# Unsupervised learning

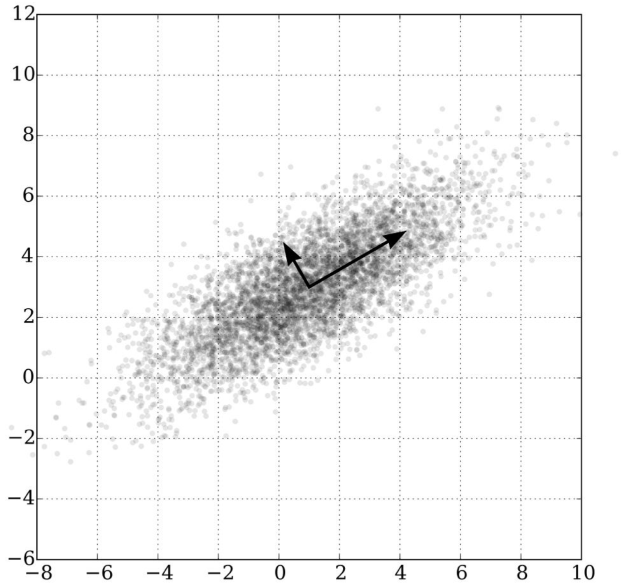
Nikita Mokrov

# Dimension reduction

- Principal Component Analysis (PCA & KernelPCA & BPCA & SPCA)
- Multidimensional scaling (MDS)
- Matrix Factorization (MF & NMF)
- Isomap
- Locally linear embedding (LLE)
- t-SNE (Umap)

# PCA

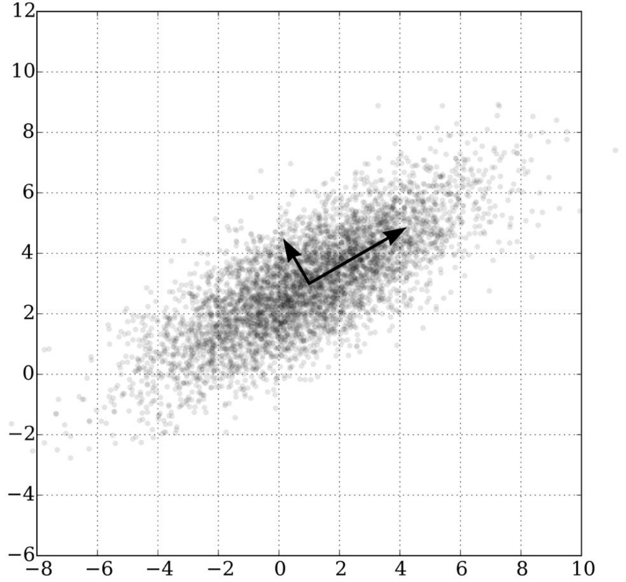
$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{(\mathbf{w}^T\mathbf{x}_i)^2\}$$



# PCA

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{(\mathbf{w}^T\mathbf{x}_\mathrm{i})^2\}$$

$$\mathbf{w}_2 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{[\mathbf{w}^T(\mathbf{x}_i - \underbrace{\mathbf{w}_1\mathbf{w}_1^T\mathbf{x}_i}_{})]^2\}$$

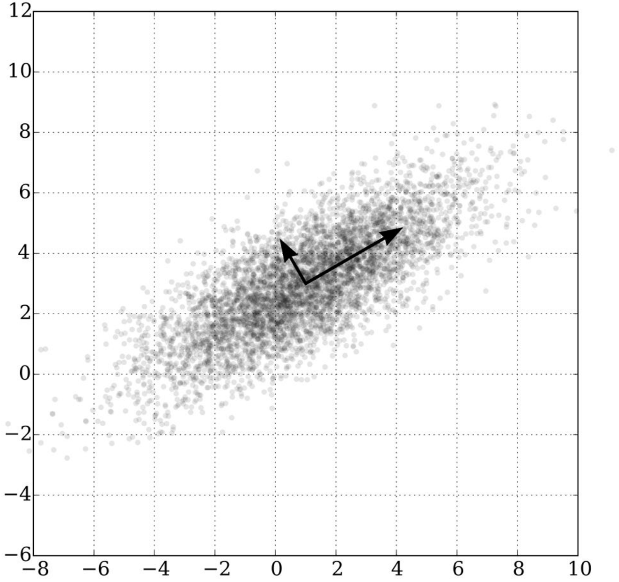


# PCA

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{(\mathbf{w}^T\mathbf{x}_\mathrm{i})^2\}$$

$$\mathbf{w}_2 = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{[\mathbf{w}^T(\mathbf{x}_i - \underbrace{\mathbf{w}_1\mathbf{w}_1^T\mathbf{x}_i}_{})]^2\}$$

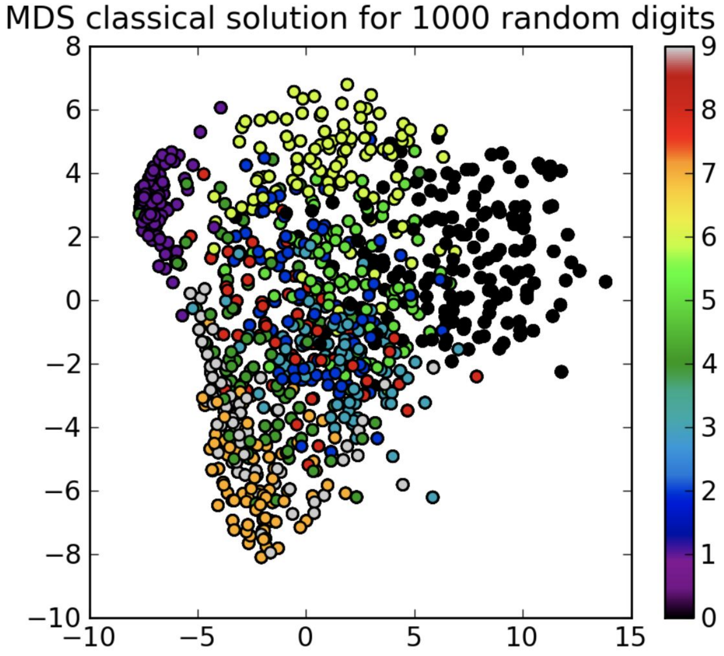
$$\mathbf{w}_k = \arg\max_{\|\mathbf{w}\|=1} \frac{1}{m}\sum_{i=1}^{m}\{[\mathbf{w}^T\underbrace{(\mathbf{x}_i - \sum_{j=1}^{k-1}\mathbf{w}_j\mathbf{w}_j^T\mathbf{x}_i)}_{}]^2\}$$



# MDS

$$\min_Y \sum_{i=1}^{t} \sum_{i=1}^{t} (\mathtt{d}_{ij}^{(X)} - \mathtt{d}_{ij}^{(Y)})^2$$

where $\mathtt{d}_{ij}^{(X)} = ||x_i - x_j||$ and $\mathtt{d}_{ij}^{(Y)} = ||y_i - y_j||$.

