AI – Machine Learning

Artificial Intelligence
Research Group

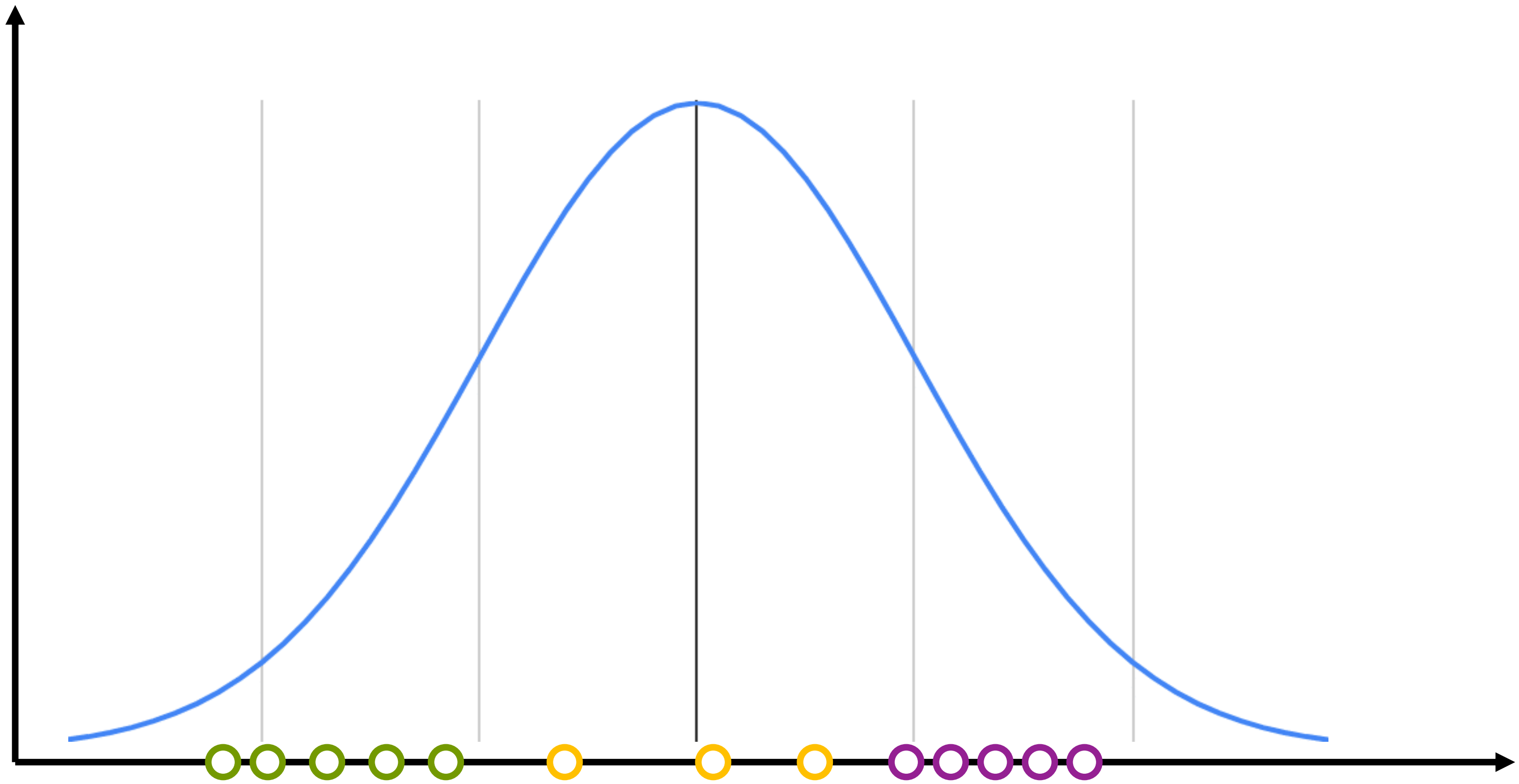
Motivation

$$f(\theta|X) = \underbrace{\log[p(X|\theta)]}_{\textit{Log-likelihood}} = \sum_{i=1}^{N} \log[p(x_i|\theta)]$$

