Fundamentals of AI

Introduction and the most basic concepts

Notion of mean point in the data

Why bother about mean point?

- **Defining mean point** can be considered as a simple application of unsupervised learning approach
- Calculating mean point is the extreme case of **dimensionality reduction**: $R^N \rightarrow R^0$
- In **complex data spaces** the definition of mean point is non-trivial task
- Definition of mean depends on the **metrics of data space**
- General definition of mean leads to **important generalizations**

Notion of average (mean) point

Arithmetic mean

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n} \quad *$$

Geometric mean

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} = \sqrt[n]{a_1 a_2 \cdots a_n}. \quad **$$

Harmonic mean

$$\bar{x} = n\left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1}$$

* a_i can be vectors!

** arithmetic mean of logarithms

Notion of average (mean) point

- In probability theory : ‘expected’ or ‘central’ value of the probability distribution
  - The analytical formula depends on the type of probability distribution!
  - Can be non-existent

$$\mathbf{E}[X] = \int_{\mathbb{R}} x f(x)\, dx$$

- In geometrical approach: point *m* minimizing the mean squared distance from all data points to *m*
  - this definition belongs to Maurice Fréchet (1878-1973)
  - depends on the metric structure of the feature space
  - can be non-unique

$$m = \underset{p \in M}{\arg\min} \sum_{i=1}^{N} d^2\left(p, x_i\right)$$

Notion of average (mean) point

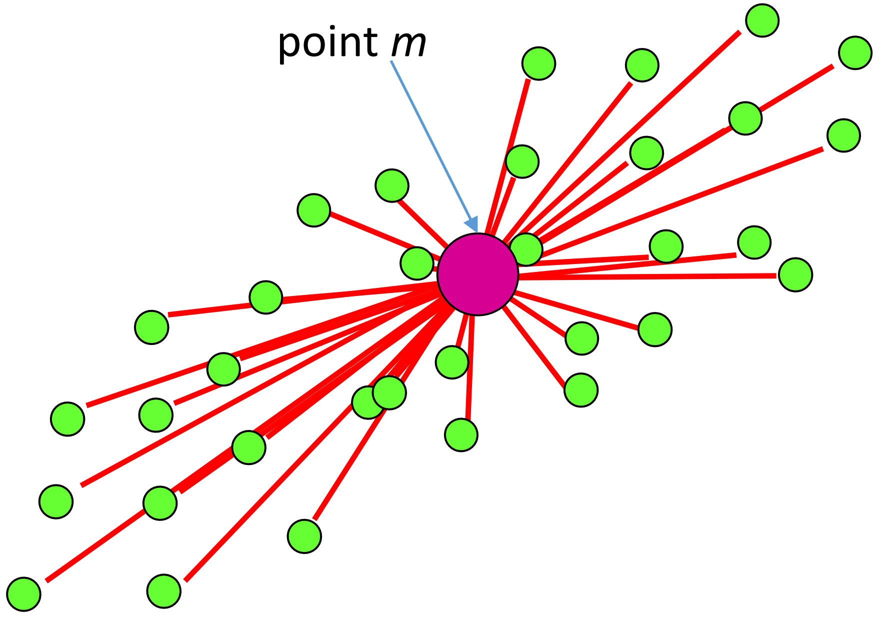
- In probability theory : ‘expected’ or ‘central’ value of the probability distribution, **first moment of the distribution**

Expected Values of Various Statistical Distributions

Distribution	Mathematical Distribution	Mean E(X)
Bernoulli	$X \sim b(1, p)$	p
Binomial	$X \sim B(n, p)$	np
Poisson	$X \sim Po(\lambda)$	λ
Geometric	$X \sim Geometric(p)$	$1/p$
Uniform	$X \sim U(a, b)$	$(a+b)/2$
Exponential	$X \sim \exp(\lambda)$	$1/\lambda$
Normal	$X \sim N(\mu, \sigma^2)$	μ
Standard Normal	$X \sim N(0, 1)$	0
Pareto	$X \sim Par(\alpha)$	$\alpha/(\alpha+1)$ if $\alpha > 1$
Cauchy	$X \sim Cauchy(x_0, \gamma)$	undefined