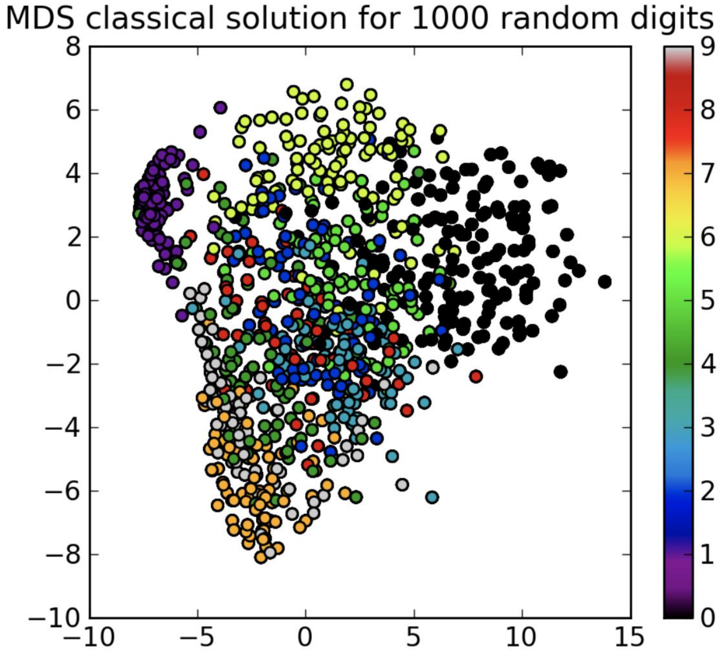


# MDS

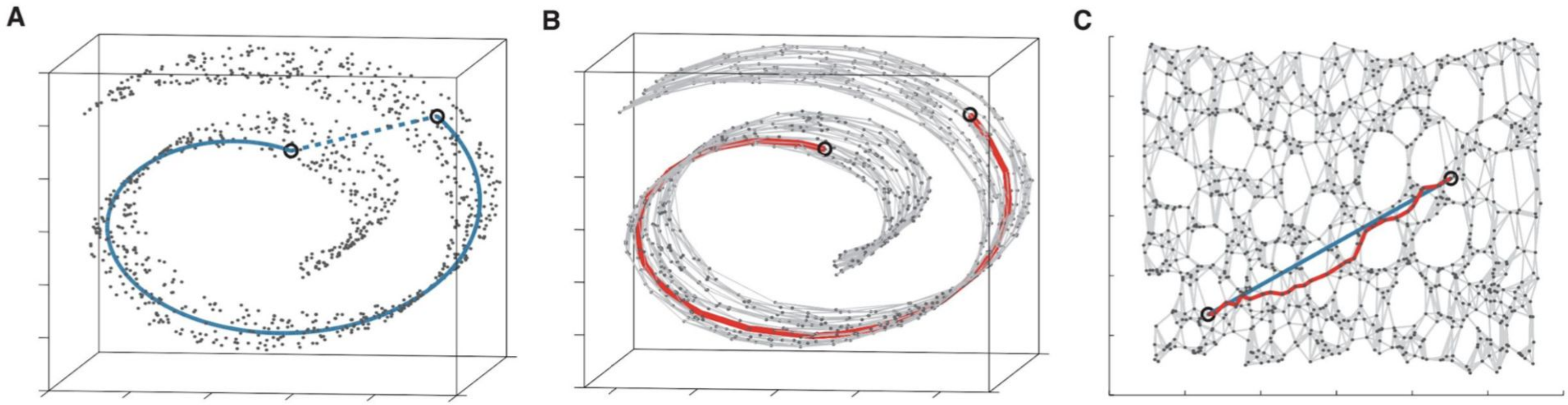
$$\min_{Y} \sum_{i=1}^{t} \sum_{i=1}^{t} (x_i^T x_j - y_i^T y_i)^2$$

