Lecture18

Thursday, November 3, 2016 10:22 PM

Exam II Syllabus

1) SVM
   - → margin, separating hyperplane
   - → soft margin
   - → kernel SVM

   $\vec{w}$?, b ?

2) Evaluation for Classification
   - → Metrics (accuracy, F1, ROC, etc.)
   - → methodology (Cross validation, t-test, Confidence Intervals)
   - → Bias-Variance
   - → Ensemble classifiers (bagging, boosting)

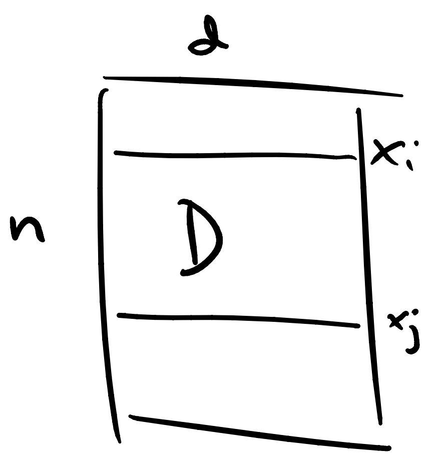
3) Least squared Regression
   - → linear
   - → kernel

   Logistic Regression

4) Neural Networks
   - → activation functions
   - → feed forward / back propagation

Clustering

Unsupervised

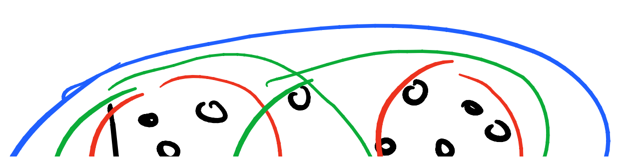


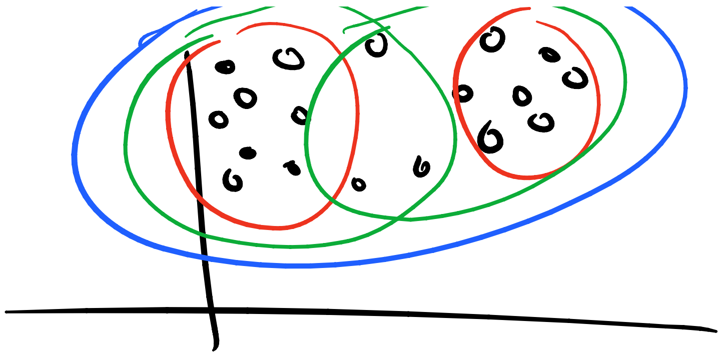
$D \in \mathbb{R}^{n \times d}$

finding "groups" in the data

↓

$\underbrace{\text{Similarity vs. Distance}}$





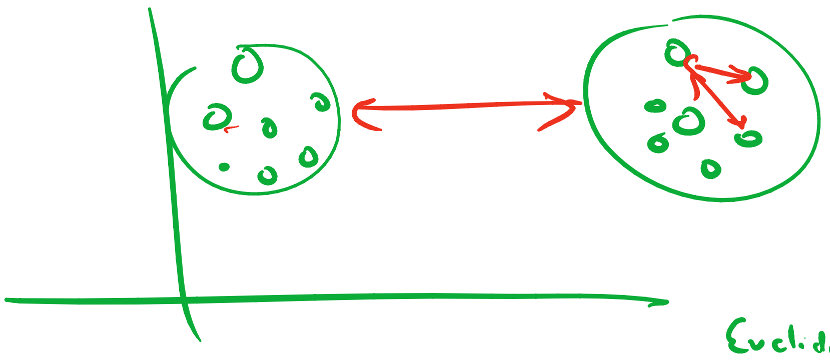
how many groups/clusters are there?

