Homework #2

2. a) We need to show that $\forall \lambda \in [0,1]$
and $\forall\, x, y \in \mathbb{R}^n$

$$\lambda h(x) + (1-\lambda) h(y) \overset{?}{\geq} h(\lambda x + (1-\lambda) y)$$

$$h(\lambda x + (1-\lambda) y) = f(\lambda A x + (1-\lambda) A y)$$

$$\leq \lambda f(Ax) + (1-\lambda) f(Ay)$$

$$= \lambda h(x) + (1-\lambda) h(y)$$

b) $h(\lambda x + (1-\lambda) y)$

$$= f(\lambda x + (1-\lambda) y) + g(\lambda x + (1-\lambda) y)$$

$$\leq \lambda f(x) + (1-\lambda) f(y) + \lambda g(x) + (1-\lambda) g(y)$$

$$= \lambda h(x) + (1-\lambda) h(y)$$

3. $\nabla f(x^*) = 0 \Leftarrow$ x^* is the unique global minimum of f.

($\Leftarrow$) Assume that x^* is the unique global minimum of f,

Assume $\nabla f(x^*) \neq 0$

$\Rightarrow f(x) \cong f(x^*) + \nabla f(x^*)(x - x^*) + O(x^3)$

$\Rightarrow \exists\, x'$ near x^* such that $f(x') < f(x^*)$

$\Rightarrow$ contradiction

$\Rightarrow \nabla f(x^*) = 0$

($\Rightarrow$)

First assume $\exists\, x'$ such that

$f(x') < f(x^*)$ (x^* not global minimum.)

By strict convexity

$\forall \lambda \in [0, 1]$

$$f(\lambda x' - (1-\lambda) x^*) < \lambda f(x') + (1-\lambda) f(x^*)$$

Define $\Delta x = x' - x^*$

$\Delta f = f(x^*) - f(x') > 0$

Then

$$f(\lambda \Delta x + x^*) < -\lambda \Delta f + f(x^*)$$

$$f(\lambda \Delta x + x^*) - f(x^*) < -\lambda \Delta f$$

$$\frac{f(\lambda \Delta x + x^*) - f(x^*)}{\lambda} < -\Delta f$$

$$\lim_{\lambda \downarrow 0} \frac{f(\lambda \Delta x + x^*) - f(x^*)}{\lambda} < -\Delta f < 0$$

$$\Rightarrow \lim_{\lambda \to 0} \frac{f(\lambda \Delta x + x^*) - f(x^*)}{\lambda}$$

$$= \nabla_{\Delta x} f(x^*) \neq 0$$

contradiction $\Rightarrow$ x^* is global minimum

It remains to be shown that x^* is the unique global minimum.

Assume $\exists\, x' \neq x^*$ s.t. $f(x') = f(x^*)$

Then define $x'' = \dfrac{x' + x^*}{2}$

By strict convexity

$$f(x'') < \frac{1}{2}\left(f(x') + f(x^*)\right)$$

$$= f(x^*)$$

$\Rightarrow$ contradiction

$\Rightarrow$ x^* is the unique global minimum.

4)

$$Q(\theta', \theta) = E[\sum \log p(y, X | \theta') | \theta Y]$$

$$\log p(y, x | \theta') = \log p(y | x, \theta') + \log p(x | \theta')$$

$$\log p(x | \theta') = \sum_{n=0}^{N-1} \sum_{k=0}^{M-1} \delta(x_n - k) \log \pi_k'$$

$$\log p(y | x, \theta') = \sum_{n=0}^{N-1} \left\{ -\frac{1}{2\gamma_{x_n}} (y_n - \mu_{x_n})^2 - \frac{1}{2} \log(2\pi \gamma_{x_n}) \right\}$$

$$p(x | y, \theta) = \prod_{n=0}^{N-1} p(x_n | y, \theta) = \prod_{n=0}^{N-1} p(x_n | y_n, \theta)$$