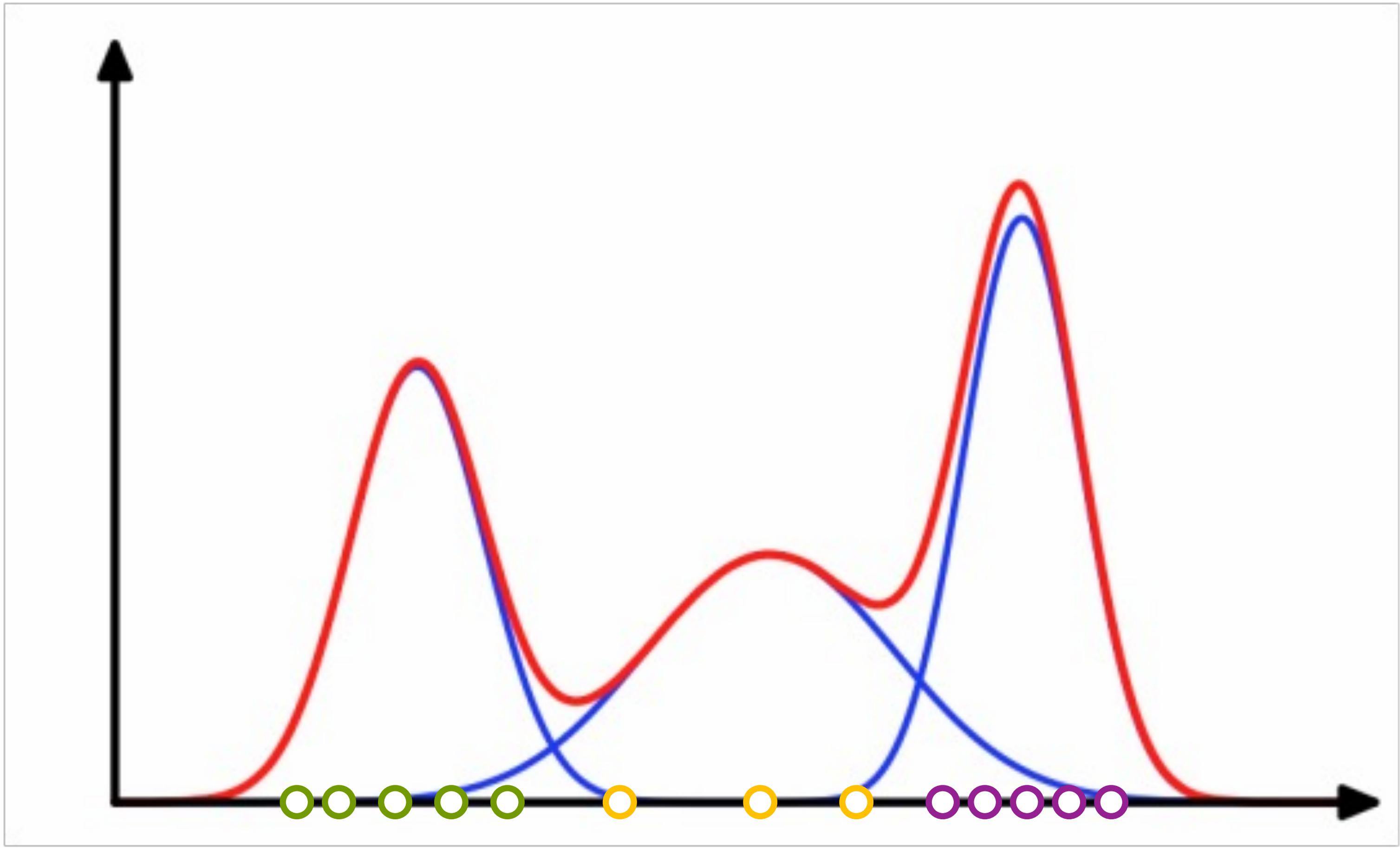
$$\underset{\theta}{\text{Argmax}} \left[\sum_{i=1}^{N} \log[\mathcal{N}(x_i|\mu,\sigma)]\right]$$

$$\mu_{\text{MLE}} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma_{\text{MLE}}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_{\text{MLE}})^2}{N}$$