Notion of average (mean) point

- In geometrical approach: point *m* minimizing the mean squared distance from all data points to *m*, *'center of mass'*



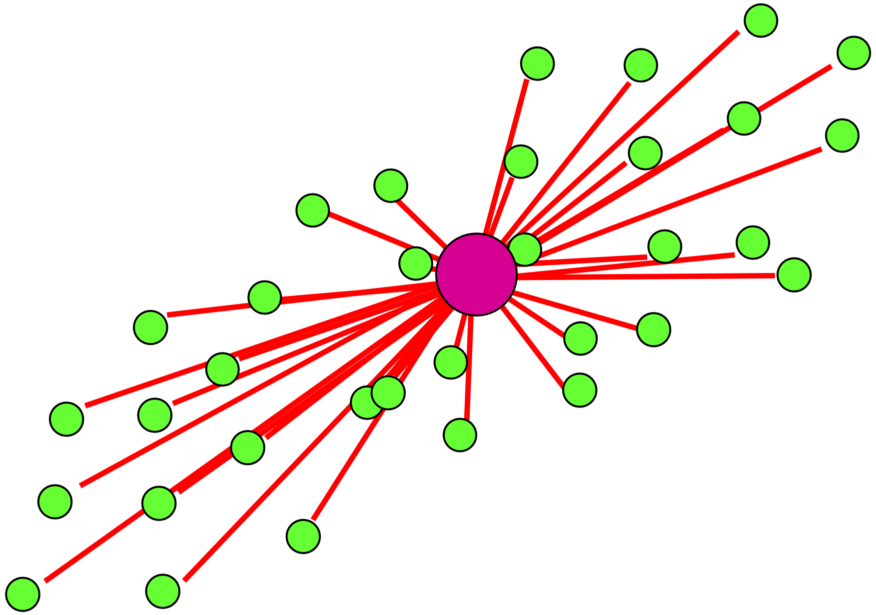
$$\sum_{i=1}^{m} \| \quad \|^2 \rightarrow \min$$

Simple exercise: what is the mean point in Euclidean space?

$$\sum_i \sum_k (x_i^k - m^k)^2 \to \min$$

i – point number

k – coordinate number



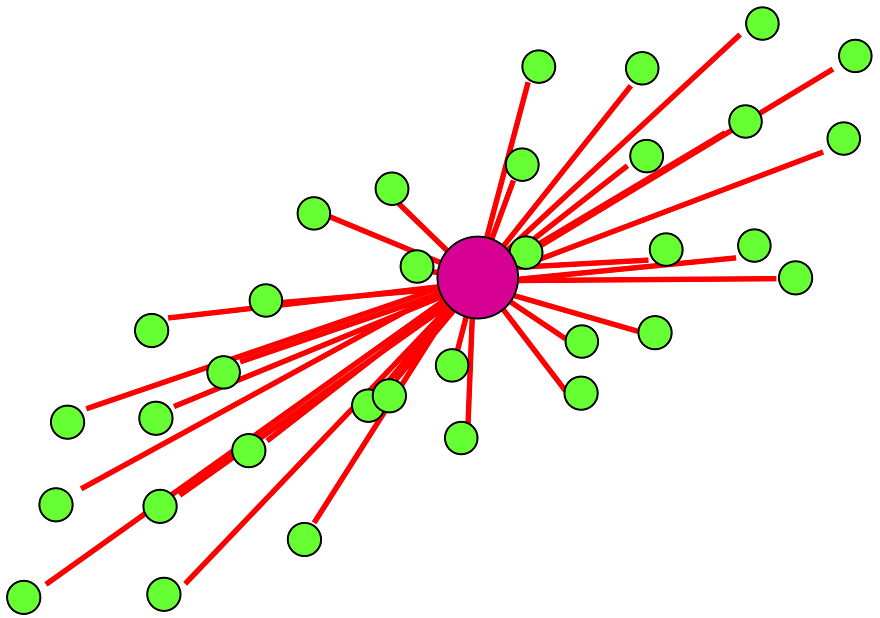
Simple exercise: what is the mean point in Euclidean space?