$$E[\log p(y, X | \theta') | Y \theta]$$

$$= E\left[\sum_{n=0}^{N-1} - \frac{1}{2\gamma_{x_n}} (y_n - \mu_{x_n}')^2 - \frac{1}{2} \log(2\pi \gamma_{x_n}') \Big| Y\theta \right]$$

$$= \sum_{n=0}^{N-1} E\left[-\frac{1}{2\gamma_{x_n}} (y_n - \mu_{x_n}')^2 - \frac{1}{2} \log(2\pi \gamma_{x_n}') \Big| Y\theta \right]$$

$$= \sum_{n=0}^{N-1} \sum_{k=0}^{M-1} \left\{ -\frac{1}{2\gamma_k} (y_n - \mu_k')^2 - \frac{1}{2} \log(2\pi \gamma_k') \right\} f(k | y_n, \theta)$$

where $f(k | y_n, \theta) \triangleq p(x_n = k | y_n, \theta)$

$$= \sum_{k=0}^{M-1} \sum_{n=0}^{N-1} \left\{ -\frac{1}{2\gamma_k} (y_n - \mu_k)^2 - \frac{1}{2} \log(2\pi \gamma_k) \right\} f(k|y_n, \theta)$$

Define

$$\bar{N}_k \triangleq \sum_{n=0}^{N-1} f(k|y_n, \theta)$$

$$\bar{\mu}_k \triangleq \frac{1}{\bar{N}_k} \sum_{n=0}^{N-1} y_n f(k|y_n, \theta)$$

$$\alpha_k \triangleq \sum_{n=0}^{N-1} (y_n - \bar{\mu}_k)^2 f(k|y_n, \theta)$$

then

$$\rightarrow = \sum_{k=0}^{M-1} \sum_{n=0}^{N-1} \left\{ -\frac{(y_n - \bar{\mu}_k)^2}{2\gamma_k'} - \frac{(\bar{\mu}_k - \mu_k')^2}{2\gamma_k'} - \frac{1}{2} \log(2\pi \gamma_k') \right\} \cdot f(k|y_n, \theta)$$

$$= \sum_{k=0}^{M-1} \left\{ - \frac{\alpha_k}{2\gamma_k'} - \frac{(\bar{\mu}_k - \mu_k')}{2\gamma_k'} \bar{N}_k - \frac{\bar{N}_k}{2} \log(2\pi \gamma_k') \right\}$$

$$E[\log p(X|\theta') \,|\, Y\theta]$$

$$= E\left[\sum_{n=0}^{N-1}\sum_{k=0}^{M-1}\delta(x_n-k)\log\pi_k' \,\Big|\, Y\theta\right]$$

$$= \sum_{n=0}^{N-1}\sum_{k=0}^{M-1}\log\pi_k' \, E[\delta(x_n-k) \,|\, Y\theta]$$

$$= \sum_{n=0}^{N-1}\sum_{k=0}^{M-1}\log\pi_k' \, p(x_n=k|y,\theta)$$

$$= \sum_{n=0}^{N-1}\sum_{k=0}^{M-1}\log\pi_k' \, f(k|y_n,\theta)$$

Define

$$\bar{N}_k = \sum_{n=0}^{N-1} f(k|y_n,\theta)$$

the

$$= \sum_{k=0}^{M-1}\log\pi_k'\left(\sum_{n=0}^{N-1} f(k|y_n,\theta)\right)$$

$$= \sum_{k=0}^{M-1}\bar{N}_k\log\pi_k'$$

So

$$Q(\theta', \theta) =$$

$$\sum_{k=0}^{M-1} \left\{ -\frac{\alpha_k}{2\gamma_k'} - \frac{\bar{N}_k}{2} \log(2\pi\gamma_k') - \frac{(\bar{\mu}_k - \mu_k')^2}{2\gamma_k'} \bar{N}_k + \bar{N}_k \log \pi_k' \right\}$$

