$\boxed{\text{HW2 Part2}}$

1. $\{x_i\}_{i=1}^{N}$ Data Likelihood of data

$$\boxed{\mathcal{L} = \prod_{i=1}^{N} p(x_i|\theta)}$$

1.1 App.1 : MLE : finds the parameters of the data likelihood

$$\boxed{\underset{\theta}{\arg\max}\ \mathcal{L}}$$

1.2 App2 : MAP : finds the parameters of data likelihood assuming a prior distribution

$$\mathcal{L}_2 = \underbrace{\prod_{i=1}^{N} p(x_i|\theta)}_{\text{data likelihood}} \cdot \underbrace{p(\theta|\alpha,\beta)}_{\text{prior is Gamma distribution}}$$

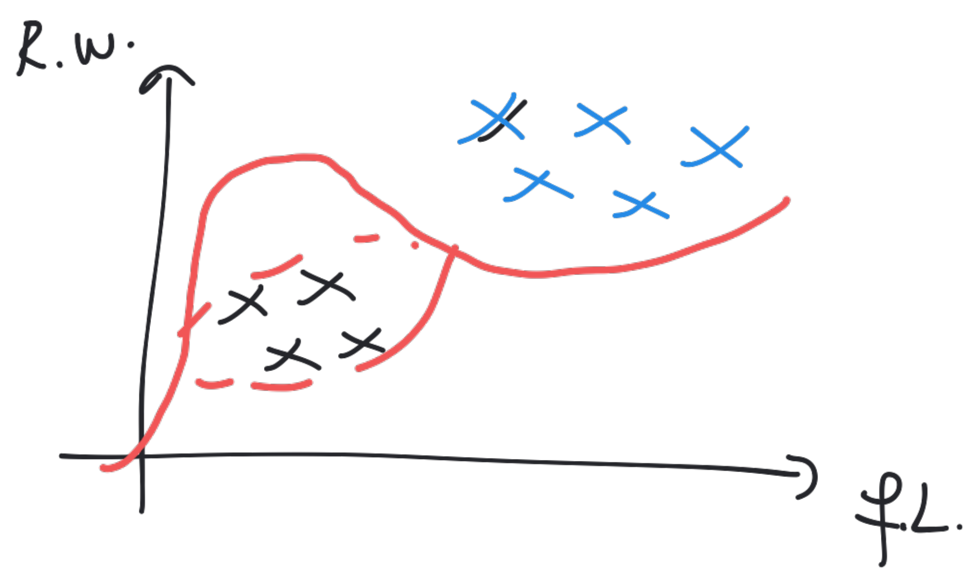
$$\boxed{\underset{\theta}{\arg\max}\ \mathcal{L}_2}$$

(2)

	Species	Frontal Lip	Rear Width	Length	Width	Depth	Male	Female
1	0							
2	1							
⋮	⋮							
140	0							
141								
⋮								
200								

Feature Matrix X

$\mathcal{L}C_0 = \text{multivariate-gaussian}(\text{training}, \mu_0, \Sigma_0).$



(1) Find mean and covariance for samples in species 0

(2) Find mean and covariance for samples in species 1

$$\underbrace{G(x|C_0)}_{\text{multivariate-gaussian}} \times p(C_0) \underset{C_1}{\overset{C_0}{\gtrless}} G(x|C_1) \times p(C_1)$$

training data

If we use 7 features,
we will obtain an error
"singular matrix"
$\Longrightarrow$ det $= 0$
$\Longrightarrow$ inverse does not exist

Multivariate Gaussian $p(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{1/2} \cdot |\Sigma|^{1/2}} \cdot \exp\left(-(x-\mu)^T \Sigma^{-1} \cdot (x-\mu)\right)$

If a matrix Σ is singular,

Σ^{-1} does not exist

$$\left(\Sigma + \alpha \cdot I\right)^{-1} \quad ,$$

Lectures 6, 7

K-Means obj. fct:

$$J(\theta, U) = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} \cdot \overbrace{d^2(x_i, \theta_k)}^{\text{Metric to define similarity between points}}$$

Metric to define similarity between points
$\downarrow$
Distance metric

Crisp assignment: $u_{ik} = \{0, 1\}$, $\sum_{k=1}^{K} u_{ik} = 1$

Membership matrix U: $N \times K$

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & \cdots & K \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_N \end{array} & \begin{bmatrix} 0 & 1 & 0 & - & 0 \\ 0 & 0 & 0 & - & 1 \\ & & & & \\ & & & & \end{bmatrix} \end{array}$$