$$\sum_i \sum_k (x_i^k - m^k)^2 \to \min$$

i – point number
k – coordinate number

$$\left[\sum_i \sum_k (x_i^k - m^k)^2\right]'_{m_k} = 0$$

$$2\sum_i (x_i^k - m^k) = 0$$



Simple exercise: what is the mean point in Euclidean space?

$$\sum_i \sum_k (x_i^k - m^k)^2 \to \min$$

i – point number

k – coordinate number

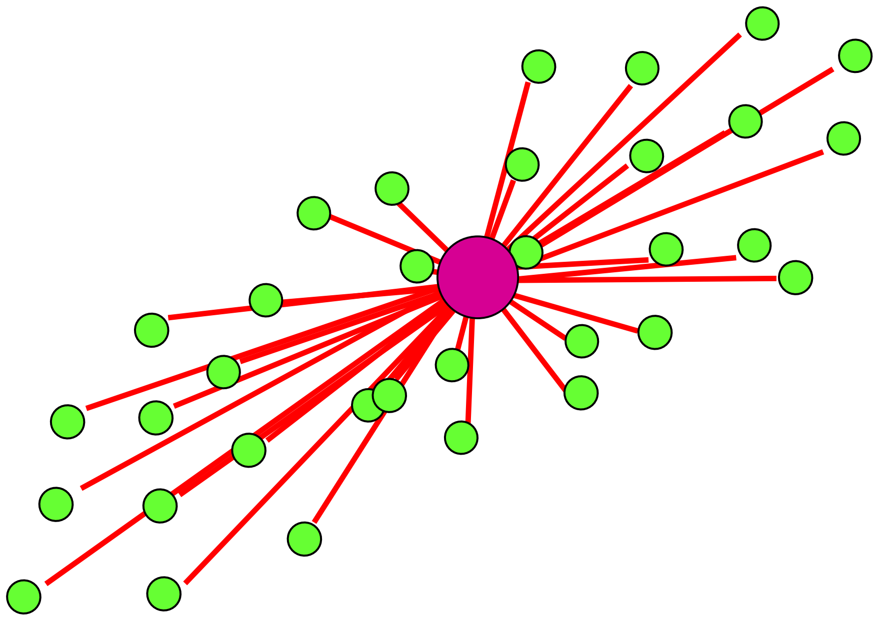
$$\left[\sum_i \sum_k (x_i^k - m^k)^2\right]'_{m_k} = 0$$

$$2\sum_i (x_i^k - m^k) = 0$$

$$\sum_i x_i^k - \sum_i m^k = 0$$

$$\sum_i x_i^k - N m^k = 0$$

$$m_k = \frac{\sum_i x_i^k}{N}$$



Simple exercise: what is the mean point in Euclidean space?

