# Practical Machine Learning

**Lecture 3**

**K-means model and Gaussian Mixture Model (GMM)**

Dr. Suyong Eum

OSAKA UNIVERSITY

# A question from the last class

$$\mathrm{MSE}(\mathrm{w}) = \frac{1}{2}\sum_{n=1}^{N}(\mathrm{w}^{\mathrm{T}}\mathrm{x}_n - \mathrm{y}_{\mathrm{n}})^2$$

$$\frac{\mathrm{d}}{\mathrm{dw}}\mathrm{MSE}(\mathrm{w}) = \sum_{n=1}^{N}(\mathrm{w}^{\mathrm{T}}\mathrm{x}_n - \mathrm{y}_n)\cdot\mathrm{x}_{\mathrm{n}}^{\mathrm{T}} = 0$$

$$= \sum_{n=1}^{N}(\mathrm{w}^{\mathrm{T}}\mathrm{x}_{\mathrm{n}}^{\mathrm{T}}\mathrm{x}_{\mathrm{n}} - \mathrm{x}_{\mathrm{n}}^{\mathrm{T}}\mathrm{y}_{\mathrm{n}}) = 0$$

$$\mathrm{w}^{\mathrm{T}}\mathrm{x}_{\mathrm{n}}^{\mathrm{T}}\mathrm{x}_{\mathrm{n}} = \mathrm{x}_{\mathrm{n}}^{\mathrm{T}}\mathrm{y}_{\mathrm{n}}$$

$$\boxed{\mathrm{w}^{\mathrm{T}} = \left(\mathrm{x}_{\mathrm{n}}^{\mathrm{T}}\mathrm{x}_{\mathrm{n}}\right)^{-1}\mathrm{x}_{\mathrm{n}}^{\mathrm{T}}\mathrm{y}_{\mathrm{n}}}$$

(mx1) [(mxn)(nxm)]$^{-1}$(mxn)(nx1)

$$\mathrm{MSE}(\mathrm{w}) = \frac{1}{2}(\mathrm{XW} - \mathrm{Y})^{\mathrm{T}}(\mathrm{XW} - \mathrm{Y})$$

(1x1) = [(nxm)(mx1)-(nx1)]$^{-T}$ [(nxm)(mx1)-(nx1)]

$$= \frac{1}{2}((\mathrm{XW})^{\mathrm{T}} - \mathrm{Y}^{\mathrm{T}})(\mathrm{XW} - \mathrm{Y})$$

$$= \frac{1}{2}(\mathrm{W}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}} - \mathrm{Y}^{\mathrm{T}})(\mathrm{XW} - \mathrm{Y})$$

$$= \frac{1}{2}(\mathrm{W}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{XW} - \mathrm{Y}^{\mathrm{T}}\mathrm{XW} - \mathrm{W}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{Y} + \mathrm{Y}^{\mathrm{T}}\mathrm{Y})$$

$$\boxed{\mathrm{Y}^{\mathrm{T}}\mathrm{XW} = (\mathrm{W}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{Y})^{\mathrm{T}}}$$

$$= \frac{1}{2}(\mathrm{W}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{XW} - 2\mathrm{W}^{\mathrm{T}}\mathrm{X}^{\mathrm{T}}\mathrm{Y} + \mathrm{Y}^{\mathrm{T}}\mathrm{Y})$$

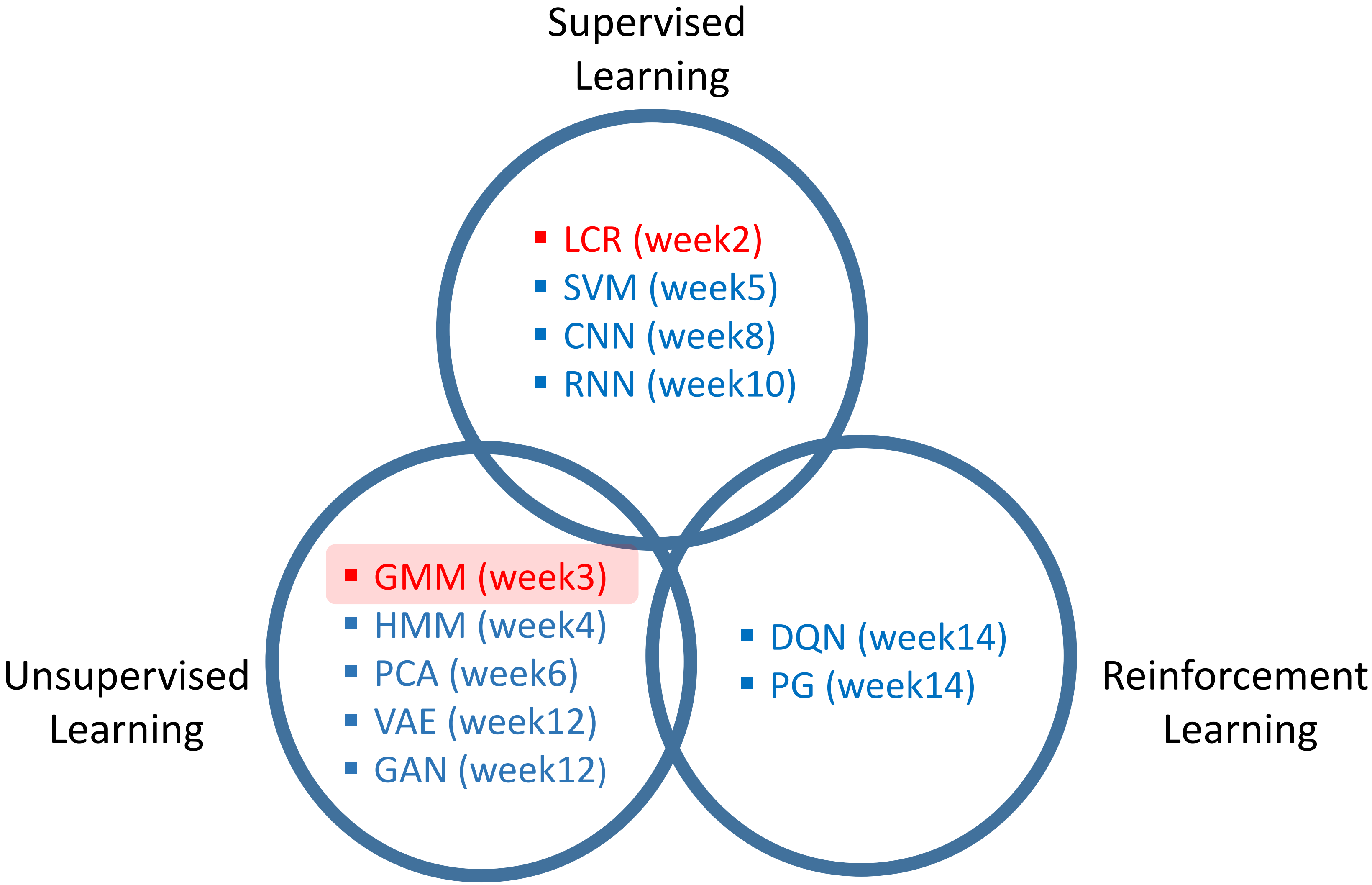
$$\frac{dE}{d\mathrm{W}^{\mathrm{T}}} = \frac{1}{2}(2\mathrm{X}^{\mathrm{T}}\mathrm{XW} - 2\mathrm{X}^{\mathrm{T}}\mathrm{Y}) = 0$$

$$\mathrm{X}^{\mathrm{T}}\mathrm{XW} = \mathrm{X}^{\mathrm{T}}\mathrm{Y} \qquad \boxed{\mathrm{W} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}}$$

(mx1) [(mxn)(nxm)]$^{-1}$(mxn)(nx1)