$$X^T X = V \Lambda V^T$$

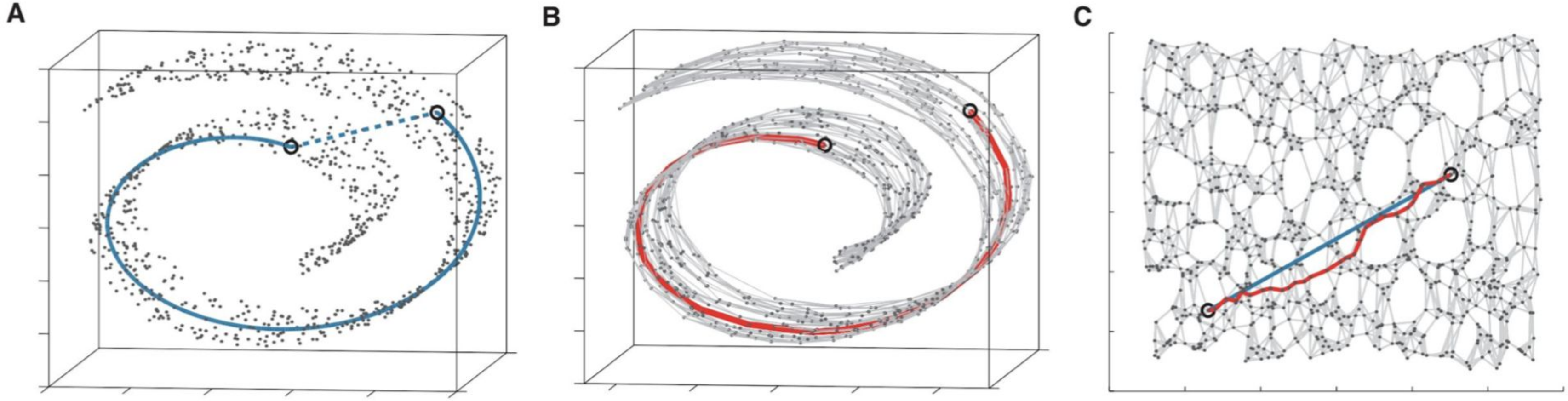
$$Y = \hat{\Lambda}^{1/2} V^T$$



# Isomap

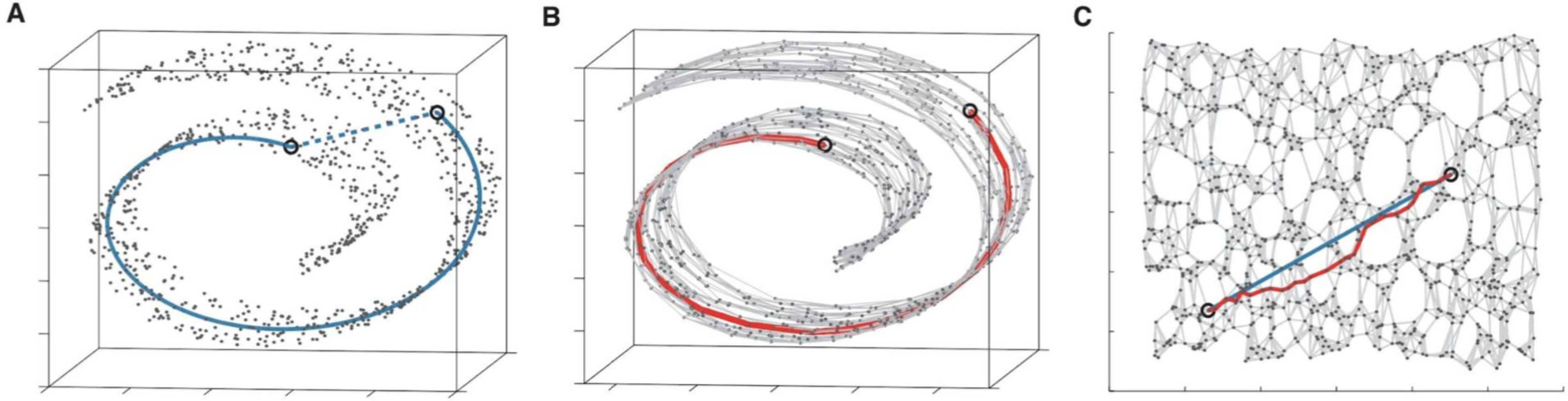


# Isomap



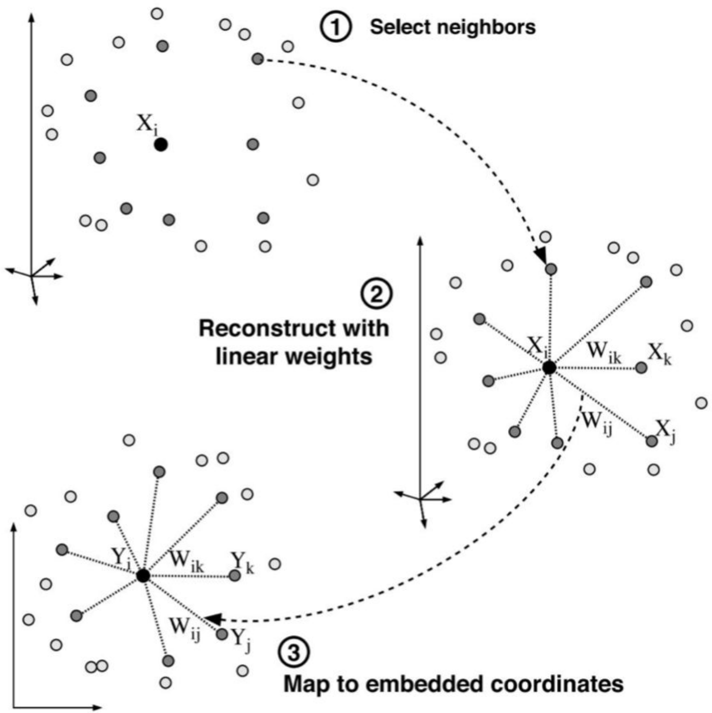

$$\min_{Y} \sum_{i=1}^{t} \sum_{i=1}^{t} (\mathtt{d}_{ij}^{(X)} - \mathtt{d}_{ij}^{(Y)})^2$$

# Isomap



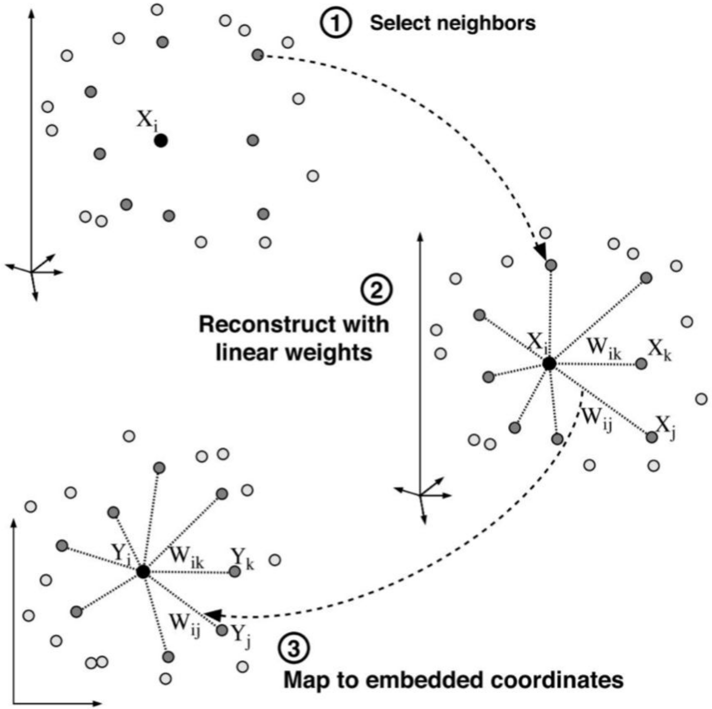

$$\min_{Y} \sum_{i=1}^{t} \sum_{i=1}^{t} (\mathtt{d}_{ij}^{(X)} - \mathtt{d}_{ij}^{(Y)})^2 \quad \longleftarrow \quad d_G(i,j)$$