Motivation



Motivation



$$\underset{\theta}{\mathrm{Argmax}}\left[\sum_{i=1}^{N}\log\left[\sum_{l=1}^{k}\alpha_l\,\mathcal{N}(x_i|\mu_l,\Sigma_l)\right]\right]$$

EM Framework

$$\theta^{(1)}, \theta^{(2)}, \cdots \theta^{(t)}$$

$$\theta^{(g+1)} = f(\theta^{(g)})$$

$$\theta^{(g+1)} = \underset{\theta}{\operatorname{Argmax}} \int \log[p(X, Z|\theta)]\, p\left(Z \middle| X, \theta^{(g)}\right) dZ$$

$$\theta = \{\alpha_1, \cdots, \alpha_k, \mu_1, \cdots, \mu_k, \Sigma_1, \cdots, \Sigma_k\}$$

X: Data observation

Z: Latent variable

Condition

- $p(X) = \int p(X|Z)p(Z)dZ$

  $$p(x_i) = \int p_\theta(x_i|z)p_\theta(z)dz$$

  $$= \sum_{l=1}^{k} \mathcal{N}(x_i|\mu_l, \Sigma_l)\alpha_l$$

- $\log\left[p\left(X|\theta^{(g+1)}\right)\right] \geq \log\left[p\left(X|\theta^{(g)}\right)\right]$

Proof

$$\log[p(X|\theta)] = \log[p(X,Z|\theta)] - \log[p(Z|X,\theta)] \qquad \because p(X) = \frac{p(X,Z)}{p(Z|X)}$$

$$E\big[\log[p(X|\theta)]\big] = E\big[\log[p(X,Z|\theta)] - \log[p(Z|X,\theta)]\big]$$

$$\int \log[p(X,Z|\theta)]\, p\big(Z\big|X,\theta^{(g)}\big)dZ - \int \log[p(Z|X,\theta)]\, p\big(Z\big|X,\theta^{(g)}\big)dZ$$

$$Q\big(\theta,\theta^{(g)}\big) - H(\theta,\theta^{(g)})$$

$$\leq Q\big(\theta^{(g+1)},\theta^{(g)}\big) - H(\theta^{(g+1)},\theta^{(g)}) \qquad \underset{\theta}{\text{Argmax}} \int \log[p(X,Z|\theta)]p\big(Z\big|X,\theta^{(g)}\big)dZ$$

If $\forall\theta\{H(\theta^{(g)},\theta^{(g)}) \geq H(\theta,\theta^{(g)})\}$ then

$$H(\theta^{(g)},\theta^{(g)}) \geq H(\theta^{(g+1)},\theta^{(g)})$$

Proof

$$H(\theta^{(g)},\theta^{(g)}) \geq H(\theta,\theta^{(g)}) \Rightarrow H\left(\theta^{(g)},\theta^{(g)}\right) - H(\theta,\theta^{(g)}) \geq 0$$

$$\int \log[p(Z|X,\theta^{(g)})]\,p(Z|X,\theta^{(g)})dZ - \int \log[p(Z|X,\theta)]\,p(Z|X,\theta^{(g)})dZ$$

$$= \int \log\left(\frac{p(Z|X,\theta^{(g)})}{p(Z|X,\theta)}\right)p(Z|X,\theta^{(g)})dZ$$

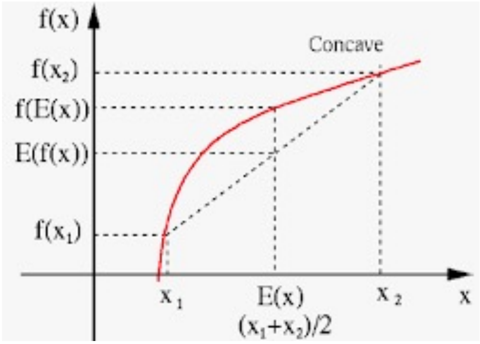
$$= -\int \log\left(\frac{p(Z|X,\theta)}{p(Z|X,\theta^{(g)})}\right)p(Z|X,\theta^{(g)})dZ$$