$$\sum_i \sum_k (x_i^k - m^k)^2 \to \min$$

i – point number
k – coordinate number

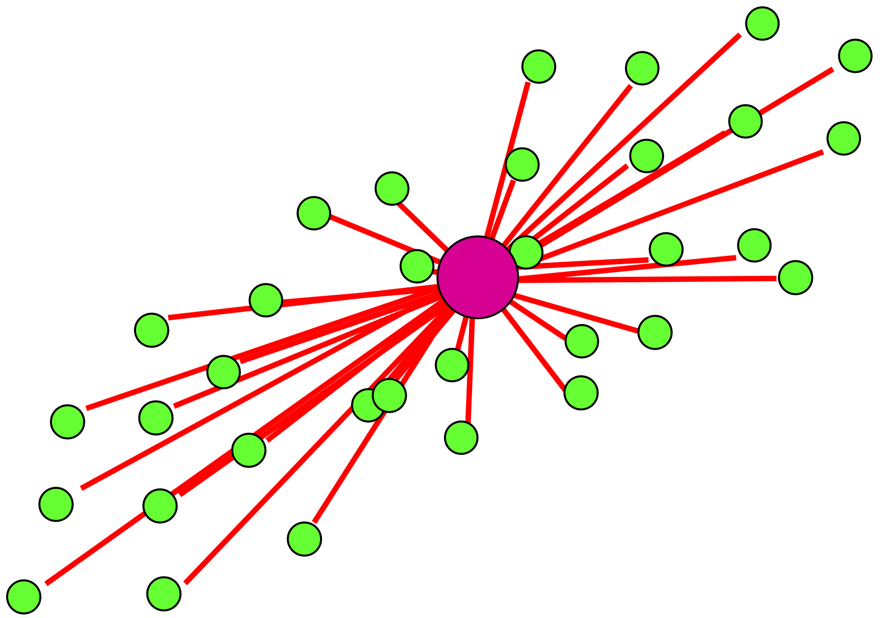
$$\left[\sum_i \sum_k (x_i^k - m^k)^2\right]'_{m_k} = 0$$

$$2\sum_i (x_i^k - m^k) = 0$$

$$\sum_i x_i^k - \sum_i m^k = 0$$

$$\sum_i x_i^k - N m^k = 0$$

$$m_k = \frac{\sum_i x_i^k}{N}$$



Arithmetic mean! ☺

What is the mean point in L1 space?

$$\sum_i \sum_k \| x_i^k - m^k \|_{L1} \rightarrow \min$$

$$\sum_i \sum_k | x_i^k - m^k | \rightarrow \min$$

What is the mean point in L1 space?

$$\sum_i \sum_k \| x_i^k - m^k \|_{L1} \to \min$$

$$\sum_i \sum_k | x_i^k - m^k | \to \min$$

$$\left(\sum_i \sum_k | x_i^k - m^k | \right)'_{m^k} = 0$$

What is the mean point in L1 space?

$$\sum_i \sum_k \| x_i^k - m^k \|_{L1} \to \min$$

$$\sum_i \sum_k | x_i^k - m^k | \to \min$$

$$\left(\sum_i \sum_k | x_i^k - m^k | \right)'_{m^k} = 0$$

$$\left(\sum_{\substack{i \\ x_i^k > m^k}} \sum_k (x_i^k - m^k) - \sum_{\substack{i \\ x_i^k < m^k}} \sum_k (x_i^k - m^k) \right)'_{m_k} = 0$$

$$-\sum_{i, x_i^k > m^k} 1 + \sum_{i, x_i^k < m^k} 1 = 0$$

What is the mean point in L1 space?

$$\sum_i \sum_k \|x_i^k - m^k\|_{L1} \to \min$$

$$\sum_i \sum_k |x_i^k - m^k| \to \min$$

$$\Big(\sum_i \sum_k |x_i^k - m^k|\Big)'_{m^k} = 0$$

$$\Big(\sum_{\substack{i \\ x_i^k > m^k}} \sum_k (x_i^k - m^k) - \sum_{\substack{i \\ x_i^k < m^k}} \sum_k (x_i^k - m^k)\Big)'_{m_k} = 0$$

$$-\sum_{i,x_i^k > m^k} 1 + \sum_{i,x_i^k < m^k} 1 = 0$$

$$\#(x_i^k < m_k) = \#(x_i^k > m_k)$$

What is the mean point in L1 space?

$$\sum_i \sum_k \| x_i^k - m^k \|_{L1} \to \min$$

$$\sum_i \sum_k | x_i^k - m^k | \to \min$$

$$\left(\sum_i \sum_k | x_i^k - m^k | \right)'_{m^k} = 0$$

$$\left(\sum_{\substack{i \\ x_i^k > m^k}} \sum_k (x_i^k - m^k) - \sum_{\substack{i \\ x_i^k < m^k}} \sum_k (x_i^k - m^k) \right)'_{m_k} = 0$$

$$-\sum_{i, x_i^k > m^k} 1 + \sum_{i, x_i^k < m^k} 1 = 0$$

$$\#(x_i^k < m_k) = \#(x_i^k > m_k)$$

This is definition of median value!