# LLE



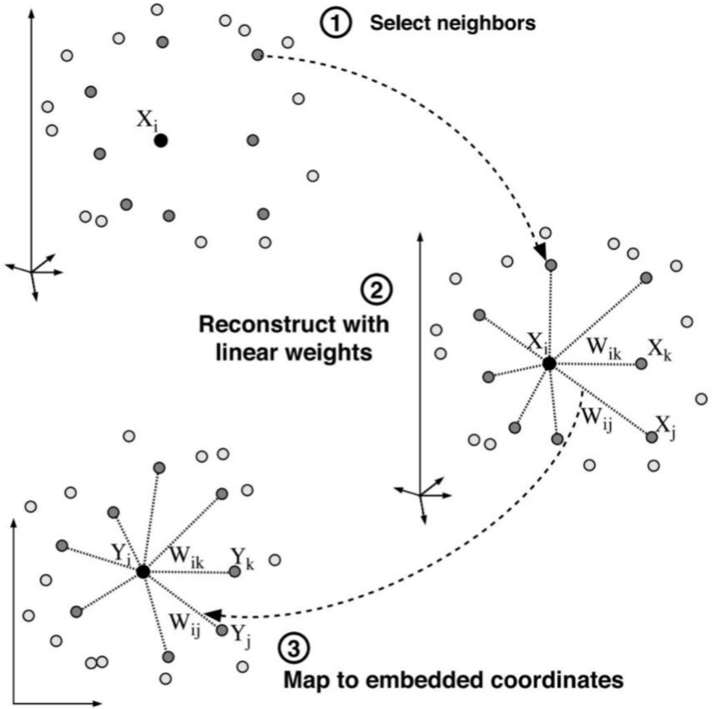
1 Select neighbors
$X_i$
2 Reconstruct with linear weights
$X_i$
$W_{ik}$
$X_k$
$W_{ij}$
$X_j$
$Y_i$
$W_{ik}$
$Y_k$
$W_{ij}$
$Y_j$
3 Map to embedded coordinates

# LLE




$$\varepsilon(W) = \sum_{i=1}^{n} ||x_i - \sum_{j=1}^{K} W_{ij} x_j||^2$$

# LLE




$$\varepsilon(W) = \sum_{i=1}^{n} ||x_i - \sum_{j=1}^{K} W_{ij} x_j||^2$$

$$\Phi(Y) = \sum_{i=1}^{n} ||y_i - \sum_{j=1}^{n} W_{ij} y_j||^2$$

# SNE

$$p_{j|i} = \frac{\exp(-\frac{||x_i - x_j||^2}{2\sigma_i^2})}{\sum_{k \neq i} \exp(-\frac{||x_i - x_k||^2}{2\sigma_i^2})}$$

$$q_{j|i} = \frac{\exp(-||y_i - y_j||^2)}{\sum_{k \neq i} \exp(-||y_i - y_k||^2)}$$

# SNE

$$p_{j|i} = \frac{\exp(-\frac{||x_i - x_j||^2}{2\sigma_i^2})}{\sum_{k \neq i} \exp(-\frac{||x_i - x_k||^2}{2\sigma_i^2})}$$

$$q_{j|i} = \frac{\exp(-||y_i - y_j||^2)}{\sum_{k \neq i} \exp(-||y_i - y_k||^2)}$$

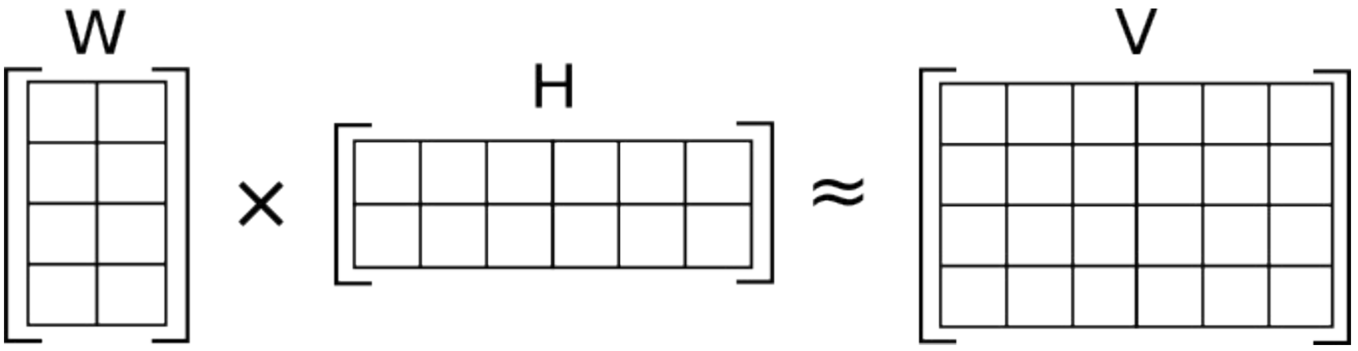
$$D_{KL}(P \,||\, Q) \to \min_Y$$

# tSNE

$$q_{ij} = \frac{\frac{1}{1+||y_i - y_j||^2}}{\sum_{k \neq i} \frac{1}{1+||y_i - y_k||^2}}$$

# MF

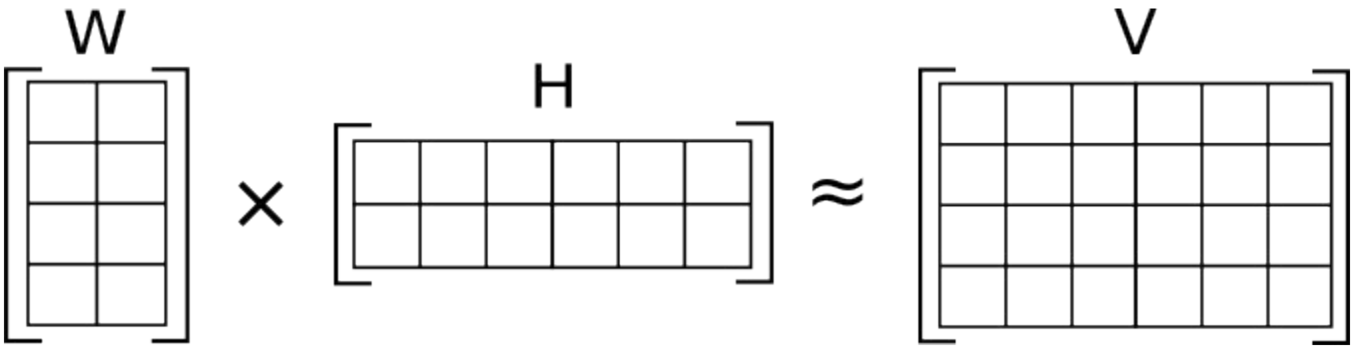
$$\mathbf{V} = \mathbf{W}\mathbf{H}\,.$$



