Parameter Estimation

$D = \{\underline{x}_1, \dots, \underline{x}_n\}$ $\quad P(x|\theta) \quad \theta = ?$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \underset{\log}{\wedge} P(D|\theta) \overset{\text{i.i.d.}}{=\!=} \arg\max_{\theta} \sum_{i=1}^{n} \log P(x_i|\theta)$$

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^d \end{bmatrix}$$

$$\begin{bmatrix} \text{math} \\ \text{phy} \\ \text{ch} \end{bmatrix}$$

ch: $E[ch]$

$E[y|x]$

observed

hidden

$$\begin{bmatrix} 20 \\ 20 \\ ? \end{bmatrix}$$

$E[ch] = 14$

0.9

0.8

$E[ch \mid phy, math]$

$E[ch \mid phy = 20, math = 20]$

$$D = \left\{ \begin{bmatrix} x_1 \\ z_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ z_2 \end{bmatrix}, \dots, \begin{bmatrix} x_n \\ z_n \end{bmatrix} \right\}$$

x_i : observed

z_i : hidden/Latent/missing

$$P\left(\begin{bmatrix} x \\ z \end{bmatrix} \middle| \theta \right) = P(x, z | \theta)$$

$\hat{\theta} = ?$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \log P(D|\theta) = \arg\max_{\theta} \underbrace{\sum_{i=1}^{n} \log P(x_i, z_i | \theta)}_{\text{log Likelihood } L(\theta)}$$

$$\mu = \begin{bmatrix} \mu_{math} \\ \mu_{phy} \\ \mu_{ch} \end{bmatrix} \begin{matrix} \to \checkmark \\ \to \checkmark \\ \end{matrix} \begin{matrix} \Big\} x \\ \} z \end{matrix}$$

$$\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \checkmark$$

$$\boxed{\hat{\mu}_z = \frac{1}{n} \sum_{i=1}^{n} z_i} \quad X$$

$$\hat{\theta} = \arg\max_{\theta} E_Z\left[\underbrace{\sum_{i=1}^{n} \log P(x_i, z_i \mid \theta)}_{\text{complete Log Likelihood } L(\theta)}\right]$$