Jensen's inequality $(1-p)f(x) + pf(y) \leq f((1-p)x + py)$

$$\geq -\log \int \left(\frac{p(Z|X,\theta)}{p(Z|X,\theta^{(g)})}\right)p(Z|X,\theta^{(g)})dZ$$

$$\geq -\log(1) = 0 \Rightarrow$$

$$H\left(\theta^{(g)},\theta^{(g)}\right) - H(\theta,\theta^{(g)}) \geq 0$$



Expectation Maximization

$$\theta^{(g+1)} = \underset{\theta}{\text{Argmax}} \int \log[p(X,Z|\theta)]p\left(Z \middle| X,\theta^{(g)}\right)dZ$$

$$p(X,Z|\theta) = \prod_{i=1}^{N} p(x_i, z_i|\theta) = \prod_{i=1}^{N} p(x_i|z_i,\theta)p(z_i|\theta) = \prod_{i=1}^{N} \mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\alpha_{z_i}$$

$$p\left(Z \middle| X,\theta^{(g)}\right) = \prod_{i=1}^{N} p(z_i|x_i,\theta^{(g)})$$

$$p\left(z_i \middle| x_i,\theta^{(g)}\right) = \frac{p(x_i|z_i)p(z_i)}{\sum_{z_i=1}^{k} p(x_i|z_i)p(z_i)} = \frac{\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\alpha_{z_i}}{\sum_{z_i=1}^{k} \mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\alpha_{z_i}}$$

Expectation Maximization

$$\int \log[p(X,Z|\theta)]p\left(Z\middle|X,\theta^{(g)}\right)dZ$$

$$\sum_{z_1=1}^{k}\sum_{z_2=1}^{k}\cdots\sum_{z_N=1}^{k}\left(\sum_{i=1}^{N}\underbrace{[\log\alpha_{z_i}+\log\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})]}_{f_i(z_i)}\underbrace{\prod_{i=1}^{N}p(z_i|x_i,\theta^{(g)})}_{p(z_1,\cdots,z_N)}\right)$$

$$\sum_{z_1=1}^{k}\sum_{z_2=1}^{k}\cdots\sum_{z_N=1}^{k}([f_1(z_1)+f_2(z_2)+\cdots+f_N(z_N)]p(z_1,\cdots,z_N))$$

$$\sum_{z_1=1}^{k}f_1(z_1)p(z_1)+\sum_{z_2=1}^{k}f_2(z_2)p(z_2)+\cdots+\sum_{z_N=1}^{k}f_N(z_N)p(z_N)$$

$$\sum_{i=1}^{N}\sum_{z_i=1}^{k}\left(\log\alpha_{z_i}+\log\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\right)p(z_i|x_i,\theta^{(g)})$$

Expectation Maximization

$$\sum_{i=1}^{N}\sum_{z_i=1}^{k}\left(\log \alpha_{z_i} + \log \mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})\right)p(z_i|x_i,\theta^{(g)}) \qquad \sum_{l=1}^{k}\alpha_l = 1$$

$$\alpha_l = \frac{1}{N}\sum_{i=1}^{N} p(l|x_i,\theta^{(g)}) \qquad p\left(l\middle|x_i,\theta^{(g)}\right) = \frac{\mathcal{N}(x_i|\mu_{z_i},\Sigma_{z_i})}{\sum_{z_j=1}^{k}\mathcal{N}(x_i|\mu_{z_j},\Sigma_{z_j})}$$

$$\mu_l = \frac{\sum_{i=1}^{N} x_i\, p(l|x_i,\theta^{(g)})}{\sum_{i=1}^{N} p(l|x_i,\theta^{(g)})}$$

$$\Sigma_l = \frac{\sum_{i=1}^{N}[x_i-\mu_l][x_i-\mu_l]^{\top}\, p(l|x_i,\theta^{(g)})}{\sum_{i=1}^{N} p(l|x_i,\theta^{(g)})}$$

Any questions?

AI Research Group
Fudan University