# NMF

$$\mathbf{V} = \mathbf{W}\mathbf{H}\,.$$

$$W \geqslant 0, H \geqslant 0.$$



# Main ideas (review)

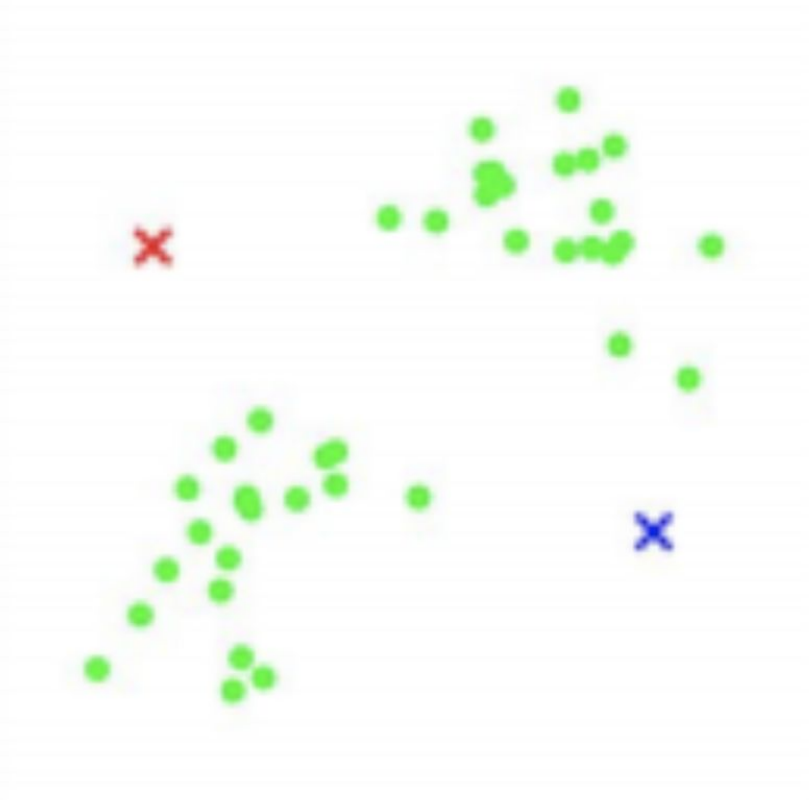
- PCA
- MDS
- MF & NMF
- Isomap
- LLE
- t-SNE

# Clustering

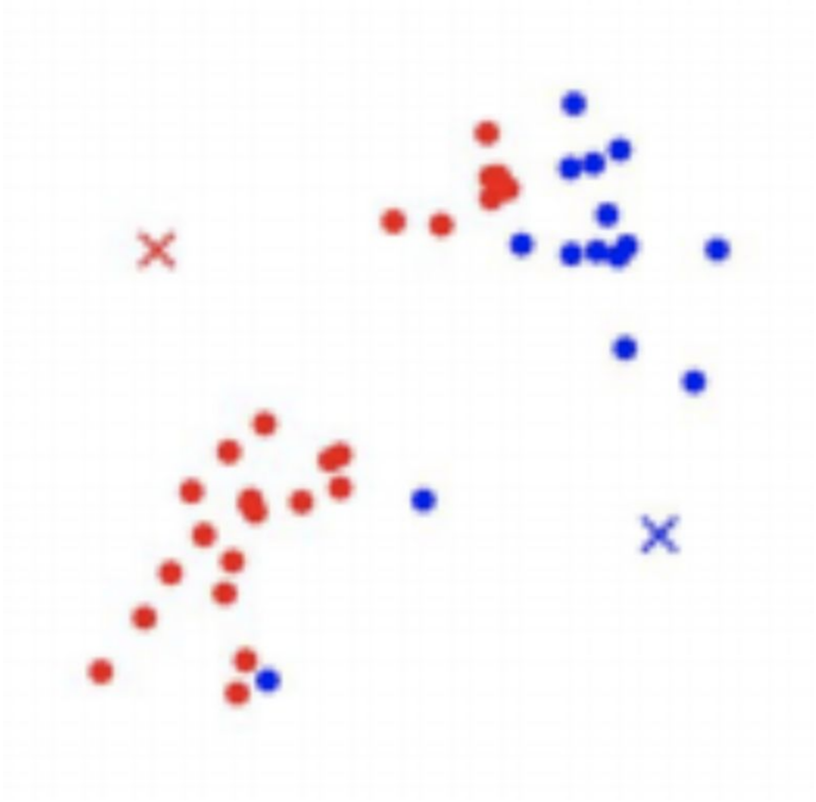
- KMeans
- Agglomerative Clustering
- DBSCAN

