$\boxed{P_\theta(X)}$ $\quad X^1, X^2, \ldots, X^m$

$\boxed{P_\theta(X,Z)}$ $\quad P_\theta(X,Z) = P_{\theta_1}(X|Z)\, P_{\theta_2}(Z)$

log-likelihood $\quad ll(\theta) = \sum_{i=1}^{m} \log P_\theta(x^i) = \sum_{i=1}^{m} \log \sum_{Z} P_\theta(x^i, z)$

$ll(\theta_0) \leq ll(\theta_1) \leq ll(\theta_2) \leq ll(\theta_3) \leq \cdots \leq ll(\theta_t) \leq \cdots$

$\boxed{\theta_t} \rightarrow$ current θ

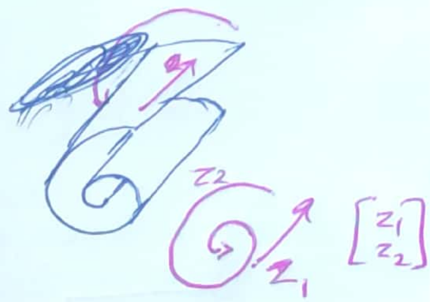
We dont have access to z!

heuristic: $z^i = \arg\max_{Z} P_{\theta_t}(Z|x^i)$ $\rightarrow$ posterior distribution $\quad$ Maximize

OR

$z^i \sim P_{\theta_t}(Z|x^i)$ $\quad$ take a sample

data



$$\theta_{t+1} = \arg\max_\theta \sum_{i=1}^{m} \log P_\theta(x^i, z^i)$$

How to compute the posterior?

$$P_\theta(X,Z) = P_\alpha(X|Z)\, P_\beta(Z) \qquad \theta = (\alpha, \beta)$$

$$\theta_t = (\alpha_t, \beta_t)$$

$$P_{\theta_t}(Z|x^i) = \frac{P_{\theta_t}(x^i, Z)}{\sum_Z P_{\theta_t}(x^i, Z)} = \frac{P_{\alpha_t}(x^i|Z)\, P_{\beta_t}(Z)}{\sum_Z P_{\alpha_t}(x^i|Z)\, P_{\beta_t}(Z)}$$

Expectation-Maximization

Compute $P_{\theta_t}(Z|x^i)$ for all Z. $\quad$ Expectation step (E-step)

$$\theta_{t+1} = \arg\max_\theta \sum_{i=1}^{m} E_{P_{\theta_t}(Z|x^i)}\left\{ \log P_\theta(x^i, Z)\right\} \rightarrow \text{expected log-likelihood}$$

$$\theta_{t+1} = \arg\max_\theta \sum_{i=1} \sum_Z P_{\theta_t}(Z|x^i) \log P_\theta(x^i, Z)$$

Maximization-Step (M-step)

EM (Expectation-Maximization): Alternate betweem E-step and M-step.

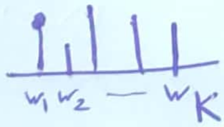
Can prove: $ll(\theta_0) \leq ll(\theta_1) \leq ll(\theta_2) \leq \cdots$

Mixture of Gaussians

→ training data $x^1, x^2, \ldots, x^m$

$$P(x) = \sum_{k=1}^{K} w_k \, N(x \mid \mu_k, \sigma_k^2) \qquad \sum_{k=1}^{K} w_k = 1$$

$$P(x,z) = \underbrace{p(x \mid z)}_{N(x;\, \mu_k, \sigma_k^2)} \; p(z) \rightarrow p(z) = \Pr(Z=z) = w_z$$



$$ll(\theta) = \sum_{i=1}^{m} \log P(x^i) = \sum_{i=1}^{m} \log \sum_{k=1}^{K} \underbrace{\Pr(Z=k)}_{w_k} N(x \mid \mu_k, \sigma_k^2)$$

$\theta = (\{\mu_k\} \{\sigma_k^2\} \; w_1, \ldots, w_K)$

θ^t

$$P(z \mid x^i) = \frac{P(x^i \mid z)\, p(z)}{\sum_z p(x^i \mid z)\, p(z)}$$

E-step

find

$$\Pr(Z=k \mid x^i) = \frac{N(x^i;\, \mu_k^t, \sigma_k^{2\,t})\, w_k^t}{\sum_{k'=1}^{K} N(x^i;\, \mu_{k'}^t, \sigma_{k'}^{2\,t})\, w_{k'}^t} = \alpha_{ik}^t$$

for $x^1, x^2, \ldots, x^m$
for $k = 1, 2, \ldots, K$

M-step

$$E_{p_{\theta_t}(z \mid x^i)} \left\{ \sum_{i=1}^{m} \log p_\theta(x^i; z) \right\} = \sum_{i=1}^{m} \sum_{z} P_{\theta^t}(z \mid x^i) \log P_\theta(x^i; z)$$

$$= \sum_{i=1}^{m} \sum_{z=1}^{K} P_{\theta^t}(z \mid x^i) \left[\log P_\theta(x^i \mid z) + \log P_\theta(z) \right]$$

$$\theta^{t+1} = \arg\max_{\theta} \underbrace{\sum_{i=1}^{m} \sum_{k=1}^{K} \alpha_{ik}^t \left[\log N(x^i \mid \mu_k, \sigma_k) + \log w_k \right]}$$

θ: $\{\mu_k\}, \{\sigma_k\}, \{w_k\}$

$$\frac{\partial}{\partial w_j} \sum_{k=1}^{K} \underbrace{\left(\sum_{i=1}^{m} \alpha_{ik}^t \right)}_{\beta_k} \log w_k \; \underline{+ \lambda\left(\sum w_k - 1\right)}$$

$$\boxed{w_j = \frac{\beta_j}{\sum_k \beta_k}}$$

$$\frac{\beta_j}{w_j} = -\lambda \Rightarrow \frac{w_j}{\beta_j} \text{ cost.} \qquad \sum w_j = 1 \qquad \sum \frac{\beta_j}{-\lambda} = 1$$

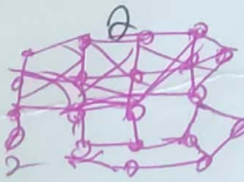
$P_\theta(X, Z)$ data $x^1, x^2, \ldots, x^m$

$\cancel{P} = P_{\theta_1}(z)\, P_{\theta_2}(x|z)$

$p(z) = \mathcal{N}(z; \mu = 0, \Sigma = I)$

$p_\theta(x|z) = \mathcal{N}(x; \mu_\theta(z), \Sigma_\theta(z))$

deep neural network

Z  $\mu_\theta(z)$, $\Sigma_\theta(z)$

Assume θ is known

take a sample from $P_\theta(x, z)$

1- $z^i \sim p(z) = \mathcal{N}(z; \mu = 0, \Sigma = I)$

z^i  Image of Noise

2- Feed z^i to neural nets to get $\mu_\theta(z^i)$, $\Sigma_\theta(z^i)$

3- sample $x^i \sim \mathcal{N}(x; \mu_\theta(z), \Sigma_\theta(z))$

$$p(x) = \sum_z p_\theta(x, z) = \sum_z p(x|z)\, p(z)$$

$$= \sum_z \mathcal{N}(x | \mu_\theta(z), \Sigma_\theta(z))\, \mathcal{N}(z; 0, I)$$

Very hard to compute the posterior! $p(z|x^i)$

Data $x^1, x^2, \ldots, x^m$