Lab 8 – K-Means

Unsupervised learning

We reason on an unsupervised scenario and address the problem of data clustering

The method

$$\min_{m_1,\dots m_k} \sum_{i=1}^{n} \min_{j=1,\dots,k} ||x_i - m_j||^2$$

Hard to solve this functional, thus we resort to an approximation...

https://www.naftaliharris.com/blog/visualizing-k-means-clustering/

UniGe | MaLGa

The algorithm

$$\min_{m_1,\ldots m_k} \sum_{i=1}^{n} \min_{j=1,\ldots,k} ||x_i - m_j||^2$$

Update   **U**   **A** *Assignment*

- **Initialize the centroids in some way**

- **At each iteration:**

  - Compute the distance between each point and the centroids

  - Associate each point with the closest centroid *Assignment*

  - Update the centroids with the new groups *Update*

Your objectives today

- Implementing the Lloyd's algorithm

- Observing the effect of different strategies for initializing the centroids

- Observing the effects of number of samples, noise, ...

UniGe | MaLGa