# kMeans

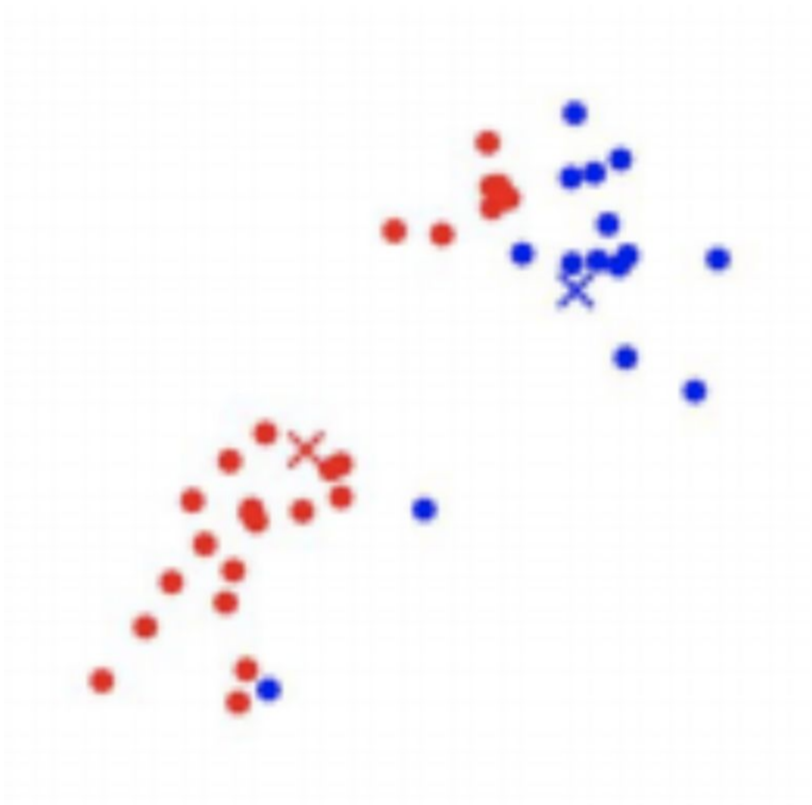
# kMeans



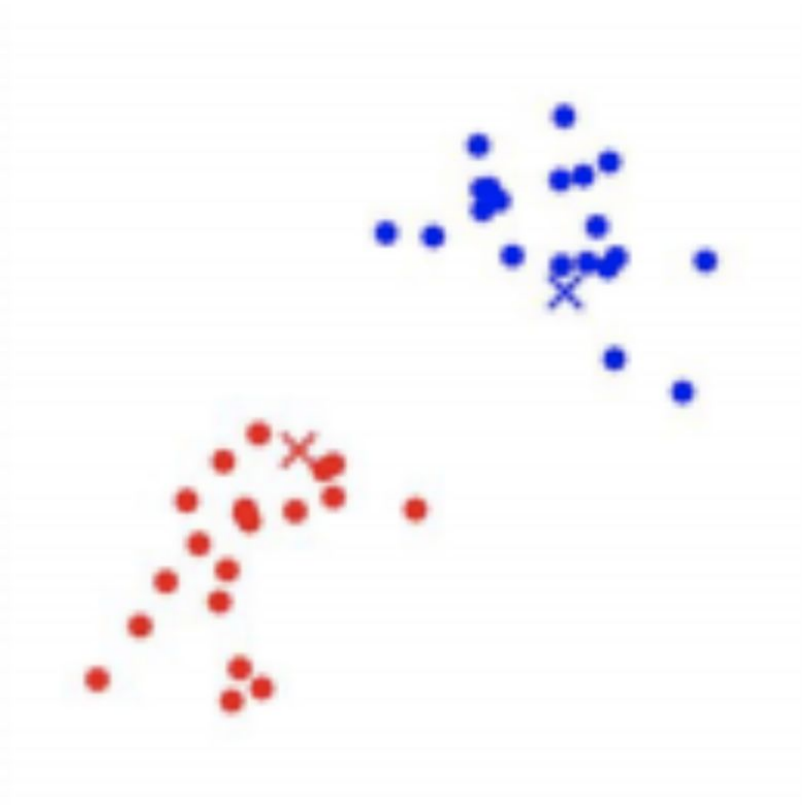
# kMeans



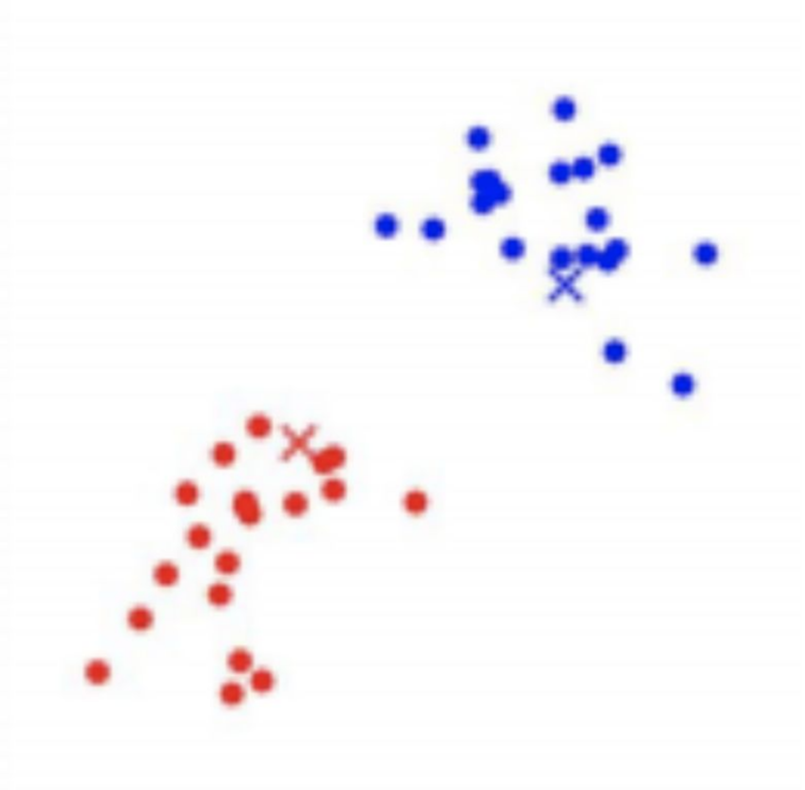
# kMeans



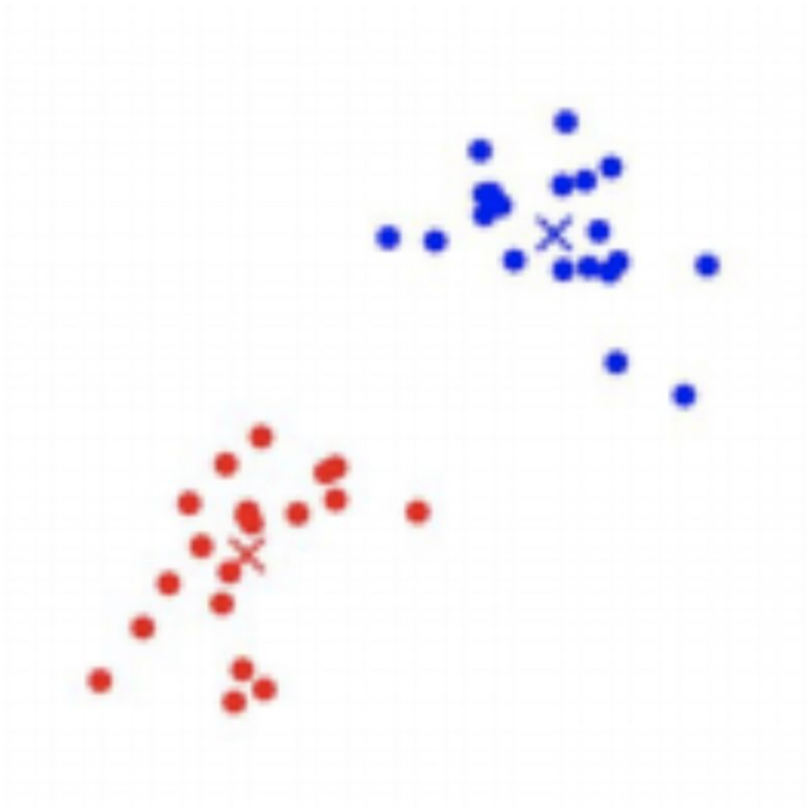
# kMeans



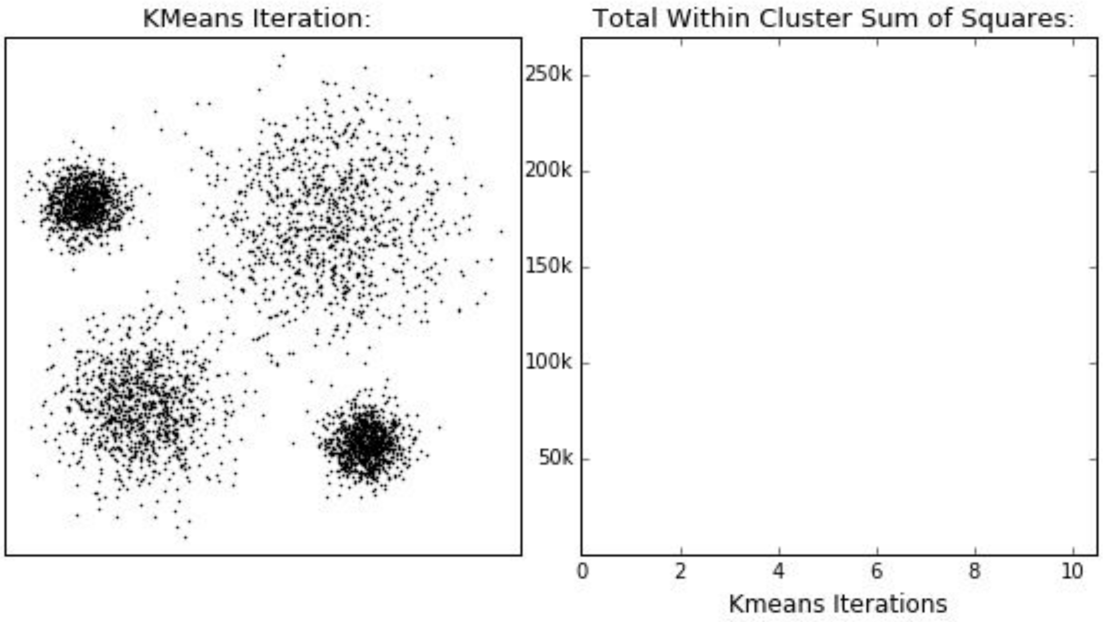
# kMeans



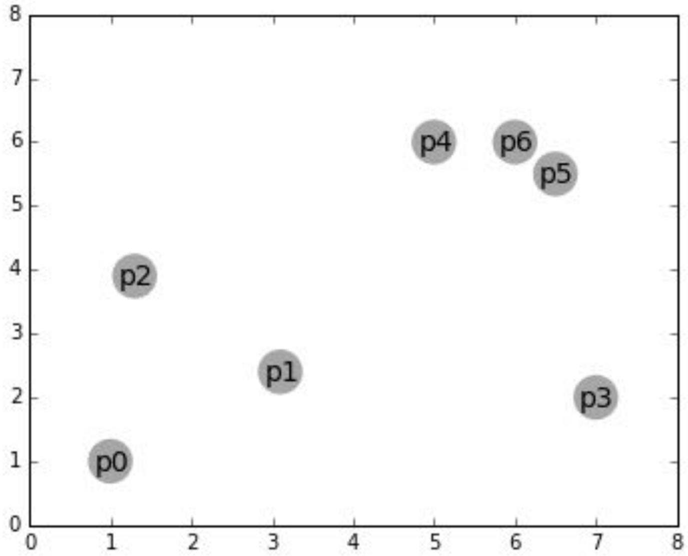