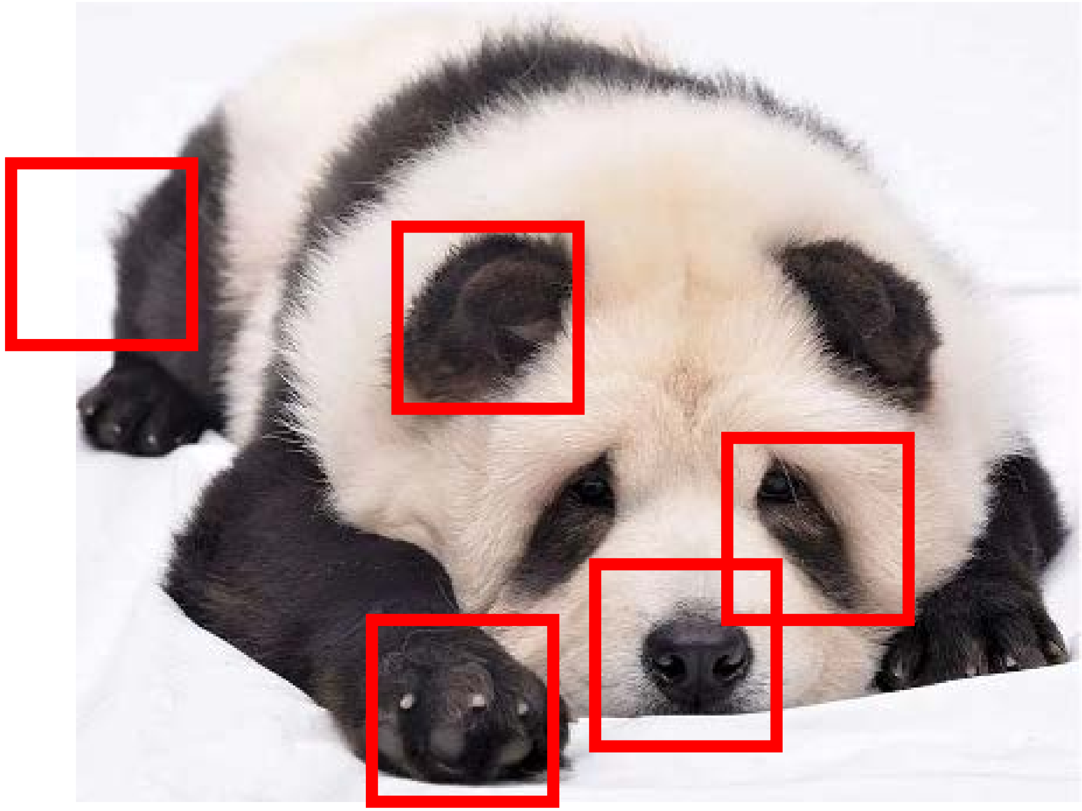
https://eli.thegreenplace.net/2014/derivation-of-the-normal-equation-for-linear-regression

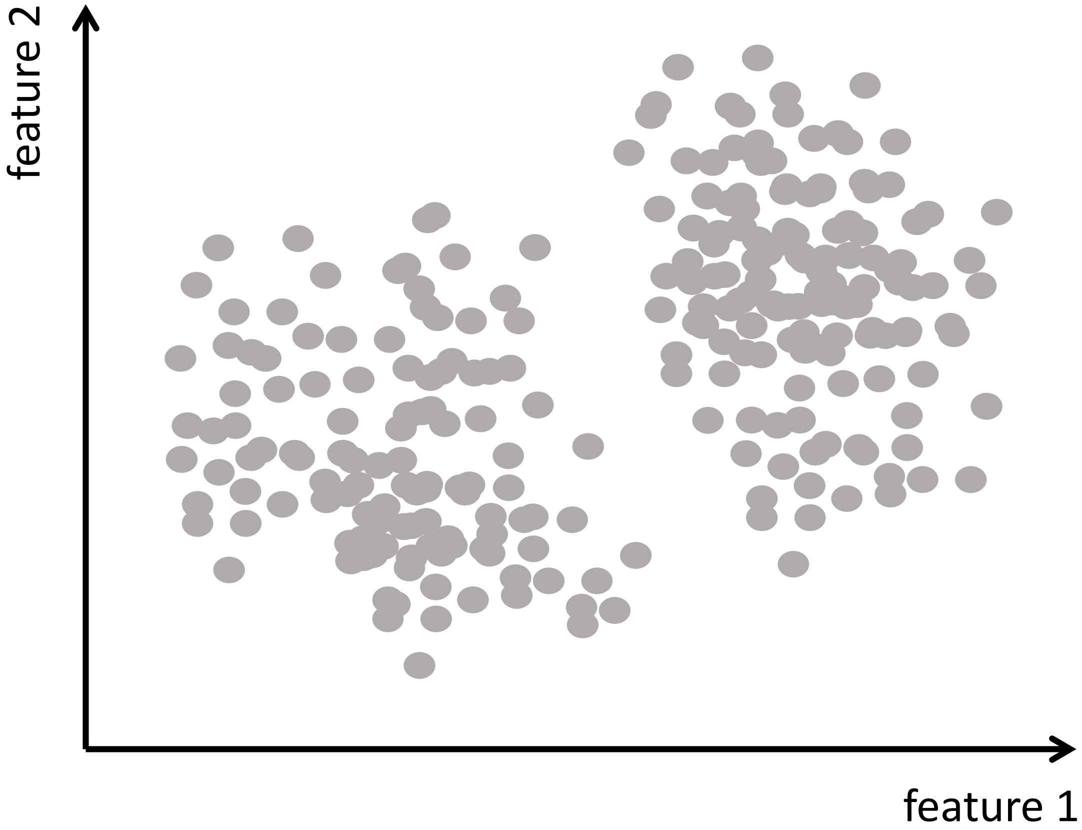
# Where we are


Supervised Learning
LCR (week2)
SVM (week5)
CNN (week8)
RNN (week10)
Unsupervised Learning
GMM (week3)
HMM (week4)
PCA (week6)
VAE (week12)
GAN (week12)
DQN (week14)
PG (week14)
Reinforcement Learning

# Unsupervised learning: clustering

- Clustering is the most fundamental learning mechanism.
- What makes you think the below is a dog not a panda?





# Lecture Outline

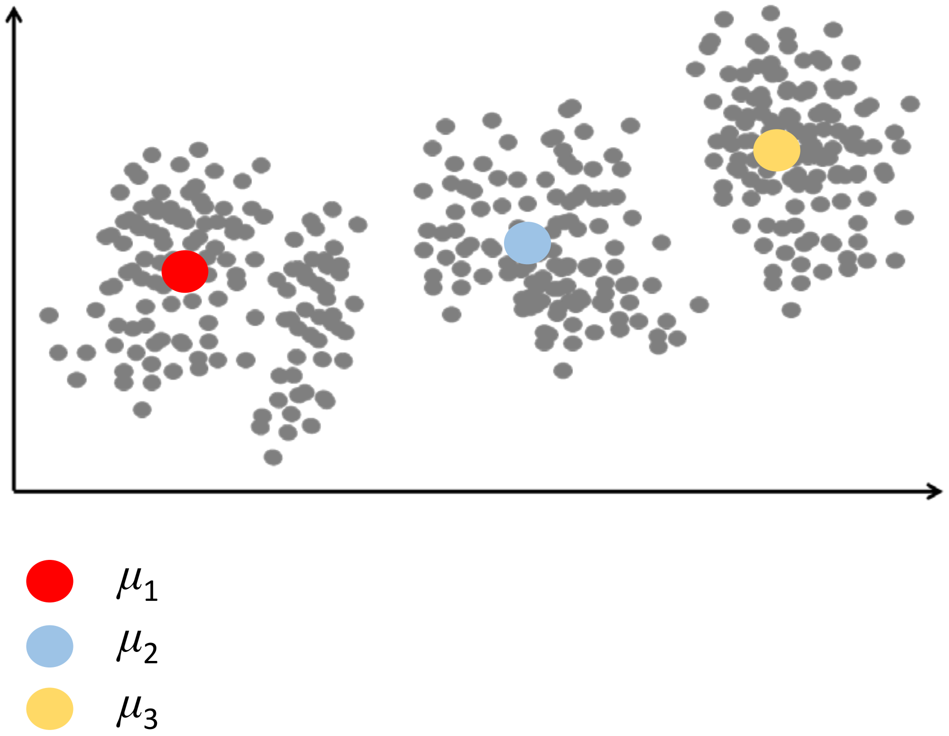
- K-means model
- Gaussian Mixture Model (GMM)
- Expectation and Maximization (EM) for GMM
- An example of EM operation
- Graphical representation of GMM

# K-means model

# K-means model

- ❑ Problem of identifying clusters of data points by minimizing the function *J*
- ❑ Clustering the data points into K clusters: “assuming that K is known”

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\left\| x_n - \mu_k \right\|^2$$




- N: the number of observed data points
- K: the number of clusters
- $x_n$: $n^{th}$ data point
- $\mu_k$: $k^{th}$ centroid corresponding to each cluster
- $r_{nk}$: {0, 1} showing whether a data point belongs to “k cluster” or not

# K-means model: which value the centroid should be?

- To minimize the error function *J*, which value the centroid should be?
- If *J* has 1L norm, then?