Mean value in L1 space - ***medoid***

What is the mean point in L1 space?

$$\sum_i \sum_k \|x_i^k - m^k\|_{L1} \to \min$$

$$\sum_i \sum_k |x_i^k - m^k| \to \min$$

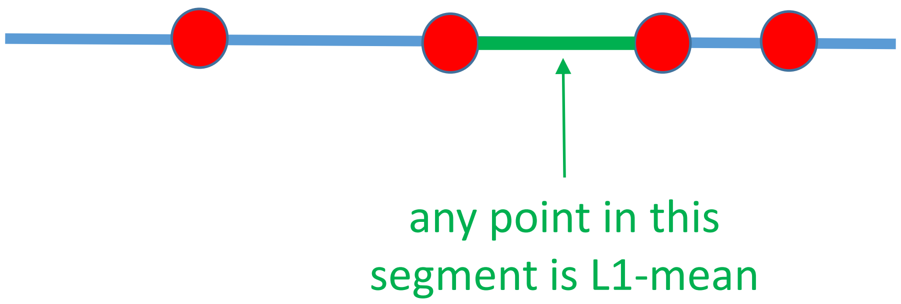
$$\left(\sum_i \sum_k |x_i^k - m^k|\right)'_{m^k} = 0$$

$$\left(\sum_{\substack{i \\ x_i^k > m^k}} \sum_k (x_i^k - m^k) - \sum_{\substack{i \\ x_i^k < m^k}} \sum_k (x_i^k - m^k)\right)'_{m_k} = 0$$

$$-\sum_{i, x_i^k > m^k} 1 + \sum_{i, x_i^k < m^k} 1 = 0$$

$$\#(x_i^k < m_k) = \#(x_i^k > m_k)$$

For even number of data points, there is infinite number of L1-means

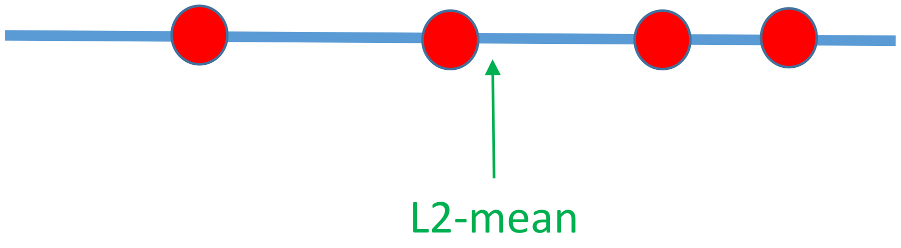


This is definition of median value!

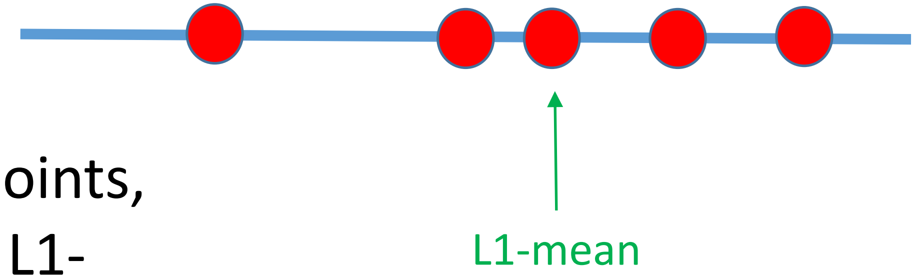
Mean value in L1 space - ***medoid***

What is the mean point in L1 space?