# kMeans

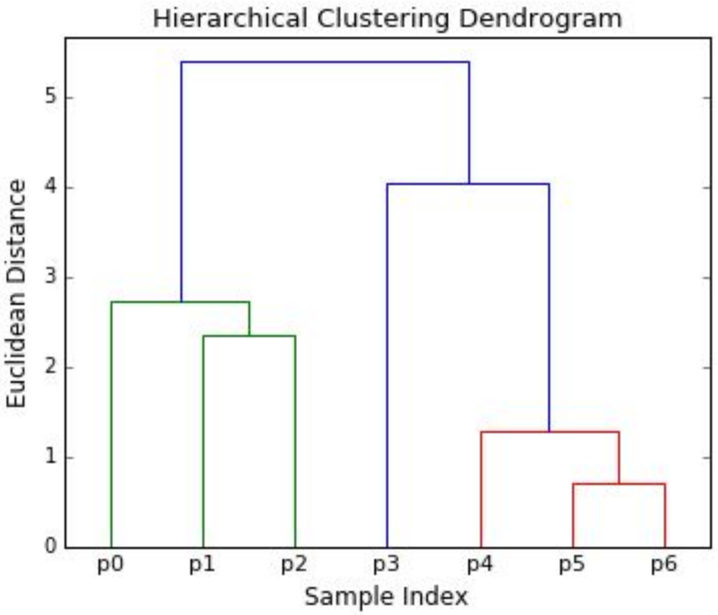


# kMeans

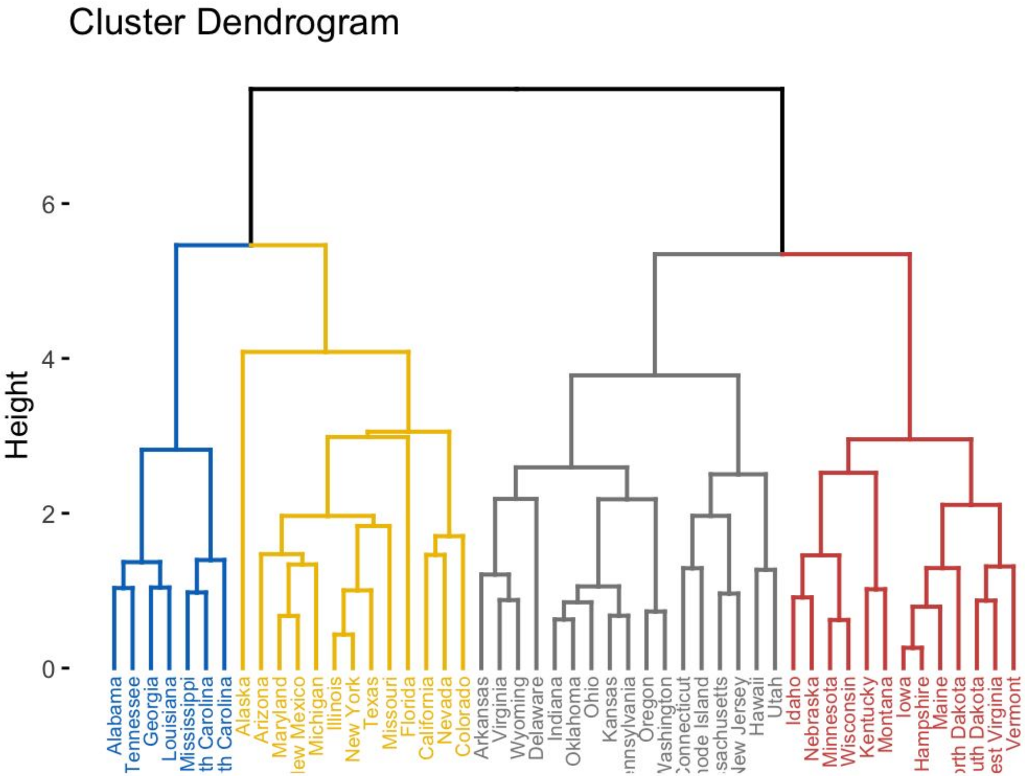

KMeans Iteration:
Total Within Cluster Sum of Squares:
250k
200k
150k
100k
50k
0
2
4
6
8
10
Kmeans Iterations

# Agglomerative clustering

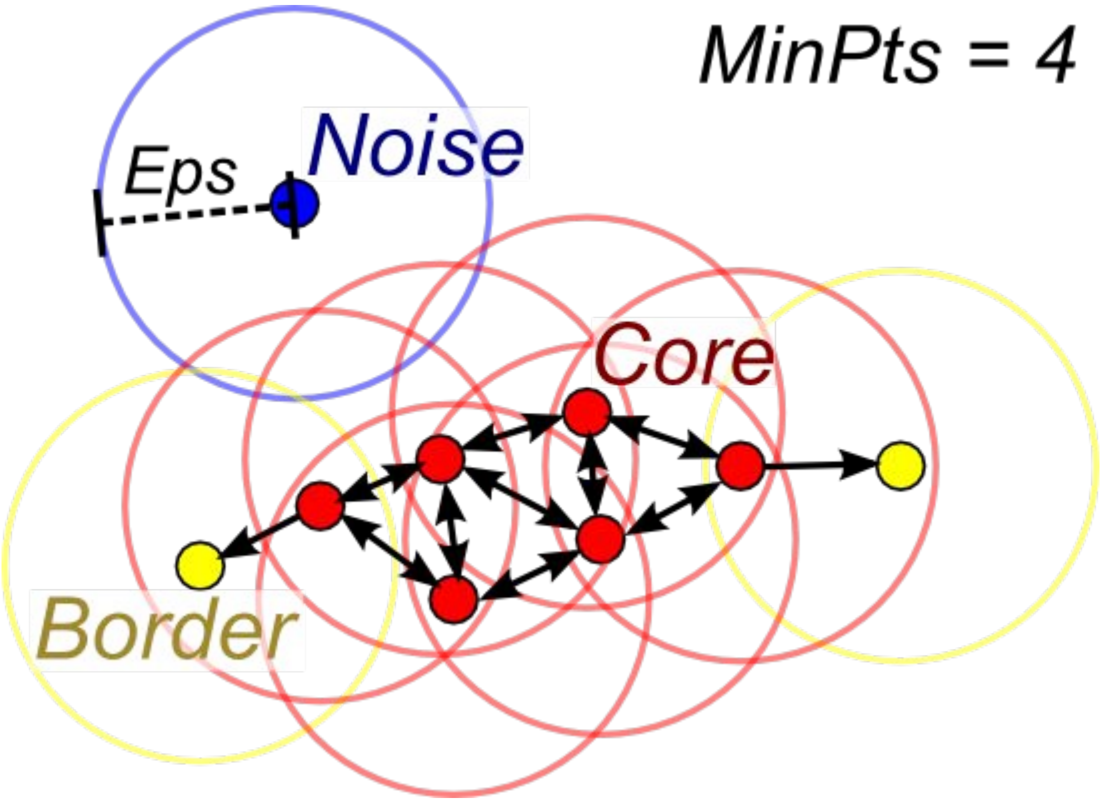