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \left\| x_n - \mu_k \right\|^2$$

$$\frac{dJ}{d\mu_k} = 2\sum_{n=1}^{N} r_{nk}(x_n - \mu_k) = 0$$

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}$$

Summation of x values in cluster k

Number of data points in cluster k

$\mu_k$ : Mean of the data points $x_k$ in cluster k

# K-means clustering: how to optimize the equation?

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \left\| x_n - \mu_k \right\|^2$$

**Expectation Step**

Expect which data points are close to each centroid

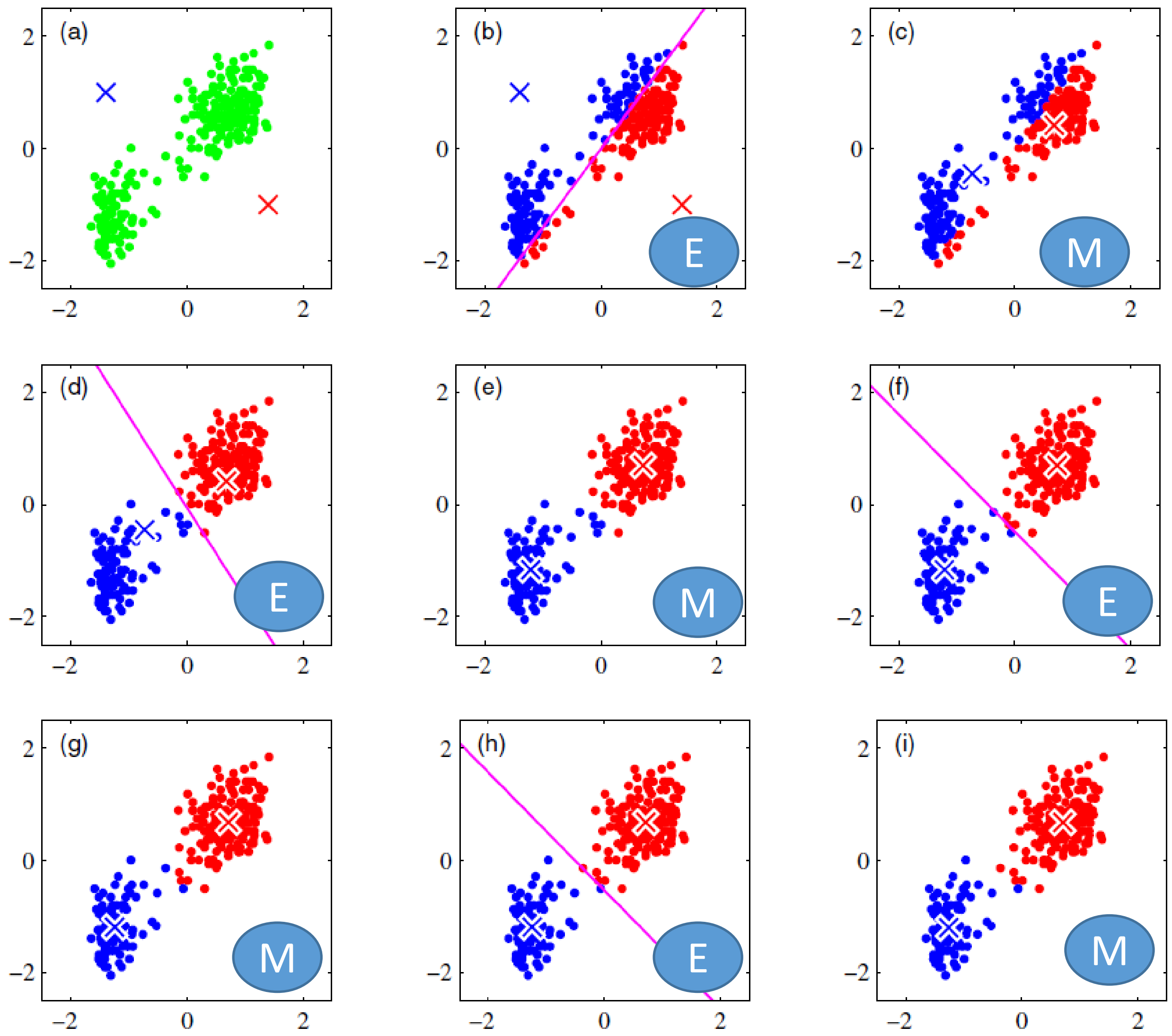
$r_{nk}$