$$\boxed{Q(\theta) = E_Z\left[L(\theta)\right]}$$ E-step

$$\frac{dQ(\theta)}{d\theta} = 0 \qquad \max_{\theta} Q(\theta)$$ M-step

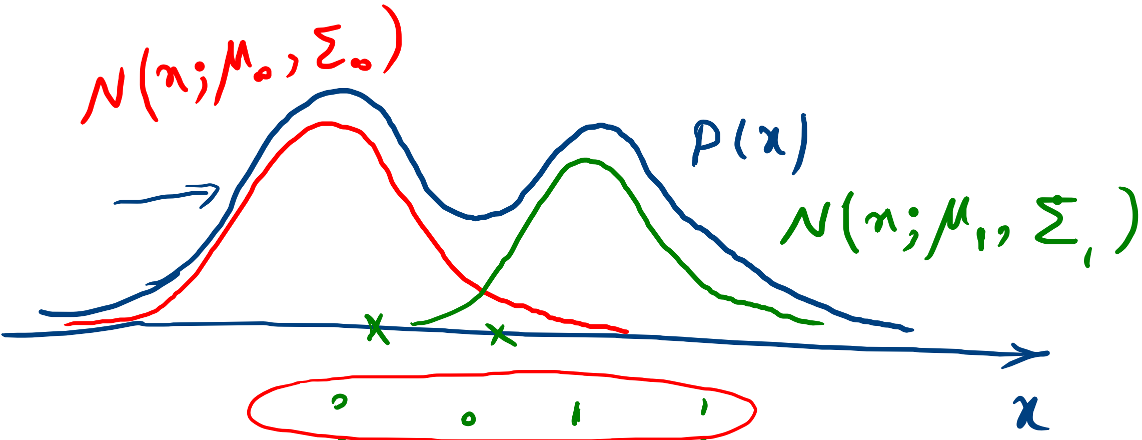
EM

$$E_z\left[L(\theta)\right] = \int_z L(\theta)\, \underbrace{p(z|x)}\, dz$$

$$\underline{p(z|x)} = \frac{p(x, z|\theta)}{p(x)}$$

$\theta^0 \quad \theta^1 \quad \theta^2$

Gaussian Mixture Model



$$z_i = \begin{cases} 0 & x \in \mathcal{N}_0 \\ 1 & x \in \mathcal{N}_1 \end{cases}$$

$$D = \{x_1, x_2, \ldots, x_n\} \rightarrow D = \left\{ \begin{bmatrix} x_1 \\ z_1 \end{bmatrix}, \ldots, \begin{bmatrix} x_n \\ z_n \end{bmatrix} \right\}$$

$$\theta = (\mu_1, \Sigma_1, \mu_2, \Sigma_2, \alpha)$$

$$p(x|\theta) = (1-\alpha)\underbrace{\mathcal{N}(x;\mu_0,\Sigma_0)}_{\text{component}} + \alpha \underbrace{\mathcal{N}(x;\mu_1,\Sigma_1)}$$

$$P(x|\theta) = (1-\alpha)N_0 + \alpha N_1$$

$$P(x,z|\theta) = \left[(1-\alpha)N(x;\mu_0,\Sigma_0)\right]^{(1-z)} \left[\alpha N(x;\mu_1,\Sigma_1)\right]^{z}$$

$$\cancel{= (1-z)\left[\text{—}\,//\,\text{—}\right] + z\left[\text{—}\,//\,\text{—}\right]}$$

$$E_z\left[L(\theta)\right] = E_z\left[\underbrace{\sum_{i=1}^{n} \log P(x_i,z_i|\theta)}\right] = \sum_{i=1}^{n} E_z\left[\log P(x_i,z_i|\theta)\right]$$

$$= \sum_{i=1}^{n} E_z\Big[\underbrace{(1-z_i)\log(1-\alpha)} + (1-z_i)\log N(x_i|\mu_0,\Sigma_0) + z_i\log\alpha$$
$$+ z_i \log N(x_i|\mu_1,\Sigma_1)\Big]$$

$$= \sum_{i=1}^{n}\Big(\log(1-\alpha)(1-E[z_i]) + \log N(x_i|\mu_0,\Sigma_0)(1-E[z_i]) + E[z_i]\log\alpha$$
$$+ E[z_i]\log N(x_i|\mu_1,\Sigma_1)\Big)$$

$$E[z] = E[z|x] = P(z=1|x) = \frac{P(x|z=1)\,P(z=1)}{P(x)}$$

$$= \frac{N(x|\mu_1^t,\Sigma_1^t)\,\alpha^t}{(1-\alpha^t)N(x|\mu_0^t,\Sigma_0^t) + \alpha^t N(x|\mu_1^t,\Sigma_1^t)} = \gamma^t$$

$$Q(\theta) = E_z\left[L(\theta)\right] = \sum_{i=1}^{n} \underline{\log(1-\alpha)(1-\gamma_i^t)} + \log N(x_i \mid \mu_0, \Sigma_0)(1-\gamma_i^t)$$

$$+ \gamma_i^t \log\alpha + \gamma_i^t \log N(x_i \mid \mu_1, \Sigma_1)$$

M-step:

$$\frac{dQ}{\partial\alpha} = 0 \qquad \sum_{i=1}^{n}\left(\frac{-1}{1-\alpha}(1-\gamma_i^t) + \frac{\gamma_i^t}{\alpha}\right) = 0$$

$$\frac{n - \sum \gamma_i^t}{1-\alpha} = \frac{\sum \gamma_i^t}{\alpha} \longrightarrow \boxed{\alpha = \frac{\sum_{i=1}^{n} \gamma_i^t}{n}}$$