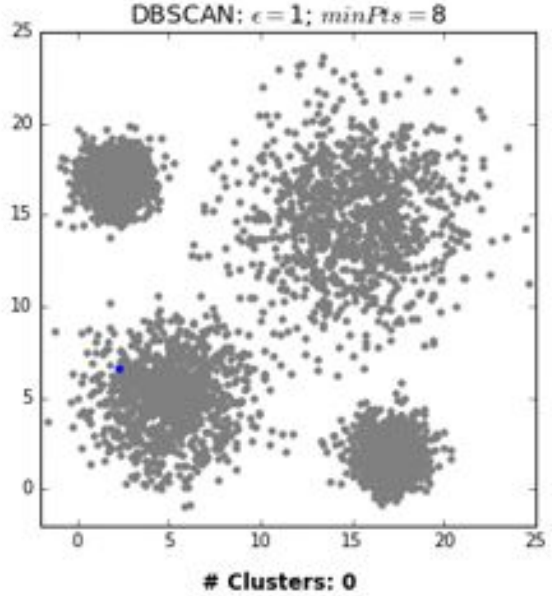
# Agglomerative clustering

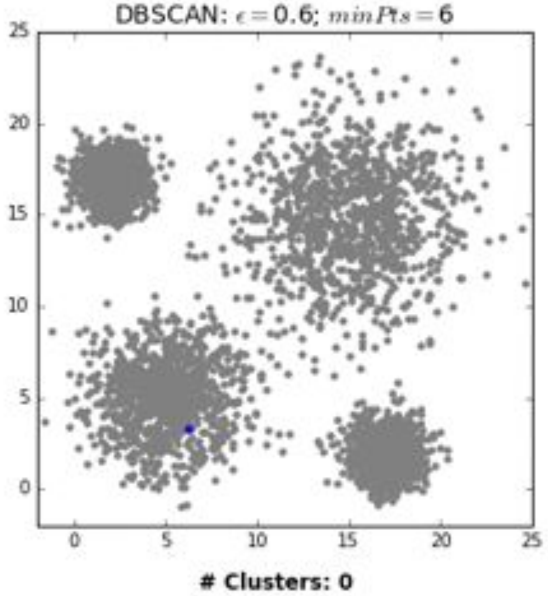


# DBSCAN



# DBSCAN


DBSCAN: $\epsilon = 1$; $minPts = 8$
# Clusters: 0


DBSCAN: $\epsilon = 0.6$; $minPts = 6$
# Clusters: 0

# Main Ideas (review)

- KMeans
- Agglomerative Clustering
- DBSCAN