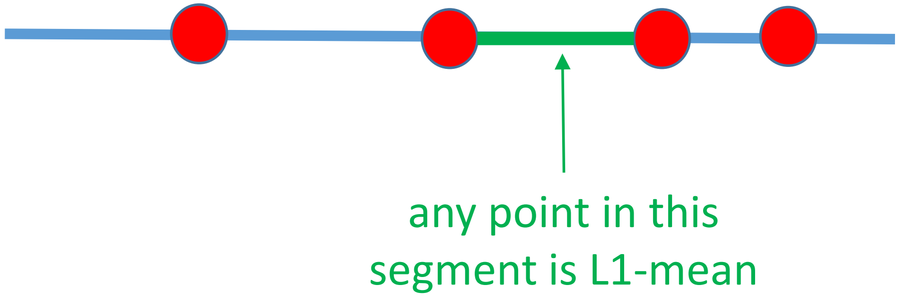
Mean in Euclidean distance is unique



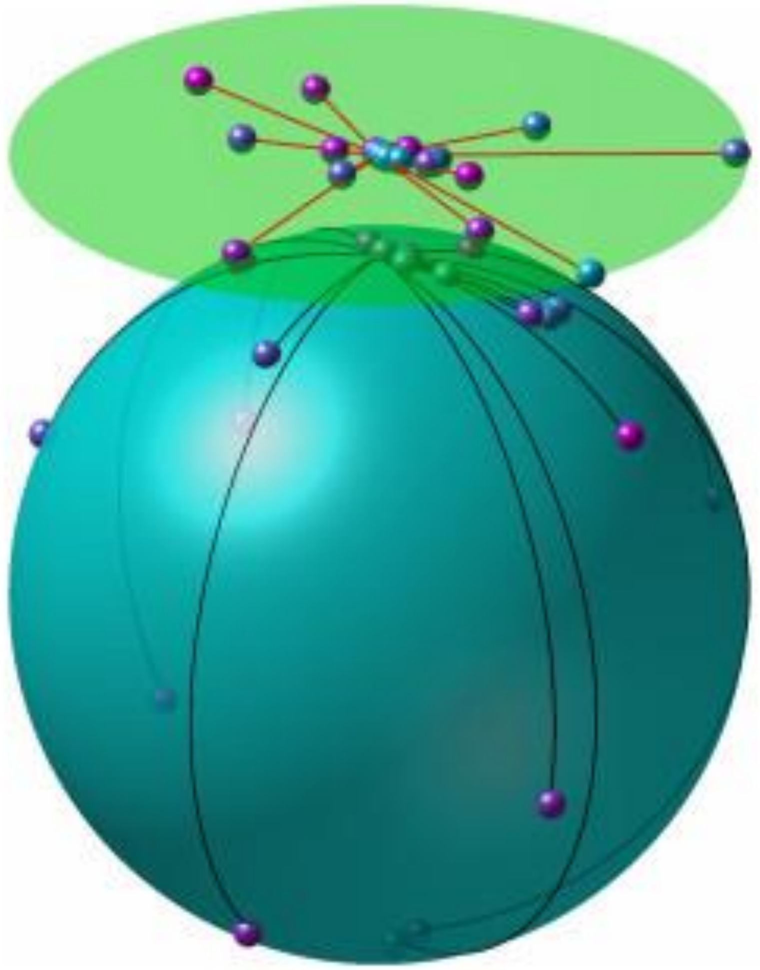
For odd number of data points, L1-mean is also unique



For even number of data points, there is infinite number of L1-means



Mean point on Rieman surface (e.g., sphere)



The distance is the length of the shortest path – of geodesics

Formula $$m = \arg\min_{p \in M} \sum_{i=1}^{N} d^2\left(p, x_i\right)$$

still holds!

Important generalizations of the mean point notion

- Mean value = best approximation of the data point cloud with **single object of zero dimension** (point)
- Best approximation of the data point cloud with **multiple objects of zero** dimension = *k-means clustering* (also called *k principal points*)
- Best approximation of the data point cloud with **multiple objects of zero** dimension in L1-space = *k-medoids clustering*
- Best approximation of the data point cloud with single object of dimension 1 = *first principal component*