**Maximization Step**

Finding a centroid for each cluster

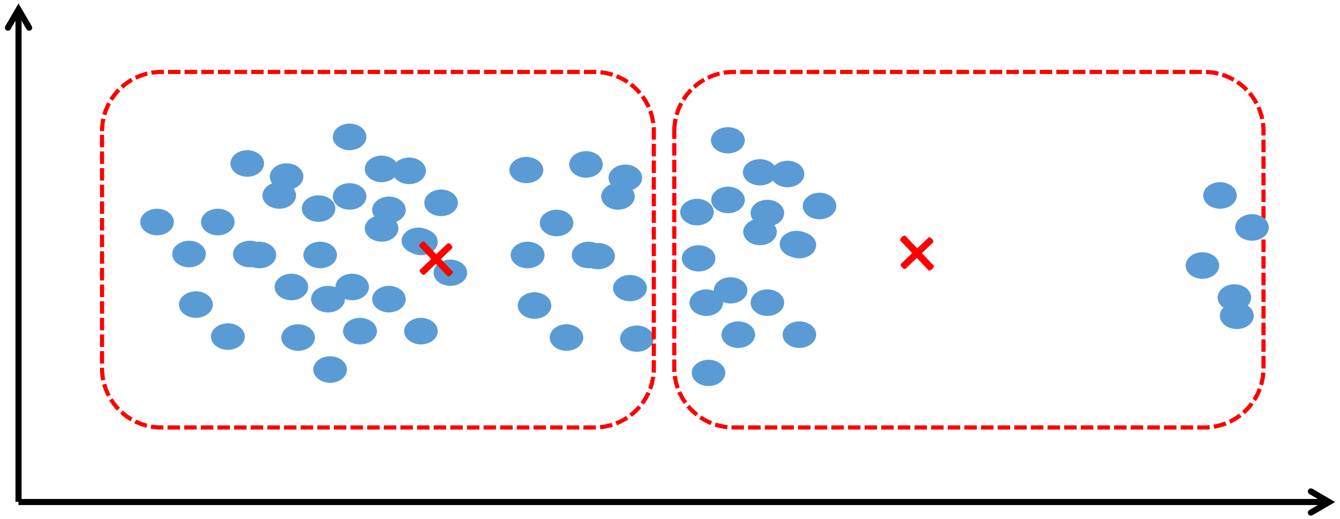
$\mu_k$

Random choice of centroids

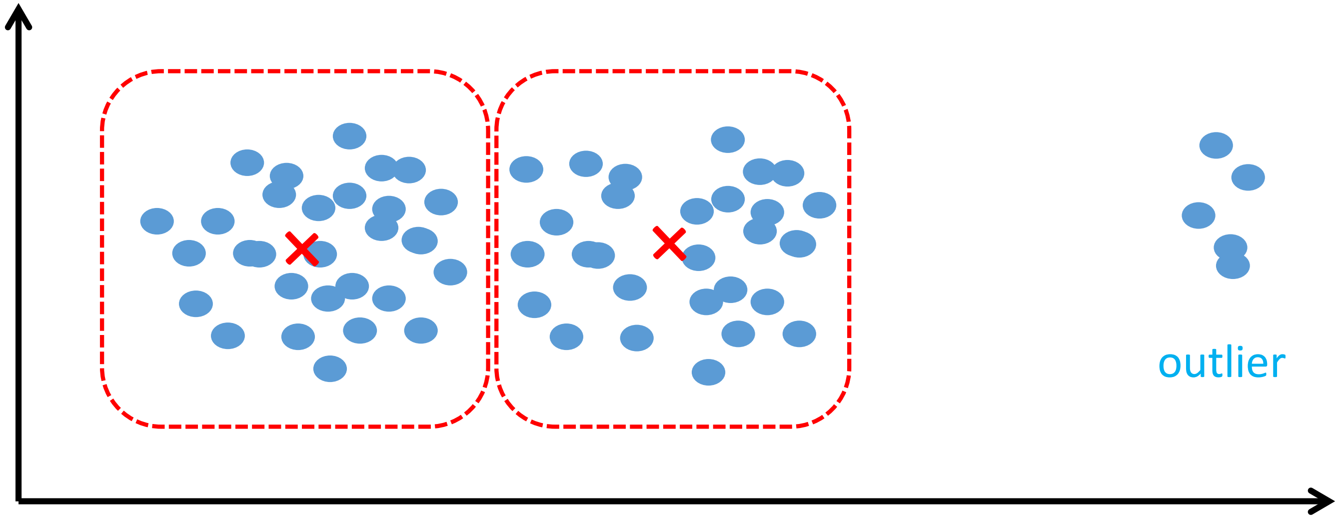


# Problems of K means: outlier or unevenly sized clusters

Undesirable clustering

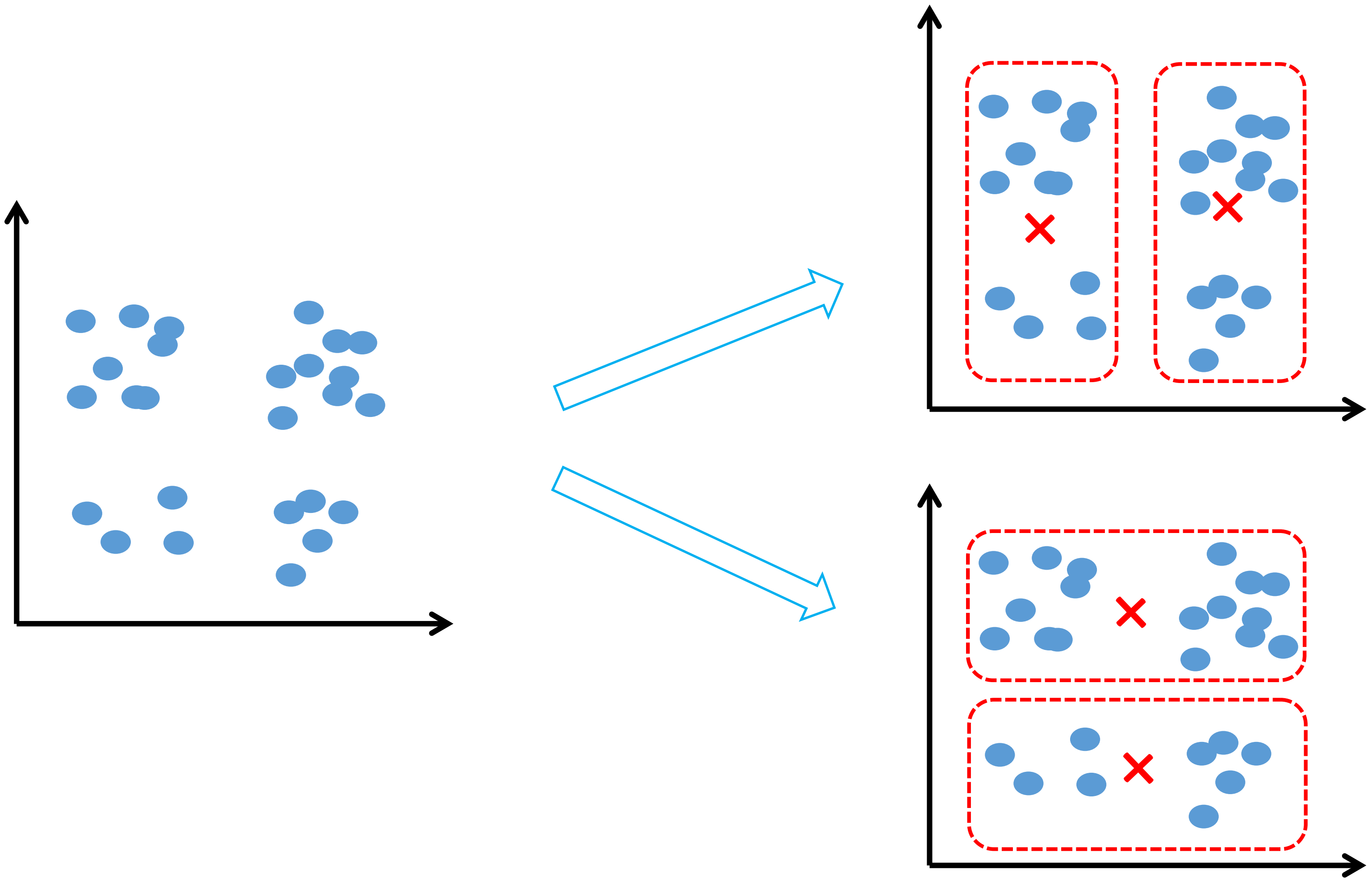


Desirable clustering



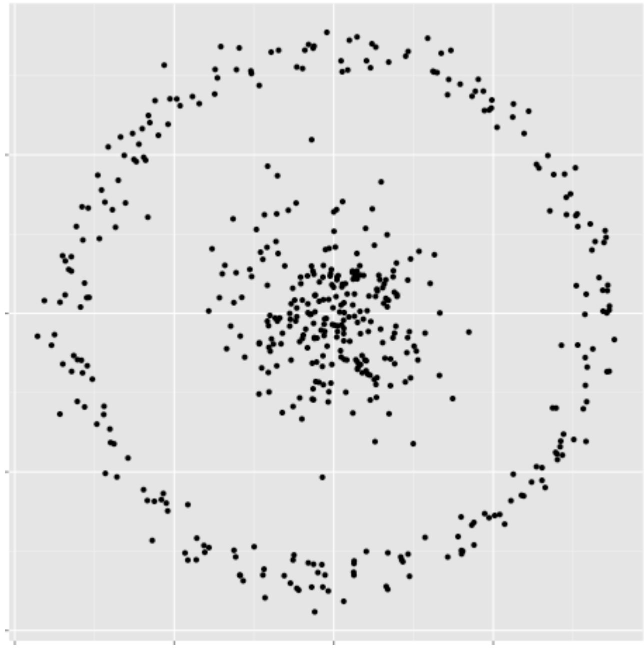
# Problems of K means: Initialization issue

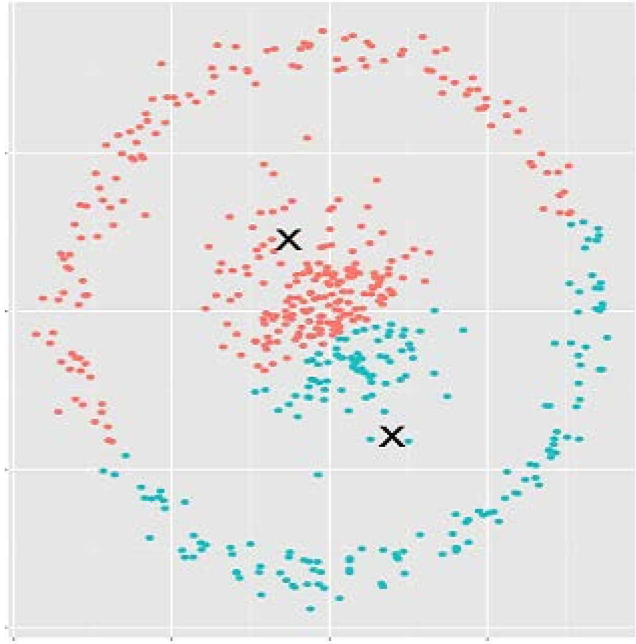
- Depending on the initialization, clustering results can be changed



# Problems of K means: Non-spherical data issue

- K means algorithm assumes that clustered data set has a shape of sphere.



