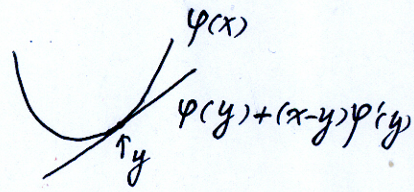
1. Jensen's Inequality

$$E[\varphi(X)] \geq \varphi(E[X]) \quad \text{for convex function } \varphi(X)$$

Proof: Suppose $\varphi(X)$ is differentiable. Because the function $\varphi(X)$ is convex, i.e.

$$\varphi(x) \geq \varphi(y) + (x-y)\varphi'(y)$$



Let $x = X$ and $y = E[X]$, we can get

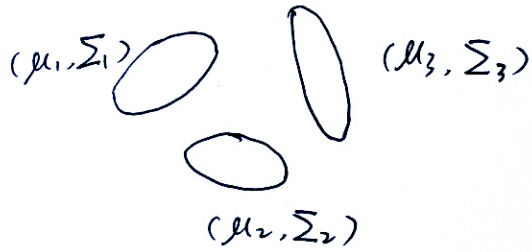
$$\varphi(X) \geq \varphi(E[X]) + (X - E[X])\varphi'(E[X])$$

Since the inequality holds for all X, we can take the expectation in both sides.

$$E[\varphi(X)] \geq \varphi(E[X]) + \underbrace{E[X-E[X]]\varphi'(E[X])}_{=0}$$

$$\Rightarrow \quad E[\varphi(X)] \geq \varphi(E[X])$$

2.



data (x, z)

$x \downarrow$ observed, $z \searrow$ hidden variable

$$p(x;\theta) = \sum_{j=1}^{m} p(x, z=j;\theta) = \sum_{j=1}^{m} p(x \mid z=j;\theta)\, p(z=j;\theta)$$

$$= \sum_{j=1}^{m} P_j\, N(x; \mu_j, \Sigma_j)$$

Complete log likelihood : assume both x and z are known.

$$l_c(x_1, \cdots x_n, z_1, \cdots, z_n; \theta)$$

$$= \sum_{k=1}^{n} \log p(x_k \mid \theta) = \sum_{k=1}^{n} \log P_{z_k} N(x_k; \mu_{z_k}, \Sigma_{z_k})$$

$$\frac{\partial l_c}{\partial \mu_j} = \sum_{z_k=j} \frac{\partial\left(-\frac{1}{2}(x_k-\mu_j)^t \Sigma_j^{-1}(x_k-\mu_j)\right)}{\partial \mu_j} = \sum_{z_k=j} \Sigma_j^{-1}(x_k-\mu_j) = 0$$

$$\Rightarrow \hat{\mu}_j = \frac{1}{n_j} \sum_{z_k=j} x_k$$

Similarly $\hat{\Sigma}_j = \frac{1}{n_j} \sum_{z_k=j} (x_k-\hat{\mu}_j)(x_k-\hat{\mu}_j)^t$

Consider l_c as a function of P_{z_k}

$$l_c(P_{z_k}) = \sum_{k=1}^{n} \log P_{z_k} + c = \sum_{j=1}^{m} n_j \log P_j + c$$

$$\propto \sum_{j=1}^{m} \frac{n_j}{n} \log P_j + \frac{c}{n} = l_c(P_j)$$

$$q_j = \frac{n_j}{n}, \quad \sum_j q_j = 1$$

$$\cancel{\max_{P_j} l_c(P_j)} \max_{P_j} l_c(P_j) = \max_{P_j} \sum_{j=1}^{m} q_j \log P_j$$

$$= \max_{P_j} \sum_{j=1}^{m} q_j \log P_j - \sum_{j=1}^{m} q_j \log q_j$$

$$= \max_{P_j} - \sum_{j=1}^{m} q_j \log \frac{q_j}{P_j} = \max_{P_j} - KL(q_j \| P_j) = \min_{P_j} KL(q_j \| P_j)$$