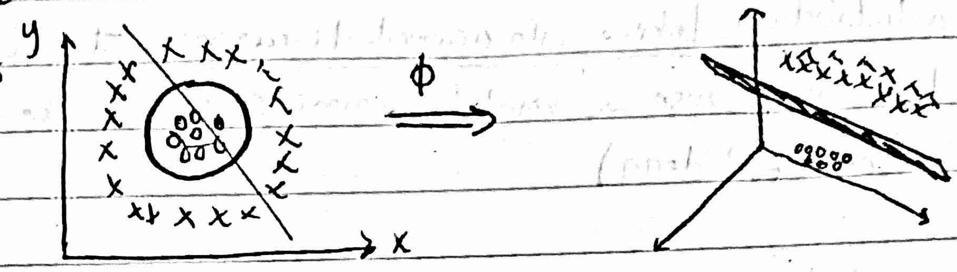
Discussion 9

Kernel clustering

1. Can the polynomial kernel $K(x_1, x_2) = (x_1^T x_2 + b)^2$ cluster concentric data?

— Regular k-means
— Goal (circle)



Linear boundary in mapped space (hyperplane)

$K(x,y) = (x^T y + b)^2$ Let $b=0$ (we already did $b=1$ before!)

$= (x^T y)^2 = \underbrace{(x_1 y_1 + x_2 y_2)}_{2D}^2$

$= x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 y_1 x_2 y_2$

$= \begin{bmatrix} x_1^2 & x_2^2 & 2x_1x_2 \end{bmatrix} \begin{bmatrix} y_1^2 \\ y_2^2 \\ 2y_1y_2 \end{bmatrix} = \phi(x)^T \phi(y)$

So, $\phi(x) = \begin{pmatrix} x_1^2 \\ x_2^2 \\ 2x_1x_2 \end{pmatrix} \Rightarrow$ Adds 3rd dimension (what we want)

As we saw in the HW, this will be a quadratic decision boundary in the input space (weighted combination of order 1 & 2 terms of original feature vector x).

(Keep in mind that $(x^T y + b)^2$ will create quadratic boundary $\forall b$)

Decision boundary: $\|x - \mu_1\|_2^2 = \|x - \mu_2\|_2^2 \Rightarrow 2(\mu_2 - \mu_1)^T x + (\mu_1^T \mu_1 - \mu_2^T \mu_2) = 0$

Form: $w^T x + b = 0$ (Linear)

Decision boundary: $\|\phi(x) - \mu_1\|_2^2 = \|\phi(x) - \mu_2\|_2^2$ (Mapped space)

Form: $w^T \phi(x) + b = 0$

2.) Alternative distance metric for assignment step in k-means

$$d_{Mahalanobis} = \sqrt{(x_i-\mu_j)^T \Sigma_j^{-1} (x_i-\mu_j)}$$

Write out k-means w/ this metric.

(Remember Mahalanobis distance represents distance of a point to a distribution, taking into account its variance. It fits in nicely w/ k-means because we want to minimize it, just like we did with Squared Euclidean)

1. Initialize : μ, Σ $\forall$ classes ($\Sigma = I$ $\forall$ classes usually)
2. Assignment:

   for $x_i \in i = \{1 \ldots n\}$:

   Assign each x_i to closest centroid according to Mahalanobis distance, i.e.

   $$\underset{c_j \in C}{\text{argmin}} \sqrt{(x_i-\mu_j)^T \Sigma_j^{-1} (x_i-\mu_j)} = \text{label}_{x_i}$$

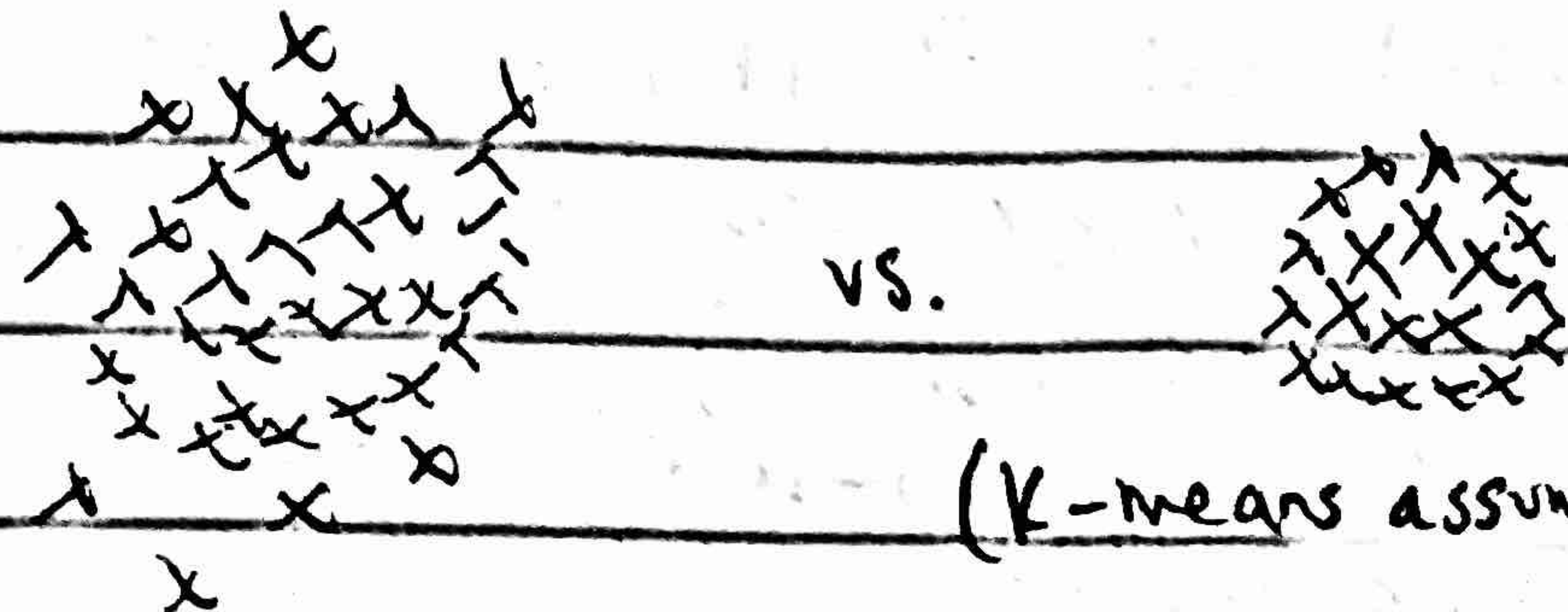
3. Update:

   $$\mu_j = \frac{1}{n_j} \sum_{i \in c_j} x_i$$

   $$\Sigma_j = \frac{1}{n_j} \sum_{i \in c_j} (x_i-\mu_j)(x_i-\mu_j)^T$$

4. Repeat until convergence

This distance metric might be useful because it takes into account the variance of a cluster, i.e.



(K-means assumes covariance = identity)