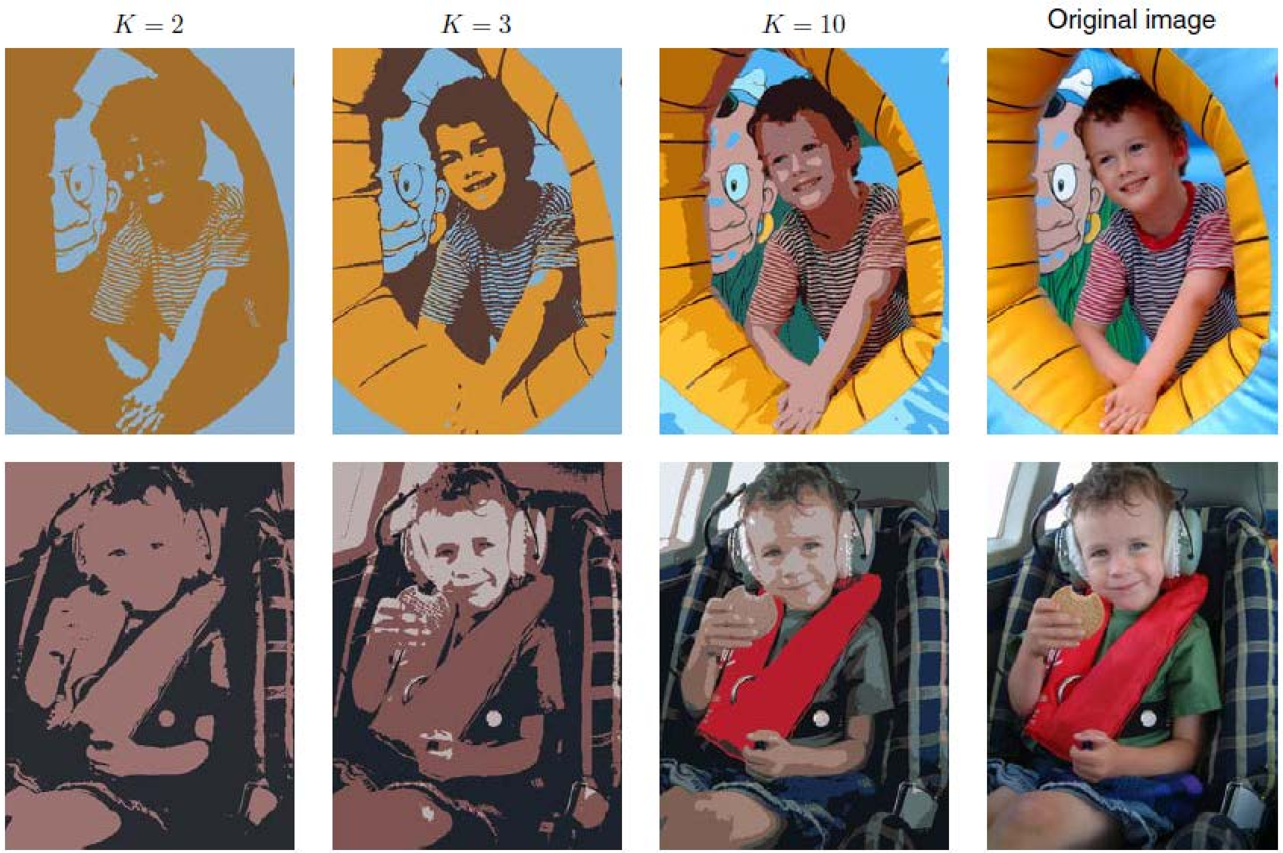


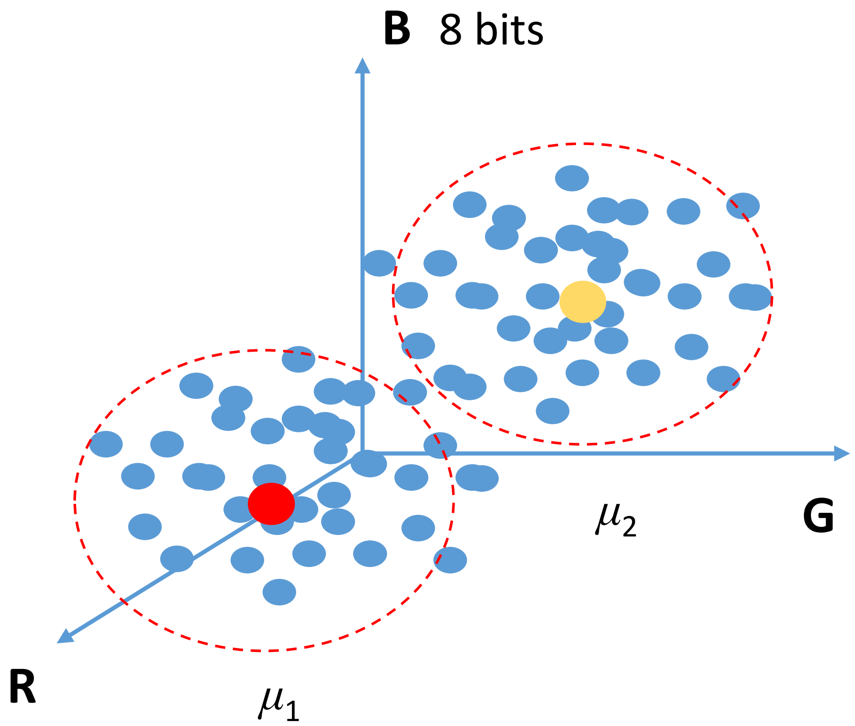


http://varianceexplained.org/r/kmeans-free-lunch/

# An application of K-means algorithm

- Image segmentation and compression




- Original: 24 bits per pixel
- K clustering: Log2 K bits per pixel

# Gaussian Mixture Model (GMM)

# Prerequisite items you need to know before GMM

- Likelihood function
- Maximum likelihood estimation
- Multivariate Gaussian distribution

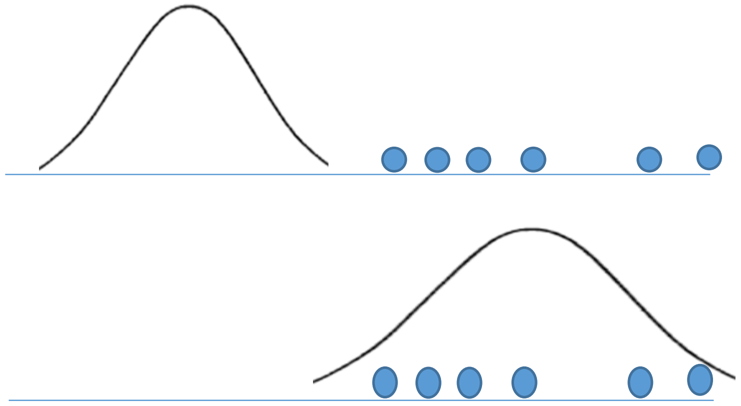
# Likelihood function

- A likelihood function is a probability mass or density function having parameter(s).
- We often take log both sides of the likelihood function and call it log-likelihood function.
- Given a set of data, the parameter(s) of the probability model is estimated by maximizing the log-likelihood function, which is called a maximum likelihood estimation.

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$L(\mathrm{X} \mid \mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln L(\mathrm{X} \mid \mu, \sigma^2) = \sum_{i=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$



# Maximum likelihood estimation

- Maximum likelihood estimation is a procedure that finds the parameter(s) of the probability model by maximizing the (log)-likelihood function.
- Some cases are easy to obtain an analytical solution. However, some cases are not.

$$\ln L(\mathrm{X} \mid \mu, \sigma^2) = \sum_{i=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln L(\mathrm{X} \mid \mu, \sigma^2) = \sum_{i=1}^{N} \left[ \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x_i-\mu)^2}{2\sigma^2} \right]$$

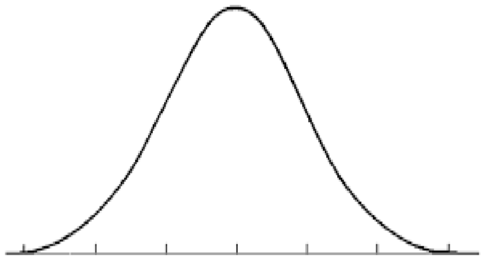
$$\underset{\mu,\sigma^2}{\arg\max} \ln L(\mathrm{X} \mid \mu, \sigma^2)$$

$$\frac{d}{d\mu} \ln L(\mathrm{X} \mid \mu, \sigma^2) = \sum_{i=1}^{N} \left[ \frac{(x_i-\mu)}{\sigma^2} \right] = 0$$

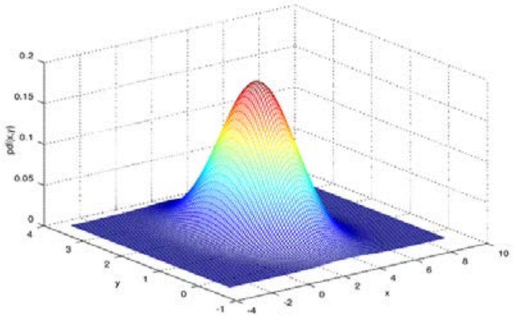
$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

# Multivariate Gaussian distribution

- A generalization of one-dimensional Gaussian distribution to higher dimensions
- Two parameters: mean (μ) and covariance (∑)
- Notation: N(μ,∑)

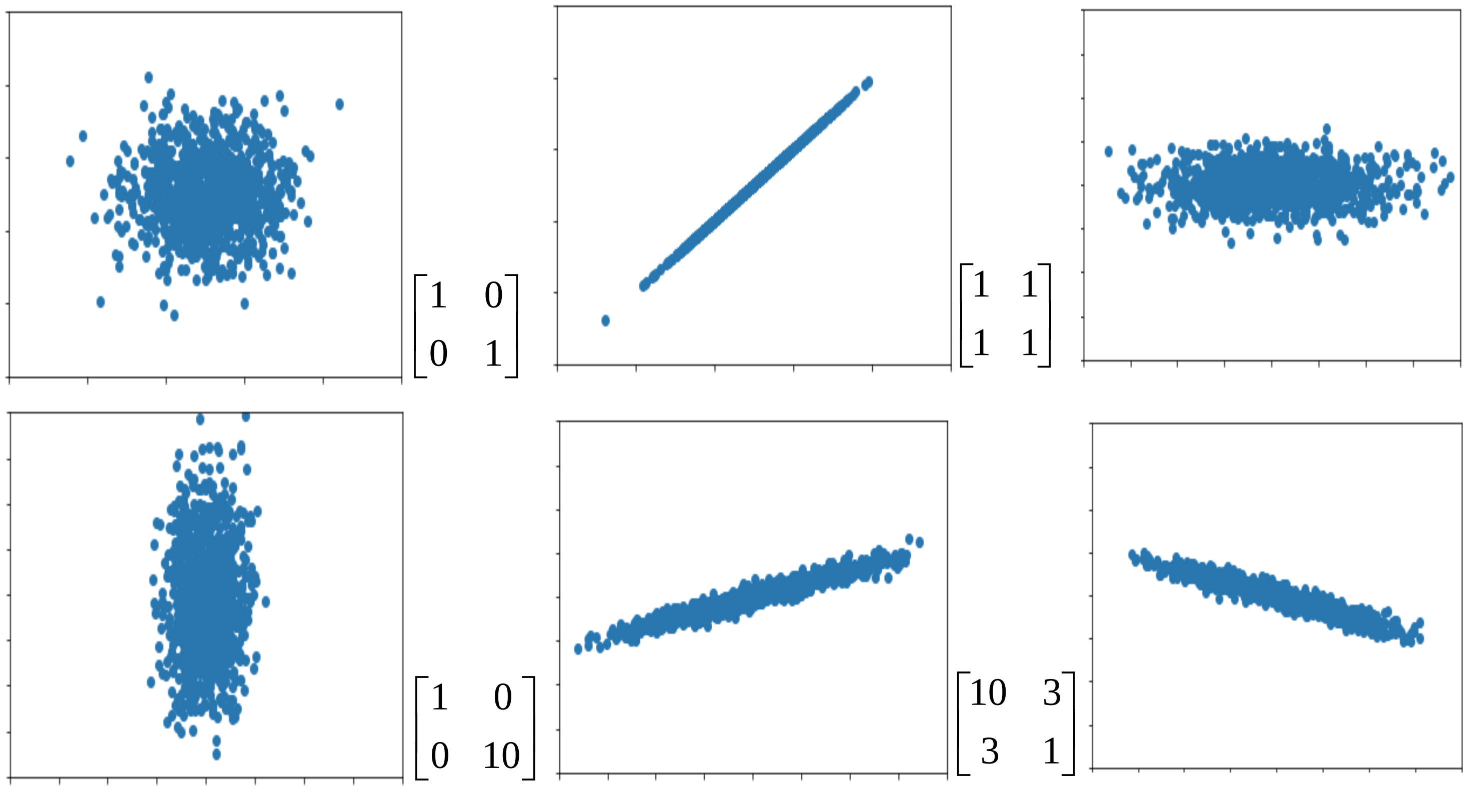


single variable Gaussian



2-variables Gaussian

isotropy



$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ $\begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$