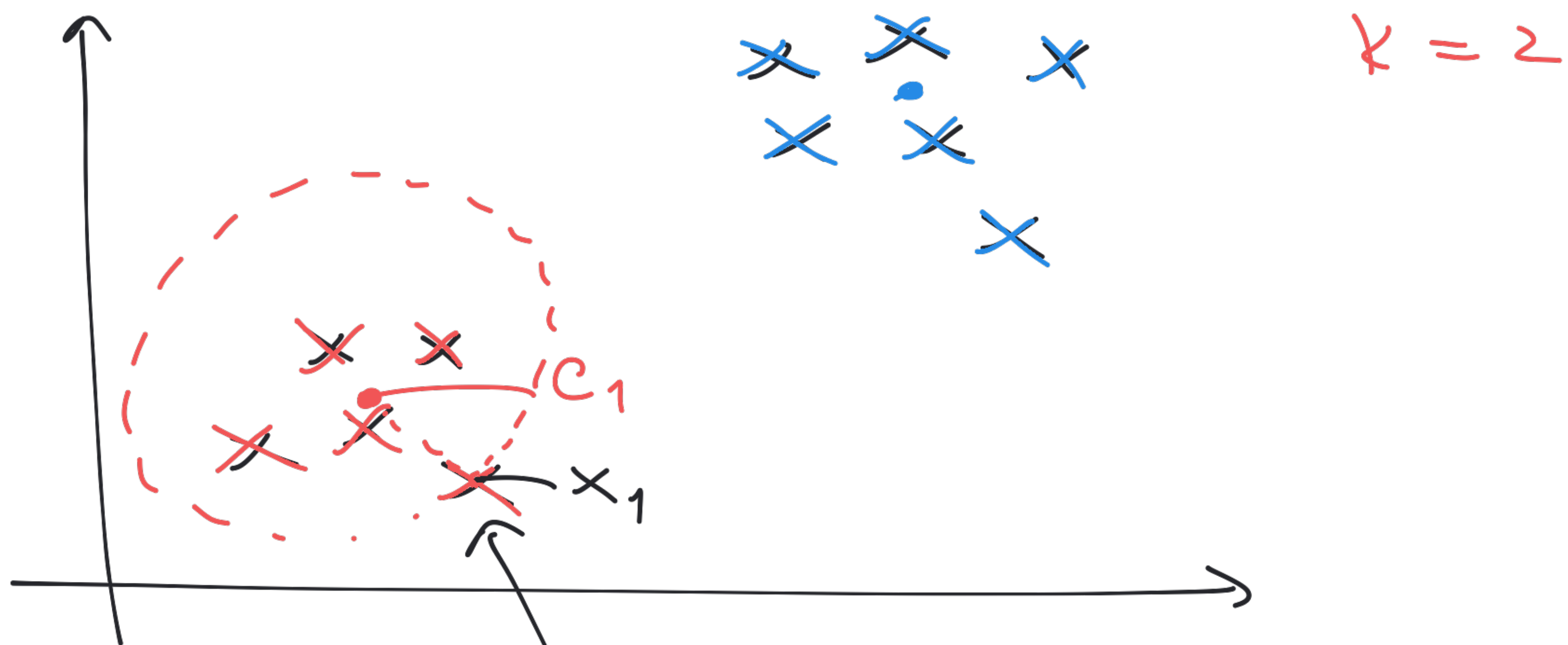
If $x_1 \in C_2$

If $x_2 \in C_K$

$$\sum_{i=1}^{N} u_{ik} \equiv N_k$$

$N_k \equiv$ # of samples in cluster k

If $d^2(x_i, \theta_k) = \|x_i - \theta_k\|^2$



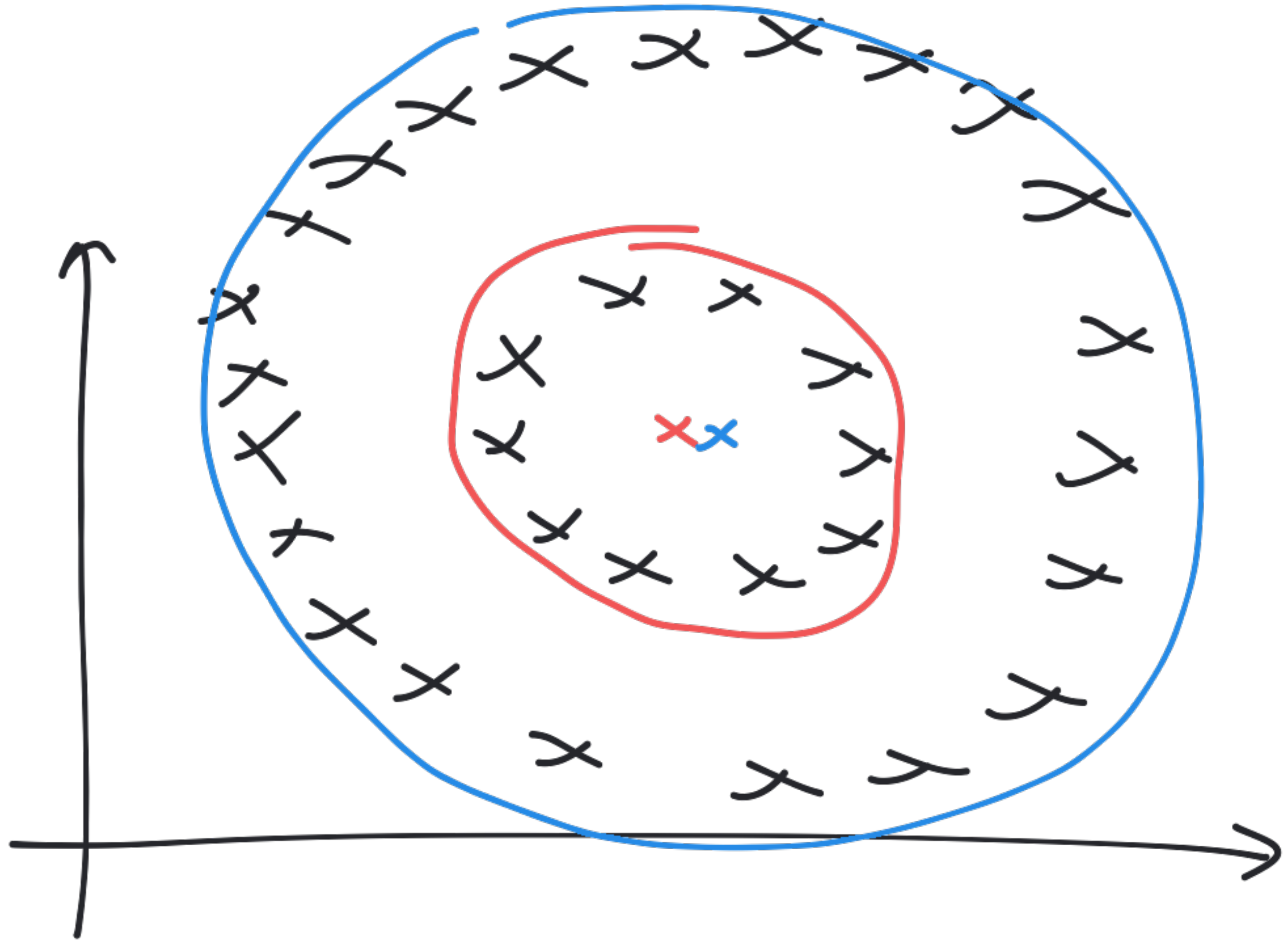
$d^2(x_1, c_1) = 2$

this point is the furthest point for centroid assign to this cluster

If distance is Euclidean then the clusters will be spherical.

We do not have samples.

Yet, K-means will cluster any point in that region as belonging to group Red.

