# Unsupervised Learning

- Clustering
  - $\rightarrow$ kmeans $\rightarrow$ latent variable modeling
  - $\rightarrow$ Spectral $\leftrightarrow$ mixture models
  - ⋮
- latent variable modeling
  - $\downarrow$ mixture models [Expectation Maximization]
  - $\downarrow$ Factor Analysis
  - Principal Component Decomposition (PCA)
  - $\updownarrow$ SVD $\rightarrow$ recommender systems.

---

## Latent Variable Models.

$p(\;)$

Topics $\rightarrow$ Technology, ~~Sci~~ Politics, Health, Sports

$p(Z)$

Z is a random variable that denotes the article topic

$p(X|Z) \rightarrow$ generates article given the topic.

| article | topic |
|---|---|
| $X_1$ | $Z_1$ |
| $X_2$ | $Z_2$ |
| $X_3$ | $Z_3$ |
| $X_4$ | ⋮ |
| ⋮ | ⋮ |
| $X_N$ | $Z_N$ |

$\rightarrow$ Latent variable model

find Z given X

# Mixture Models.

$z \in \{1, 2, \ldots, k\}$ $\quad p(z) \sim \text{Multinomial}(\pi)$

$p(x|z) = p(x|z=k) = p_k(x)$ Each $p_k()$ will have its own parameters.

~~$p(x|\theta$~~

let us consider one data point

$$\boxed{p(x, z) = p(x|z)\, p(z)}$$

We want to find $p(z|x)$ and we want to find the parameters.

$\boxed{EM}$

# Mixture Models.

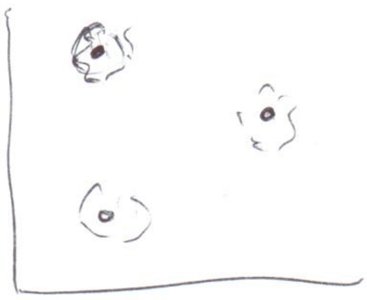
$x \rightarrow$ observation $\quad x \in R^D$

$z \rightarrow$ latent variable $\quad z \in \{\cancel{0}, 1, 2, \dots K\}$ $\rightarrow$ components.

$p(x \mid z=k) \longrightarrow$ Gaussian $(\mu_k, \Sigma_k)$

$p_k(x)$ $\qquad \mu_k$ (D×1) $\quad \Sigma_k$ (D×D)

There are K Gaussian distributions (components)



for K=3

We also have a mixing probability vector $\rightarrow \pi \rightarrow (\pi_1, \pi_2, \dots, \pi_k)$

$\boxed{\sum \pi_k = 1}$

---

① Estimate $\mu_1, \Sigma_1$, $\mu_2\ \Sigma_2$ and $\pi$

given $\begin{bmatrix} x_1 & z_1 \\ x_2 & z_2 \\ \vdots & \vdots \end{bmatrix}$

Yes

Use $z_i$ to separate data into two parts.

Use MLE to get $\mu_1\ \Sigma_1$ and $\mu_2\ \Sigma_2$

Use the counts of Zs to get $\pi$

---

② Given $\mu_1\ \Sigma_1$, $\mu_2\ \Sigma_2$ and $\pi$ $\rightarrow$ $\overset{N}{\text{pdf}}(x_i \mid \mu_1, \Sigma_1)$

what is z for a given x?

$$p(z_i = 1 \mid x_i) = \frac{p(x_i \mid z_i = 1)\, \overset{\pi}{p(z_i = 1)}}{p(x_i \mid z_i = 1)\, p(z_i = 1) + p(x \mid z_i = 2)\, p(z_i = 2)}$$

Yes $p(z_i = 2 \mid x_i) =$

Choose $z_i$ with highest value

What if we do not know $\underbrace{\mu_1, \Sigma_1, \mu_2, \Sigma_2, \pi}$ and $z_i$?

We will assume that we know $K$.

We will use $\underline{\theta}$ to denote all our parameters

---

What is the log likelihood of data?

$x_1$
$x_2$
$\vdots$
$x_N$

$$ll(\theta) = \sum_{i=1}^{N} \log p(x_i|\theta)$$

$$\boxed{p(x_i|\theta) = \sum_{k=1}^{K} p_k(x_i|\theta)\; p(z_i=k)}$$

$p(z_i=k) \rightarrow \pi_k$

$p_k(x_i|\theta) \rightarrow \mathcal{N}(\ |\mu, \Sigma)$

$\log(x+y)$

$\frac{1}{(x+y)}$

$$ll(\theta) = \sum_{i=1}^{N} \log\left[\sum_{k=1}^{K} \left(p_k(x_i|\theta)\, p(z_i=k)\right)\right]$$

Technically, we can maximize $ll(\theta)$ w.r.t $\theta$.