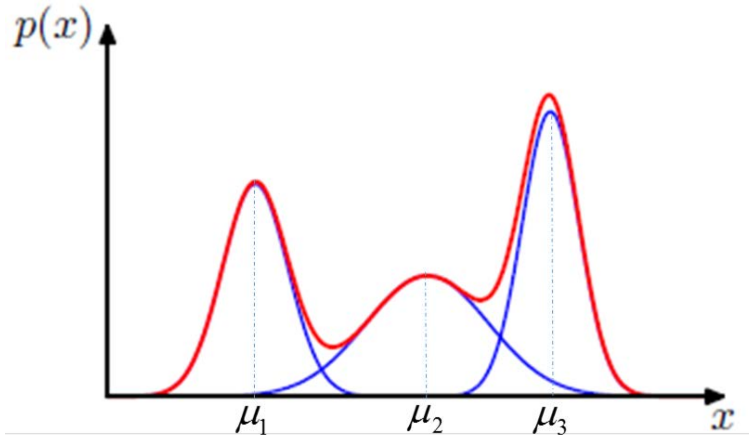
$\begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}$ $\begin{bmatrix} 10 & 3 \\ 3 & 1 \end{bmatrix}$ $\begin{bmatrix} 10 & -3 \\ -3 & 1 \end{bmatrix}$

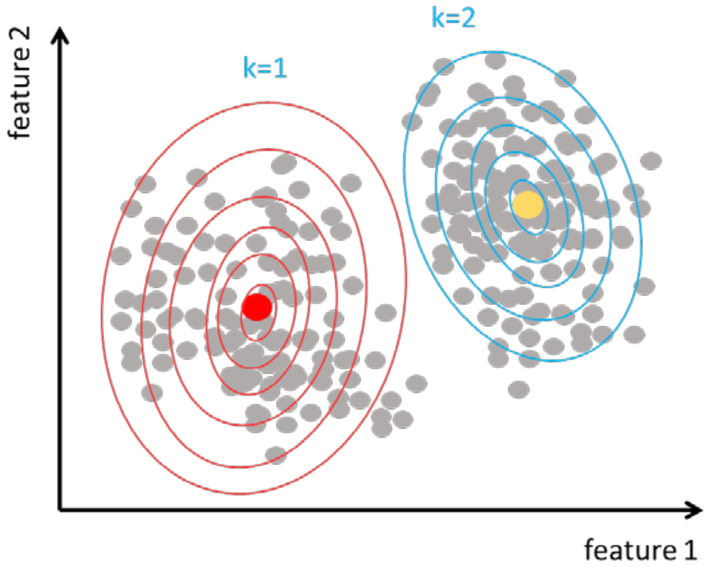
Covariance

# Gaussian Mixture Models (GMM)

- A probability model that multivariate Gaussian distributions are mixed or linearly superposed.

$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k N(\mathrm{x} \mid \mu_k, \Sigma_k)$$




- $\pi_k$ : mixing coefficient - probability that $k^{th}$ multivariate Gaussian being selected
- $\mu_k$ : mean of $k^{th}$ multivariate Gaussian
- $\Sigma_k$ : covariance of $k^{th}$ multivariate Gaussian

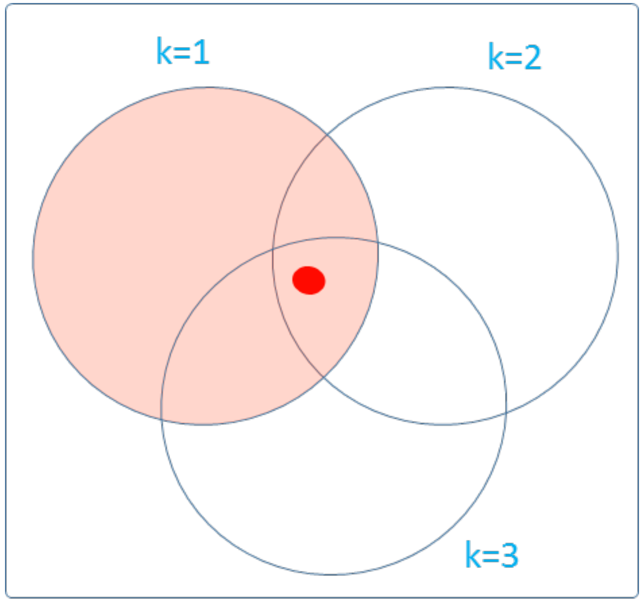
# Gaussian Mixture Models (GMM): a hidden or a latent variable

- GMM has a hidden or a latent variable in the model.
- It is denoted as "z", which has K-dimensional binary random variable having 1-of-K representation.
- The latent variable shows which cluster is active, which is governed by the mixing coefficient $\pi_k$

$$\mathrm{z} = (z_1, z_2, \ldots, z_k) \qquad z_k \in \{0,1\}$$

$$p(z_k = 1) = \pi_k$$

Probability that $k^{th}$ Gaussian is active.




$z=(z_1, z_2, z_3)=(1,0,0)$

# Gaussian Mixture Models (GMM): how to find all parameters of GMM?

- We can find all parameters of GMM using maximum likelihood estimation
- Log-likelihood function of GMM is given as follows:

$$L(\mathrm{X} \mid \pi, \mu, \Sigma) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k N(\mathrm{x}_n \mid \mu_k, \Sigma_k)$$

: Likelihood function

: N times of the GMM probabilities

$$\arg\max_{\pi,\mu,\Sigma} \ln L(\mathrm{X} \mid \pi, \mu, \Sigma)$$

$$\ln L(\mathrm{X} \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(\mathrm{x}_n \mid \mu_k, \Sigma_k) \right\}$$

: log likelihood function

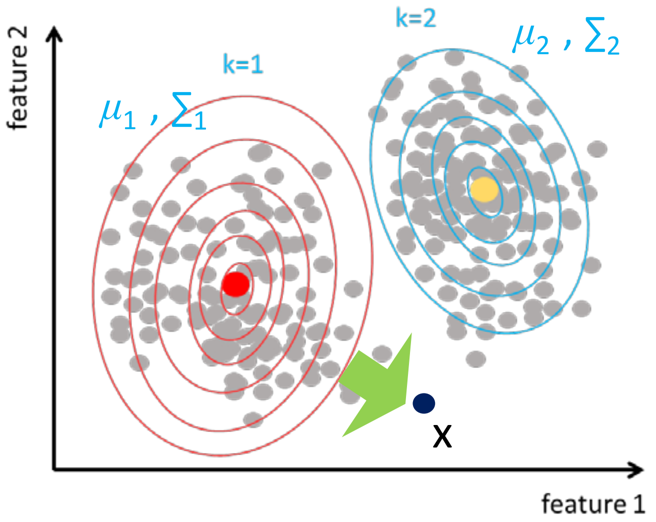
- There is not any analytical solution for this maximization problem. So,
  - Neural network approach: using the negative log likelihood function as an error function
  - Expectation Maximization (EM) approach

# Expectation and Maximization (EM) for GMM

# Gaussian Mixture Models (GMM): responsibility $\gamma(z_k)$

- Different from K-means algorithm, GMM model tells the probabilities that a given data point belongs to individual classes.
  - The probability is called "responsibility", which is denoted "$\gamma(z_k)$"
  - The probability is also called "posterior", which is denoted "$p(z_k=1|x)$"

$$\gamma(z_k) = \frac{\pi_k N(\mathrm{x}\,|\,\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(\mathrm{x}\,|\,\mu_j,\Sigma_j)}$$




$$p(\mathrm{x},\mathrm{z}) = p(\mathrm{z},\mathrm{x})$$
$$p(\mathrm{x})\,p(\mathrm{z}\,|\,\mathrm{x}) = p(\mathrm{z})\,p(\mathrm{x}\,|\,\mathrm{z})$$
$$p(\mathrm{z}\,|\,\mathrm{x}) = \frac{p(\mathrm{z})\,p(\mathrm{x}\,|\,\mathrm{z})}{p(\mathrm{x})} = \frac{\pi_k N(\mathrm{x}\,|\,\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j N(\mathrm{x}\,|\,\mu_j,\Sigma_j)}$$

$$\gamma(z_1) = \frac{\pi_1 N(\mathrm{x}\,|\,\mu_1,\Sigma_1)}{\pi_1 N(\mathrm{x}\,|\,\mu_1,\Sigma_1) + \pi_2 N(\mathrm{x}\,|\,\mu_2,\Sigma_2)}$$

# Gaussian Mixture Models (GMM): three parameters of GMM model

- Well, this part normally involves slightly(?) heavy mathematical derivation.
- An idea is that you can find the parameters 1) $\pi_k$ , 2) $\mu_k$, 3) $\Sigma_k$ when responsibility $\gamma(z_k)$ is given.
- And vice versa !