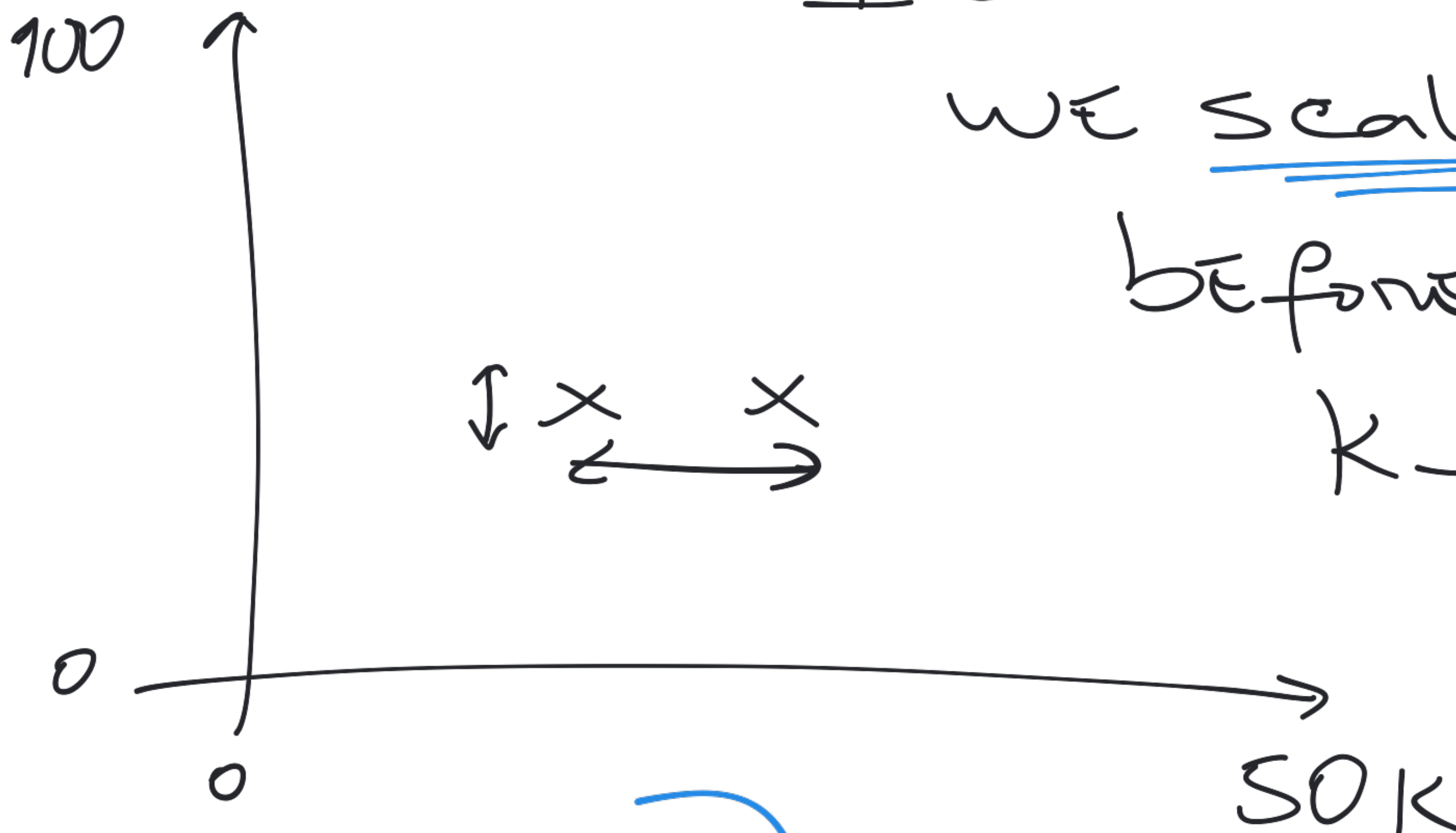


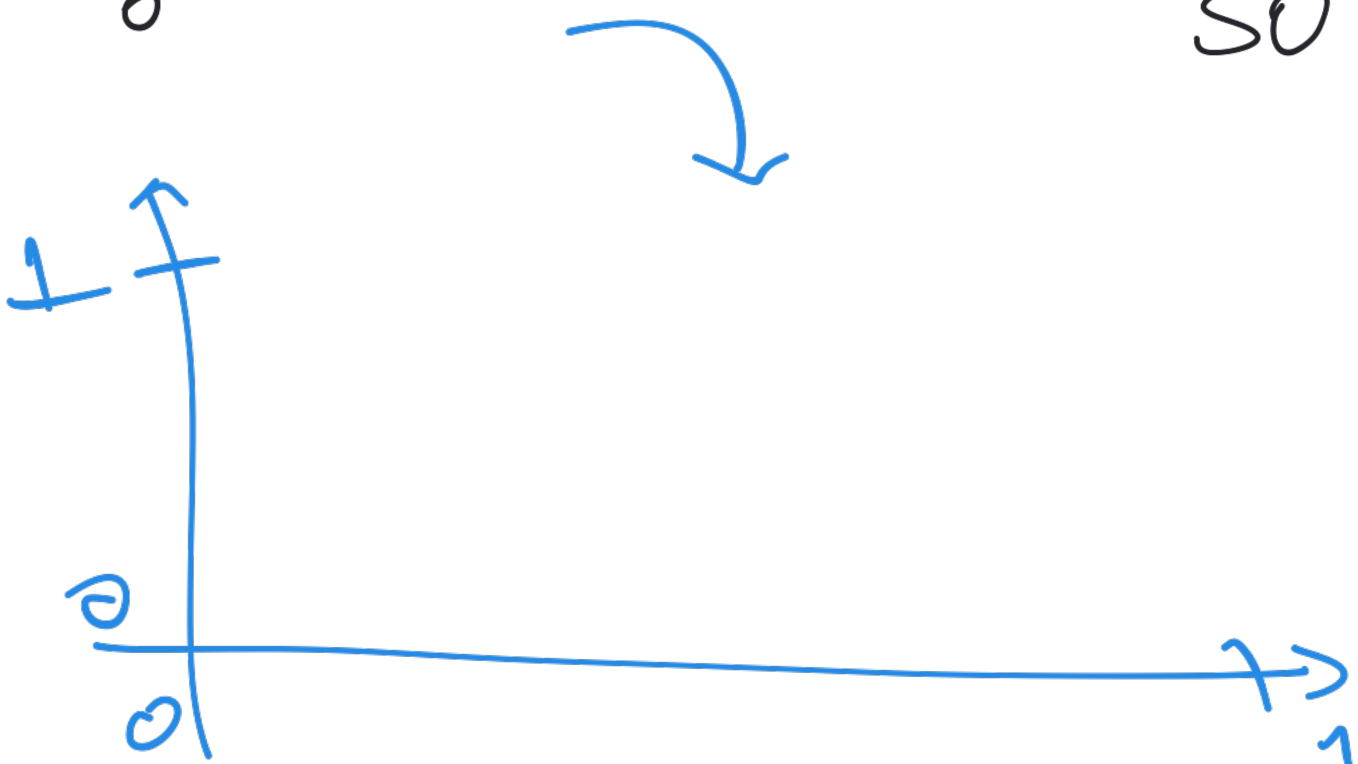
In practice we will have high-dimensions and so we will not be able to plot.

So we need to define *quantitative* metrics to validate the clustering results.

$$J = \sum_i \sum_k u_{ik} \, d^2(x_i, \theta_k)$$

It is crucial that we scale the data before applying K-means.





① Find Membership Matrix:

$$\frac{\partial J}{\partial u_{ik}} = 0 \Longrightarrow$$ Assign point x_i to the cluster with closest centroid ϑ_k

② $$\frac{\partial J}{\partial \vartheta_k} = 0 \iff \sum_{i=1}^{N} u_{ik} \cdot \underbrace{2 \cdot (-1)}\cdot (x_i - \vartheta_k) = 0$$

$$\iff \sum_{i=1}^{N} u_{ik} (x_i - \vartheta_k) = 0$$

$$\iff \sum_{i=1}^{N} u_{ik} \cdot x_i - \sum_{i=1}^{N} u_{ik} \cdot \vartheta_k = 0$$

NEXT page →

$$\underbrace{\sum_{i=1}^{N} u_{ik} x_i}_{(*)} = \theta_k \underbrace{\sum_{i=1}^{N} u_{ik}}_{N_k}$$

$$\Leftrightarrow \sum_{n \in C_k} x_i = \theta_k \cdot N_k$$

$$\Leftrightarrow \boxed{\theta_k = \frac{1}{N_k} \cdot \sum_{n \in C_k} x_i}$$

↳ average for samples belonging to cluster k.

$$U = \begin{array}{c} \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \dots & k \end{array} \\ \left[\begin{array}{cccc} 1 & & & \\ 0 & & & \\ \vdots & & & \\ & & & \end{array} \right] \end{array}$$

N_k = # of points assigned to cluster k

$$(*) \sum_{i=1}^{N} u_{ik} x_i = 0 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 + \dots$$

$$= \sum_{n \in C_k} x_n$$