---

Expectation Maximization (EM)

At iteration $t-1 \rightarrow$ we have some $\theta^{(t-1)}$ and $z^{(t-1)}$

$$\boxed{\sum \log p(x_i, z_i|\theta)}$$

$$p(x_i, z_i) = p_{z_i}(x_i)\, p(z_i)$$

$$\boxed{Q = \mathbb{E}_{\theta}\left[\sum_{i=1}^{N} \log p(x_i, z_i|\theta)\right]}$$

$Q \rightarrow$ will be a function of $\theta$

Maximize $Q$ w.r.t. $\theta$

$f(x)$

$\mathbb{E}[f(x)] = \int f(x)\, p(x)\, dx$

# EM

$\underline{\underline{K=2}}$

| $x_i$ |
|---|
| 7 |
| 8 |
| 2 |
| 9 |
| 3 |
| 1 |

$$\Theta \to \begin{cases} \mu_1 \quad \sigma_1 \\ \mu_2 \quad \sigma_2 \\ \pi \end{cases}$$

(E) Find $z_i$'s

$z_1$
$z_2$
$z_3$
$z_4$
$z_5$
$z_6$

Assume we know $\Theta$

$$p(z_i = 1 \mid x_i, \Theta)$$

$$= \cancel{p(\Theta)} \frac{p(z_i=1)\, p(x_i \mid z_i = 1)}{(\quad)}$$

$$p(z_i = 2 \mid x_i, \Theta) = 1 - (\quad)$$

Then compute $\mathbb{E}\left[\sum \log\left[p(x_i, z_i \mid \Theta)\right]\right] = \int f(\theta)\, p(\theta)\, d\theta$

(M) Maximize $\mathbb{E}[\;]$

Example $D=1$, $K=2$, $N=6$.

$\Theta \to$ $\mu_1$ $\sigma_1^2$, $\mu_2$ $\sigma_2^2$, $\pi_1, \pi_2 \to$ scalar

$\pi_1 + \pi_2 = 1$

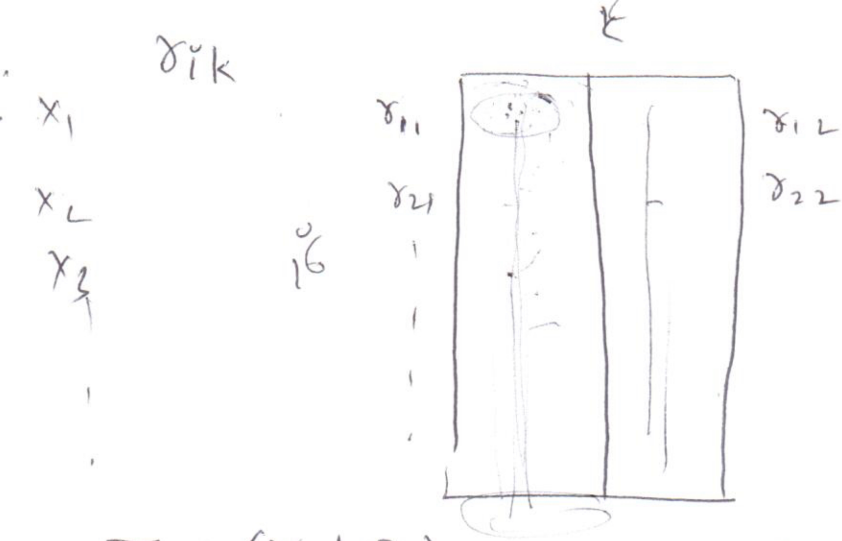
Initialize $\Theta$ $\quad \pi \to$ a number between 0 & 1

Usually we initialize $\pi_1 = \pi_2 \to 0.5$

Just set $\mu_1 = \mu_2 =$ sample mean $\to$ 5

$\sigma_1^2 = \sigma_2^2 = 1$

## E step




$$\gamma_{11} = \frac{\pi_1\, p(x_1 \mid \theta_1)}{\pi_1\, p(x_1 \mid \theta_1) + \pi_2\, p(x_1 \mid \theta_2)} = \frac{0.5 * \frac{1}{\sqrt{2\pi}\sigma_1} \exp - \left[\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2\right]}{(0 \qquad) + (\qquad)}$$

## M step

$$\pi_1 = \frac{1}{6}\left(\gamma_{11} + \gamma_{21} + \gamma_{31} + \cdots\right)$$

$$\pi_2 = \frac{1}{6}\left(\gamma_{12} + \gamma_{22} + \gamma_{32} + \cdots\right)$$

$$\mu_1 = \frac{\left(\gamma_{11} x_1 + \gamma_{21} x_2 + \gamma_{31} x_3 + \cdots\right)}{\left(\gamma_{11} + \gamma_{21} + \cdots\right)}$$

$$\mu_2 = \underline{\qquad\qquad}$$

$$\sigma_1^2 = \underline{\qquad\qquad}$$

$$\sigma_2^2 = \underline{\qquad\qquad}$$