K : # of clusters

user-defined

hierarchical clusters

Overlapping clusters

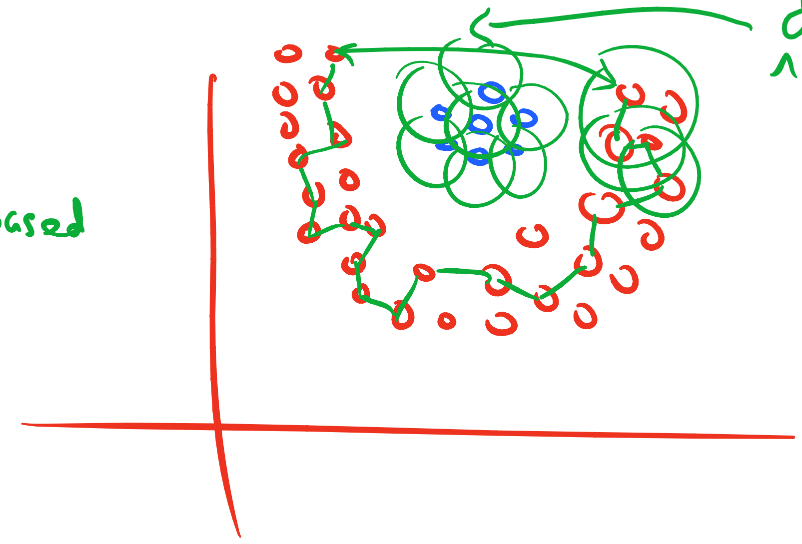


maximize inter-cluster distance

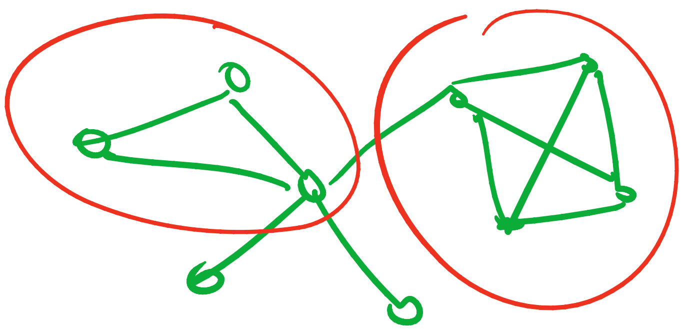
minimize intra-cluster distance

Euclidean distance is larger

density-based



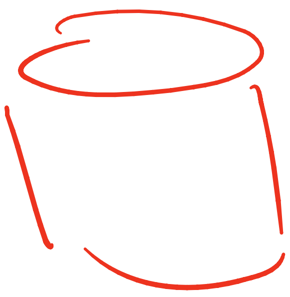
Convex shapes



graph clustering

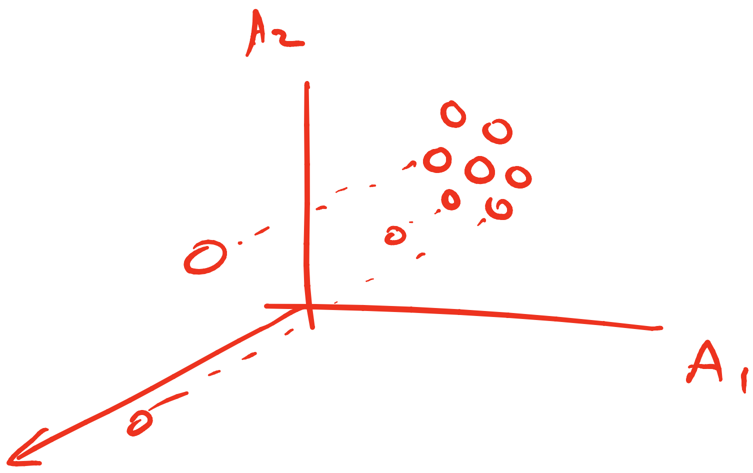
"distance" : length of shortest path

$d = 100, 1000, \dots$



In high dimensions:

data is scattered along the boundary & in the corners



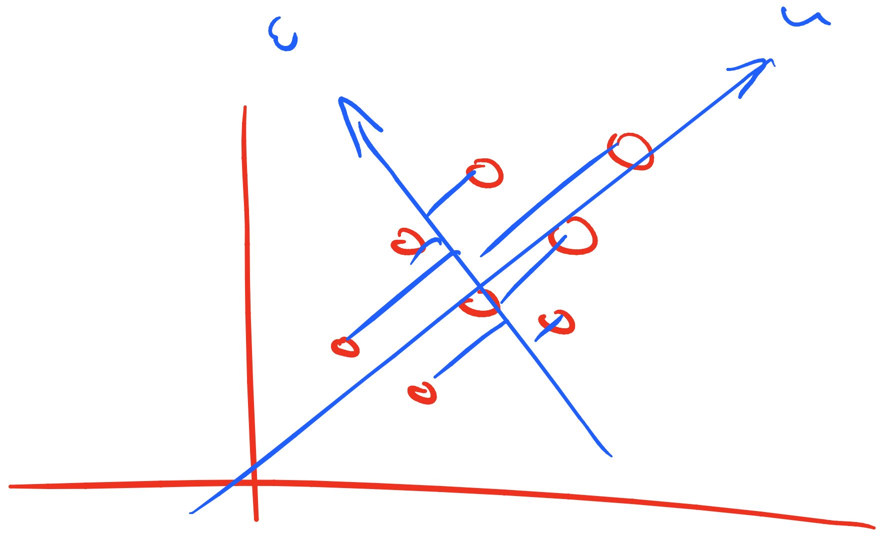
Z

$D \in \mathbb{R}^{n \times d}$

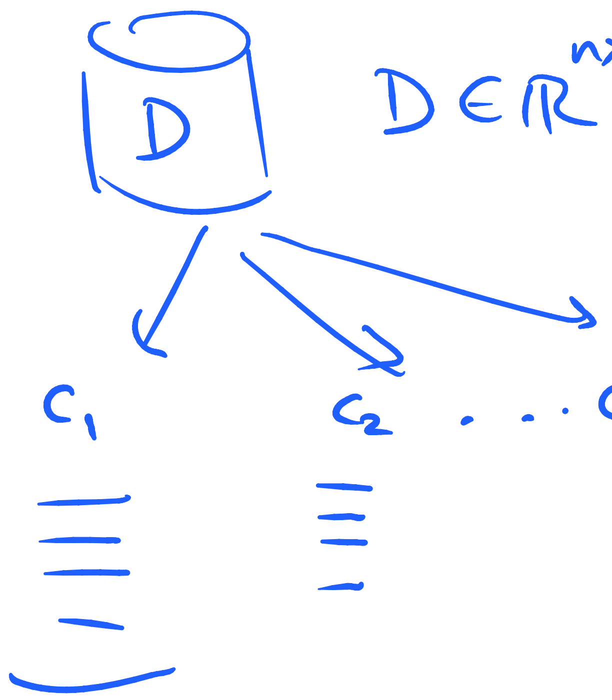
find the good subspace where the data clusters well

Subsets

linear Combinations



K-Means Algorithm



$D \in \mathbb{R}^{n \times d}$

$C_1 \quad C_2 \quad \dots \quad C_k$

$k \leftarrow$ # of clusters to find

partitioning based clustering

into k parts

representative point

$\vec{\mu}_1 \qquad \vec{\mu}_2 \quad \dots \quad \vec{\mu}_k$

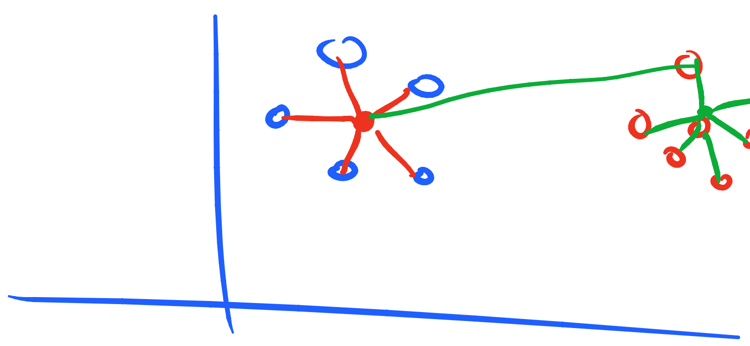
"best" partitioning

$$\underline{SSE}: \sum_{i=1}^{k} \sum_{x_j \in C_i} \| \vec{x}_j - \vec{\mu}_i \|^2$$

$$\min_{\{C_1, C_2, \dots, C_k\}} SSE$$

over all partitionings

even for $k = 2$



even for $k = 2$

NP_hard problem.