$$\ln L(\mathrm{X} \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(\mathrm{x}_n \mid \mu_k, \Sigma_k) \right\}$$

: log likelihood function

$$N(\mathrm{x} \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_k|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathrm{x}-\mu_k) \Sigma_k^{-1} (\mathrm{x}-\mu_k) \right\}$$

1 $\frac{d}{d\pi_k} \ln L(\mathrm{X} \mid \pi, \mu, \Sigma) = 0$ → (Mixing coefficient; Lagrange method) → $\pi_k = \frac{N_k}{N}$

- Mixing coefficient
- Lagrange method

2 $\frac{d}{d\mu_k} \ln L(\mathrm{X} \mid \pi, \mu, \Sigma) = 0$ → (Mean of data in a class k) → $\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathrm{x}_n$

3 $\frac{d}{d\Sigma_k} \ln L(\mathrm{X} \mid \pi, \mu, \Sigma) = 0$ → (Covariance of data in a class k) → $\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathrm{x}_n - \mu_k)(\mathrm{x}_n - \mu_k)^T$

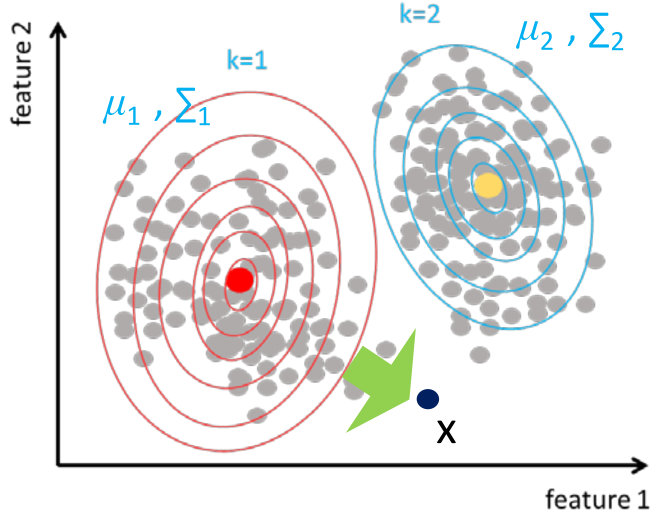
$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

# Gaussian Mixture Models (GMM): E-step

- Three parameters of GMM model is from M-step (or randomly initialized in the first iteration).
  - 1) $\pi_k$ , 2) $\mu_k$, 3) $\Sigma_k$

- Expect the responsibility "$\gamma(z_k)$"

$$\gamma(z_k) = \frac{\pi_k N(\mathrm{x} \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(\mathrm{x} \mid \mu_j, \Sigma_j)}$$




$$\gamma(z_1) = \frac{\pi_1 N(\mathrm{x} \mid \mu_1, \Sigma_1)}{\pi_1 N(\mathrm{x} \mid \mu_1, \Sigma_1) + \pi_2 N(\mathrm{x} \mid \mu_2, \Sigma_2)}$$

$$\gamma(z_2) = \frac{\pi_2 N(\mathrm{x} \mid \mu_2, \Sigma_2)}{\pi_1 N(\mathrm{x} \mid \mu_1, \Sigma_1) + \pi_2 N(\mathrm{x} \mid \mu_2, \Sigma_2)}$$

# Gaussian Mixture Models (GMM): M-step

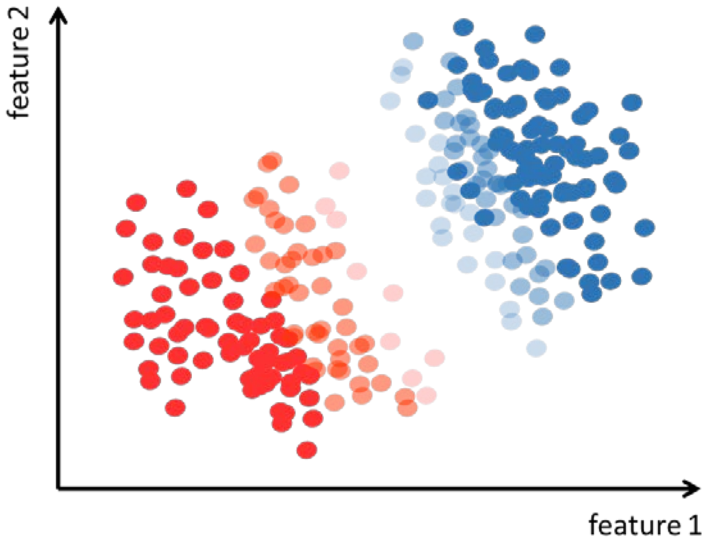
- The responsibility "$\gamma(z_k)$" is from E-step.
- Three parameters of GMM model is calculated using the equations below:

1. $$\pi_k = \frac{N_k}{N}$$

2. $$\mu_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathrm{x}_n$$

3. $$\Sigma_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathrm{x}_n - \mu_k)(\mathrm{x}_n - \mu_k)^T$$

$$N_k = \sum_{n=1}^{N}\gamma(z_{nk})$$




| | $x_1$ | $x_2$ | ... | $x_n$ | |
|---|---|---|---|---|---|
| Cluster (k=1) | $\gamma(z_{11})$ | $\gamma(z_{21})$ | ... | $\gamma(z_{n1})$ | $N_1$ |
| Cluster (k=2) | $\gamma(z_{12})$ | $\gamma(z_{22})$ | ... | $\gamma(z_{n2})$ | $N_2$ |

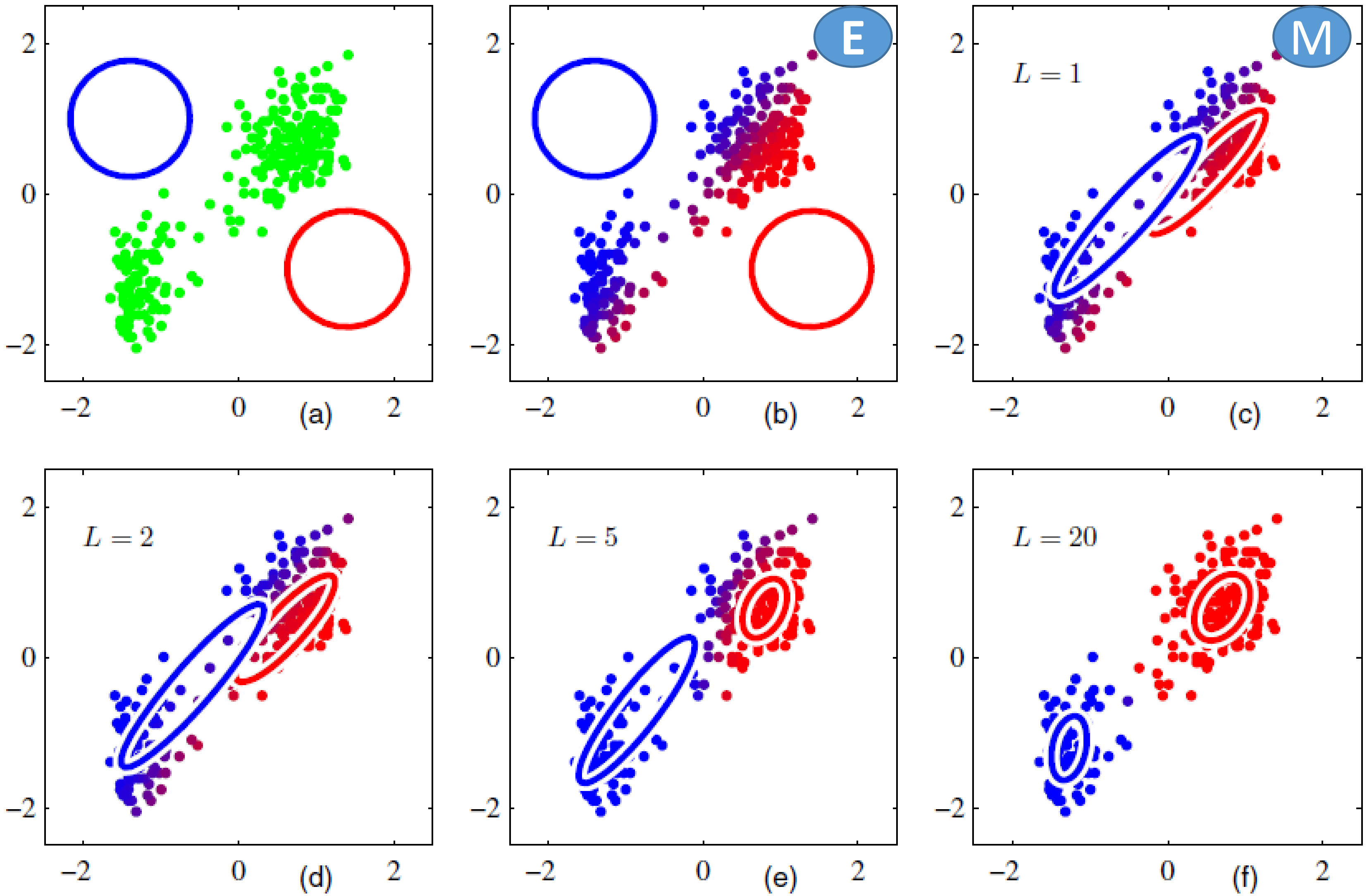
# Gaussian Mixture Models (GMM): operation

1) $\pi_k$,
2) $\mu_k$,
3) $\Sigma_k$

**Initial random setup**

Responsibility Expectation

Three parameters calculation




**Iteration = 2** **Iteration = 5** **Iteration = 20**

# Gaussian Mixture Models (GMM) vs K-means

| K-means | GMM |
| --- | --- |
| ❑ Hard clustering: {Yes or No } | ❑ Soft clustering: {Probability} |
| ❑ Centroid ($\mu_k$ ) | ❑ Mean and Covariance ($\mu_k$, $\Sigma_k$) |
| ❑ $r_{nk}$: {0,1} | ❑ Mixing coefficient ($\pi_k$ ): probability |
| ❑ Reducing the distance | ❑ Maximizing log likelihood function |
| ❑ Simple and Fast | ❑ Complex and Slow |

- ❑ Therefore, common to run the *K*-means algorithm in order to find a suitable initialization for a Gaussian mixture model that is subsequently adapted using EM.
- ❑ “K” needs to be decided.

