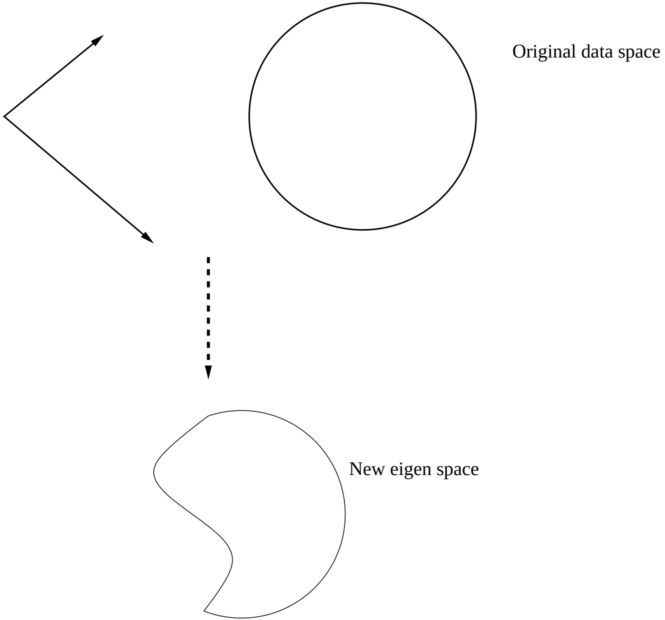
Soft clustering



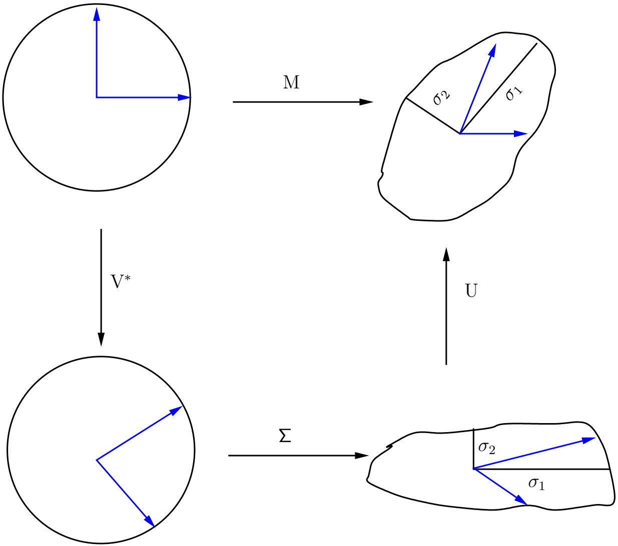
Eigenvalues and subspaces



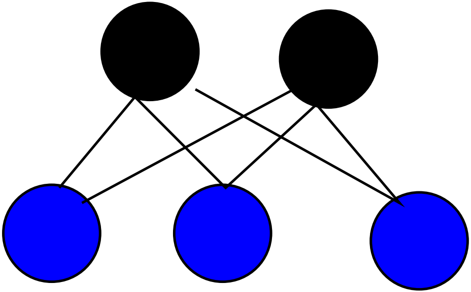
Eigenvalues and subspaces



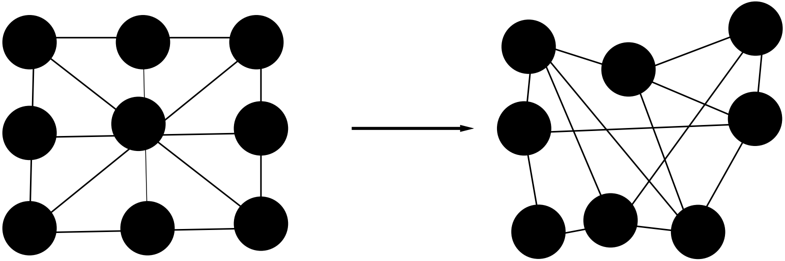
Eigenvalues and subspaces



Boltzmann Machines



Self-organizing Maps



Combining supervised & unsupervised learning

Further reading: Books

Vladimir Vapnik, Statistical learning theory, Adaptive and learning systems for signal processing communications and control, 1998.

Vic Barnett, Comparative statistical inference (1999) John Wiley & Sons

John Maynard Keynes, A treatise on probability (1962) Harper & Row Publishing

Bernardo, José M and Smith, Adrian FM, Bayesian theory (2009) John Wiley & Sons

Harald Cramér, Mathematical methods of statistics, Princeton Univ (1999) Press, Princeton, NJ

Simon Haykin, Neural networks and learning machines 3rd edition (2009) Pearson Education

Further reading: Peer-reviewed articles

Nagy, George, State of the art in pattern recognition, Proceedings of the IEEE, 1968, 56.5: 836-863.

David Hand, Deconstructing statistical questions, Journal of the Royal Statistical Society: Series A (Statistics in Society), 1994, 157.3: 317-338.

David Hand, Statistics and the theory of measurements, Journal of the Royal Statistical Society: Series A (Statistics in Society), 1996, 159.3: 445-473.