K-Means

1) Initialization step:

guess the initial centers

$\mu_1, \mu_2, \ldots, \mu_k$

Should do multiple runs

Choose the answer with smallest SSE

2) for all $i = 1, \ldots, k$ ⟵ all clusters k

for all $j = 1, \ldots, n$ ⟵ all points in D n

$O(n \cdot k \cdot d) \times t$ total complexity

t # of iterations

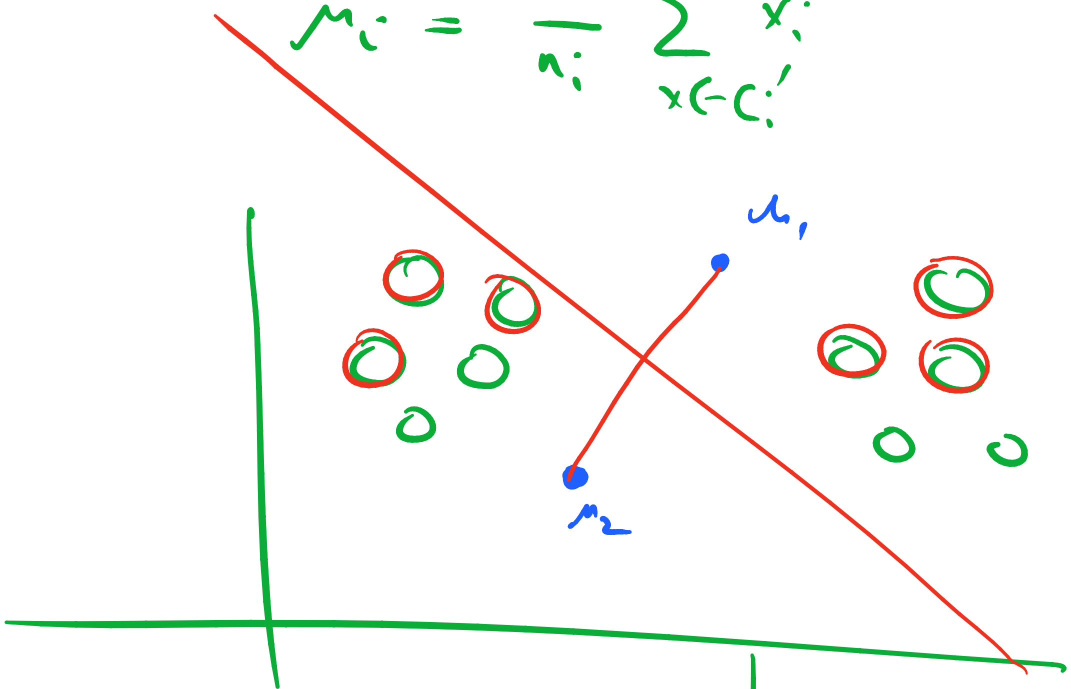
$$S(i,j) = \underbrace{\| \vec{x}_j - \vec{\mu}_i \|^2}_{d}$$

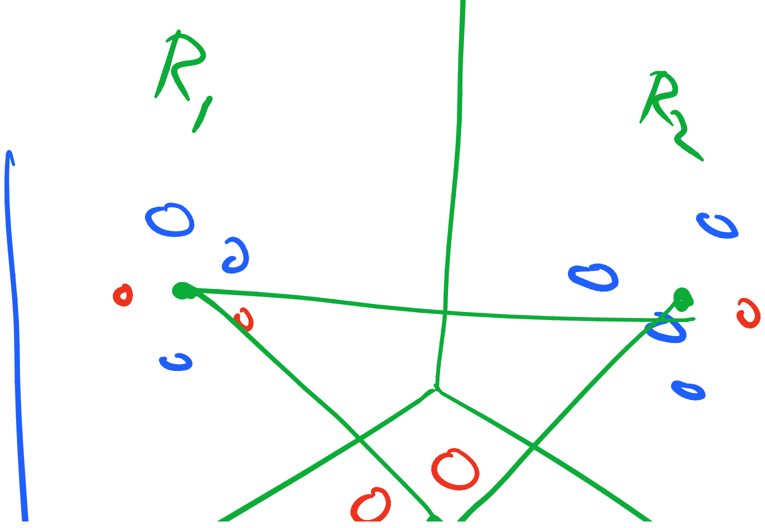
use $S(i,j)$ to figure the closest mean for point x_j