# An example of EM operation

# EM algorithm: an example – smoking and cancer

- Your task is to find out a group of people over 70 who has high risks of cancer.
- Your initial belief is that
  - 70% of cancer patients are a smoker.
  - 30% of non-cancer patients are a smoker.
- Then, a survey is carried out to five groups of people as follows:

| | smoker | Non-smoker |
|---|---|---|
| Group1 | 6 | 4 |
| Group2 | 7 | 3 |
| Group3 | 5 | 5 |
| Group4 | 9 | 1 |
| Group5 | 8 | 2 |

# EM algorithm: an example: E-step

| | smoker | Non-smoker |
|---|---|---|
| Group1 | 6 | 4 |
| Group2 | 7 | 3 |
| Group3 | 5 | 5 |
| Group4 | 9 | 1 |
| Group5 | 8 | 2 |
| | **35** | **15** |

- Initially, the model parameter is guessed (your belief) as follows:
  - Cancer patient: p(smoker) = 0.7
  - Non-cancer patient: p(smoker) = 0.3
- Calculate the probability from each class (cancer and non-cancer)
  - The class is modelled using Binomial distribution.
- Expect the posterior: p(cancer|smoker)
  - Responsibility of each class based on the given model parameter and data

Probability that {6,7,5,9,8} out of 10 are a smoker when they are cancer patients

Probability that {6,7,5,9,8} out of 10 are a smoker when they are non-cancer patients

Posterior showing how much responsible each class has for data set

$$\frac{\text{cancer}}{\text{cancer} + \text{non\_cancer}} \qquad \frac{\text{non\_cancer}}{\text{cancer} + \text{non\_cancer}}$$

| | Cancer | Non-cancer | Cancer | Non-cancer |
|---|---|---|---|---|
| G1 | $C(10,6)(0.7)^6(1-0.7)^4=0.200$ | $C(10,6)(0.3)^6(1-0.3)^4=0.037$ | 0.844 | 0.156 |
| G2 | $C(10,7)(0.7)^7(1-0.7)^3=0.267$ | $C(10,7)(0.3)^7(1-0.3)^3=0.009$ | 0.967 | 0.033 |
| G3 | $C(10,5)(0.7)^5(1-0.7)^5=0.103$ | $C(10,5)(0.3)^5(1-0.3)^5=0.103$ | 0.5 | 0.5 |
| G4 | $C(10,9)(0.7)^9(1-0.7)^1=0.121$ | $C(10,9)(0.3)^9(1-0.3)^1= 0.00013$ | 0.998 | 0.002 |
| G5 | $C(10,8)(0.7)^8(1-0.7)^2=0.233$ | $C(10,8)(0.3)^8(1-0.3)^2=0.00145$ | 0.993 | 0.007 |

# EM algorithm: an example: M-step

- Posteriors: p(cancer|smoker) and p(non-cancer|smoker) are given from E-Step
- The parameter of the Binomial distribution is calculated to maximize its likelihood function.

| | smoker | Non-smoker |
|---|---|---|
| Group1 | 6 | 4 |
| Group2 | 7 | 3 |
| Group3 | 5 | 5 |
| Group4 | 9 | 1 |
| Group5 | 8 | 2 |
| | **35** | **15** |

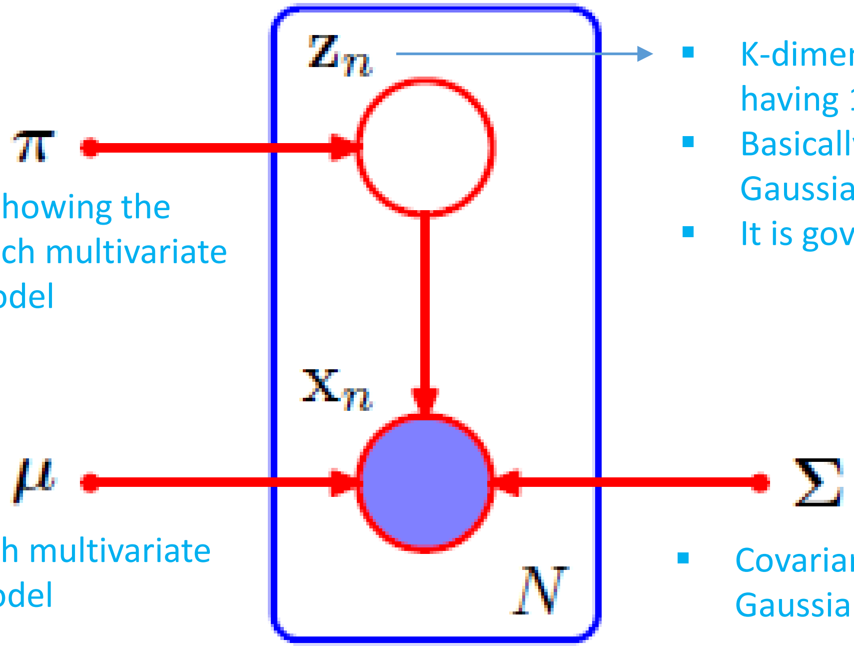
| | Cancer | | Non-cancer | |
|---|---|---|---|---|
| | Smoker | Non-smoker | Smoker | Non-smoker |
| G1 | 6 x 0.844 =5.069 | 4 x 0.844 =3.379 | 6 x 0.156 = 0.931 | 4 x 0.156 = 0.621 |
| G2 | 7 x 0.967 = 6.772 | 3 x 0.967 = 2.902 | 7 x 0.033 = 0.228 | 3 x 0.033 = 0.098 |
| G3 | 5 x 0.5 = 2.500 | 5 x 0.5 = 2.500 | 5 x 0.5 = 2.500 | 5 x 0.5 = 2.500 |
| G4 | 9 x 0.998 = 8.990 | 1 x 0.998 = 0.999 | 9 x 0.002 = 0.010 | 1 x 0.002 = 0.001 |
| G5 | 8 x 0.993 = 7.951 | 2 x 0.993 = 1.988 | 8 x 0.007 = 0.049 | 2 x 0.007 = 0.012 |
| | 31.28 | 11.77 | 3.72 | 3.23 |
| | p(smoker)=31.28/(31.28+11.77)=**0.73** | | p(smoker)=3.72/(3.72+3.23)=**0.54** | |

- Comparing to the previous values: Cancer patient, p(smoker)=0.7, Non-cancer patient, p(smoker)=0.3
- If the values do not change much, go to E-step. Otherwise, stop.

Graphical representation of a GMM

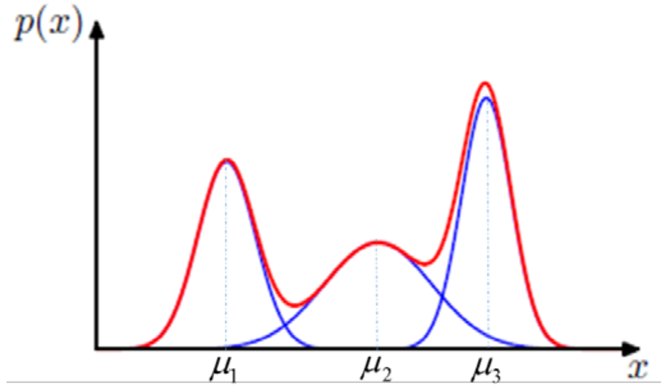
# Graphical representation of a GMM

- Probability showing the weight of each multivariate Gaussian model
- Mean of each multivariate Gaussian model




- K-dimensional binary random variable having 1-of-K representation
- Basically, tell you which multivariate Gaussian model is active.
- It is governed by $\pi$

- Covariance of each multivariate Gaussian model




$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k N(\mathrm{x} \mid \mu_k, \Sigma_k)$$

Multivariate Gaussian distribution

- ❑ Select one multivariate Gaussian distribution using $\pi$
- ❑ From the selected multivariate Gaussian distribution with $\mu$ and $\Sigma$, generate a sample