get new groups $C_1', C_2', \ldots, C_k'$

3) Use new groups to find the new mean points

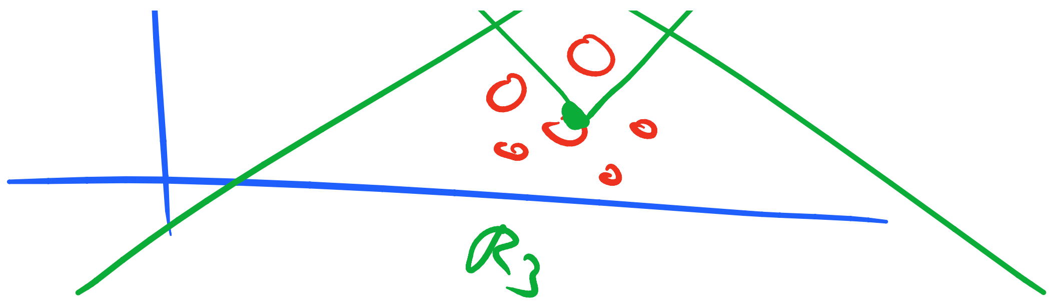
$$\mu_i' = \frac{1}{n_i} \sum_{x \in C_i'} x_i$$



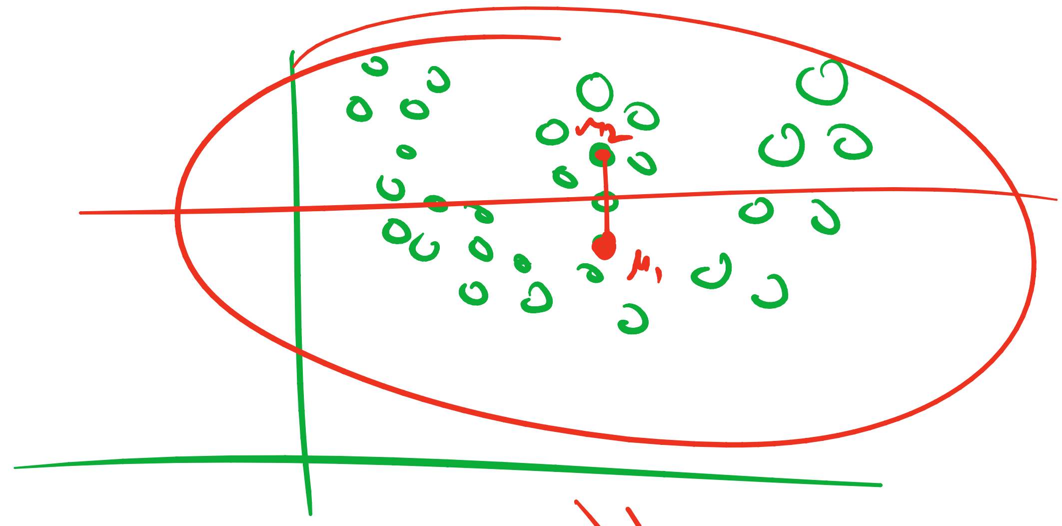


Geometry

Voronoi partitions



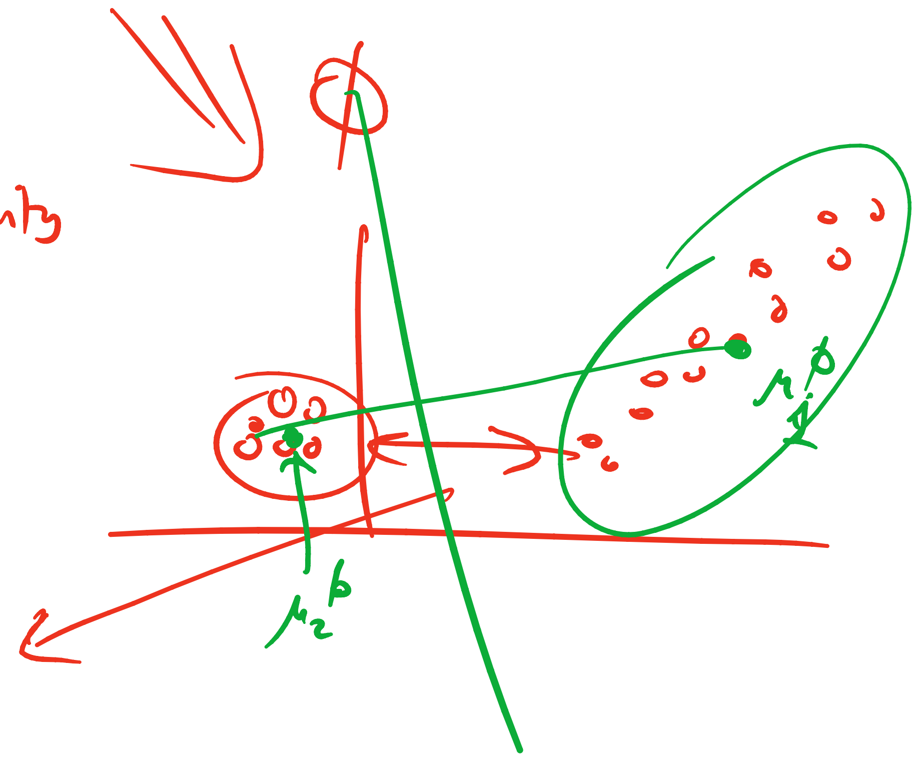
partitions : Convex



"blessing of dimensionality"

$\Downarrow$

use the scattering to find groups



Kernel Kmeans

all operations should involve only K matrix
Kernel

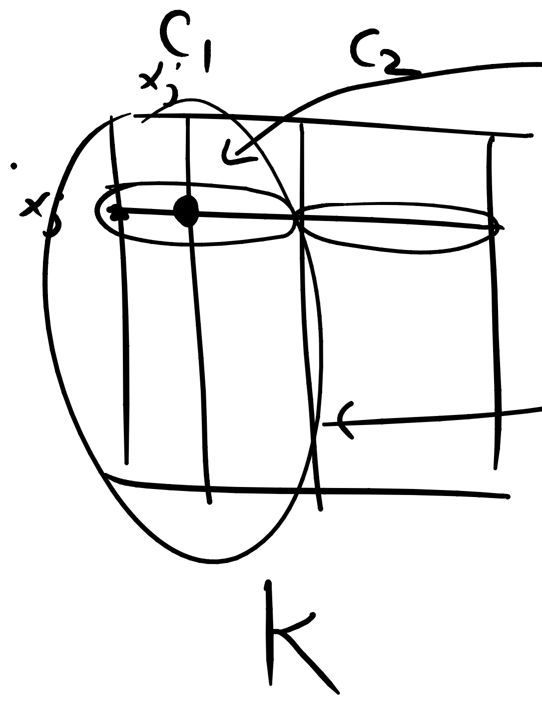
$x \longrightarrow \phi(x)$
?

$\mu_i^\phi \longleftarrow$ mean in feature space for C_i

$$\boxed{\mu_i^\phi = \frac{1}{n_i} \sum_{x_j \in C_i} \phi(x_j)}$$

$$\delta(\phi(x_j), \mu_i^\phi) = \|\phi(x_j) - \mu_i^\phi\|^2$$

$$= \underbrace{K(x_j, x_j)} - \underbrace{\frac{2}{n_i} \sum_{x_a \in C_i} K(x_a, x_j)} + \underbrace{\frac{1}{n_i^2} \underbrace{\sum_{x_a} \sum_{x_b}}_{C_i} K(x_a, x_b)}_{\text{sq. mean}}$$



~~① guess the means~~ $\underbrace{\mu_1^\phi, \mu_2^\phi, \dots, \mu_k^\phi}$

⓪ Compute **K**

① start with random partitioning of the points
balanced

$C_1, C_2, \dots, C_k \leftarrow$ initial guess

② cluster assignment step

for all $i = 1 \dots k$
  for all $j = 1 \dots n$
    $s(i,j) = K(x_j, x_j) - \frac{2}{n_i} \sum_{C_i} K(x_a, x_j) + \frac{1}{n_i^2} \underbrace{\sum\sum}_{C_i} K(x_a, x_b)$
  assign x_j to closest center

results in a new partitioning
$C_1', C_2', \dots, C_k'$

③ Stopping criteria:

Compare prev: $C_1, C_2, \dots, C_k$
new: $C_1', C_2', \dots, C_k'$

C_1 $\{1\,2\,3\}$ C_2 $\{4\,5\}$
$\downarrow$
C_1' $\{2\,5\}$ C_2' $\{1\,3\,4\}$

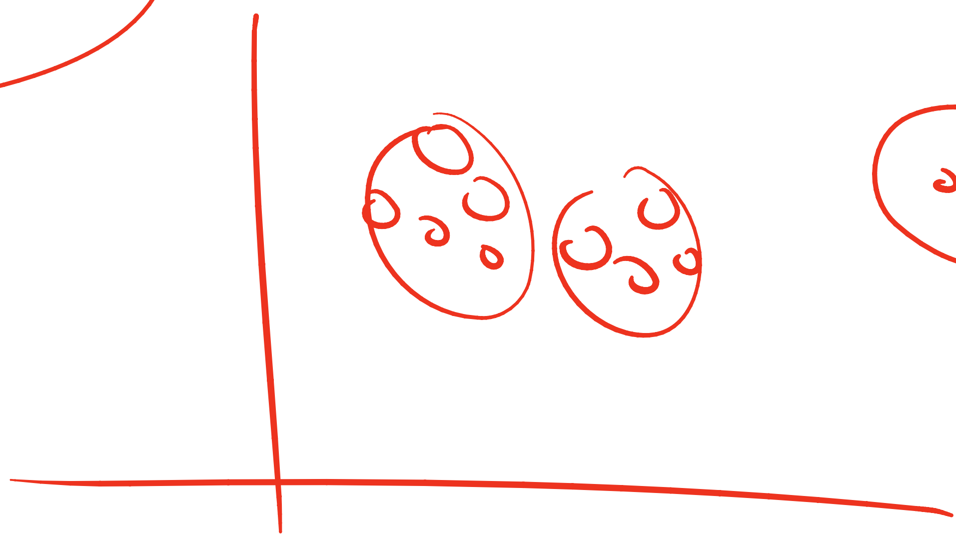
C_1' C_2'

$K(i,j) = x_i^T x_j$

→ linear kernel ⟹ regular k-mean

K $\quad \underbrace{O(n^2)}_{\text{kernel kmeans}} \times t$ ← # of iterations

$n^2 \gg n.k.d$



Probabilistic Mixture Model

→ Gaussian Mixture

→ EM algo for MLE

max. likelihood Estimate

Expectation Maximization

---

① Assume that clusters are Gaussian in shape

C_i parameter $\{\vec{\mu}_i, \Sigma_i, P(C_i)\}$

mean covariance Prior

mean covariance matrix Prior

likelihood

$$\boxed{\underbrace{P(C_i \mid x_j)}_{\text{Posterior probability}} = \frac{P(x_j \mid C_i) \cdot P(C_i)}{P(x_j)}}$$

$$P(x_j \mid C_i) \propto N(x_j \mid \mu_i, \Sigma_i)$$

Proportional to.

$$\boxed{P(x_j) = \sum_{i=1}^{k} P(x_j \mid C_i) \cdot P(C_i)}$$

$$P(x_j) = \underbrace{P(x_j \mid C_1) \cdot P(C_1) + P(x_j \mid C_2) \cdot P(C_2) + \cdots}_{\text{Mixture of Gaussians}}$$

$$\sum_{i=1}^{k} P(C_i) = 1$$

Maximize the likelihood over all points

$$P(D \mid \theta) \leftarrow L = \prod_{j=1}^{n} P(x_j) \leftarrow \text{the clustering is implicit}$$

$$\boxed{\ln L = \sum_{j=1}^{n} \ln(P(x_j))}$$

θ = set of all k parameter sets

Θ = set of all k parameter sets

$$\{ \mu_i, \Sigma_i, p(c_i) \}$$

$$\frac{\partial \ln L}{\partial \mu_i}, \frac{\partial \ln L}{\partial \Sigma_i}, \frac{\partial \ln L}{\partial p(c_i)}$$

① random guess of $\mu_i, \Sigma_i, p(c_i)$

② Expectation step:

$$p(c_i \mid x_j) \quad \forall i \ \forall j$$

③ maximization step

new $\mu_i, \Sigma_i, p(c_i)$