$$\hat{\theta} = \arg\max_{\theta'} Q(\theta', \theta)$$

$$\hat{\mu}_k = \arg\max_{\mu_k'} \left\{ \frac{-(\bar{\mu}_k - \mu_k')^2}{2\gamma_k'} \bar{N}_k \right\}$$

$$= \arg\min_{\mu'} \left\{ \frac{N_k}{2\gamma_k'} (\bar{\mu}_k - \mu_k')^2 \right\}$$

$$= \mu_k'$$

$$\hat{\mu}_k = \frac{1}{\bar{N}_k} \sum_{n=0}^{N-1} y_n f(k \mid y_n, \theta)$$

$$\hat{\gamma}_k = \arg\max_{\gamma_k'} \left\{ -\frac{\alpha_k}{2\gamma_k'} - \frac{\bar{N}_k}{2} \log(2\pi\gamma_k') \right\}$$

$$= \frac{\alpha_k}{N_k}$$

$$\hat{\gamma}_K = \frac{1}{\bar{N}_K} \sum_{n=0}^{N-1} (y_n - \bar{\mu}_K)^2 f(K \mid y_n, \theta)$$

$$\hat{\pi} = \underset{\substack{\pi' \\ \sum_i \pi_i' = 1}}{\operatorname{argmax}} \sum_{K=0}^{M-1} \bar{N}_K \log \pi_K'$$

$$\hat{\pi}_K = \frac{\bar{N}_K}{\sum_{K=0}^{M-1} \bar{N}_K}$$

<u>EM Algorithm</u>

1. Compute

$$f(K \mid y_n, \theta) = \frac{p(y_n \mid \gamma_K, \mu_K)\, \pi_K}{\sum_{K=0}^{M-1} p(y_n \mid \gamma_K, \mu_K)\, \pi_K}$$

2. Compute updates

$$\bar{N}_K \leftarrow \sum_{n=0}^{N} f(K \mid y_n, \theta)$$

$$\mu_K \leftarrow \frac{1}{\bar{N}_K} \sum_{n=0}^{N-1} y_n f(K \mid y_n, \theta)$$

$$\gamma_K \leftarrow \frac{1}{\bar{N}_K} \sum_{n=0}^{N-1} (y_n - \bar{\mu}_K)^2 f(K \mid y_n, \theta)$$

5) $\log p(x) = \log \pi_{x_0} + \sum_{n=1}^{N} \log P_{x_{n-1} x_n}$

Define

$$h(l,m) = \sum_{n=1}^{N} \delta(x_{n-1}-l)\,\delta(x_n-m)$$

then

$$\log p(x) = \log \pi_{x_0} + \sum_{l}\sum_{m} h(l,m) \log P_{lm}$$

$$\hat{\pi} = \underset{\sum_i \pi_i = 1}{\arg\max} \; \pi_{x_0}$$

$$\Rightarrow \quad \hat{\pi}_k = \delta(k - x_0)$$

$$\hat{P}_{lm} = \underset{\substack{P_{l*} \\ \sum_i P_{li} = 1}}{\arg\max} \; \sum_m h(l,m) \log P_{lm}$$

using Lagrange multipliers

$$\frac{\partial}{\partial P_{lk}} \left(\sum_m h(l,m) \log P_{lm} \right) +$$

$$\frac{\partial}{\partial P_{lk}} \lambda \left(\sum_m P_{lm} - 1 \right) = 0$$

$$h(l,k) \frac{1}{P_{lk}} + \lambda = 0$$

$$P_{\ell k} = \frac{1}{\lambda} h(\ell, k)$$

Since $\sum_m P_{\ell m} = 1 \Rightarrow \lambda = \sum_m h(\ell, m)$

$$\hat{P}_{\ell k} = \frac{h(\ell, k)}{\sum_{k=0}^{M-1} h(\ell, k)}$$

$$\sigma_{NC}^2 = \frac{\sigma_c^2}{1+\rho^2}$$

$$\delta_n - g_n = \frac{(\delta_n - h_n) * (\delta_n - h_{-n})}{1+\rho^2}$$

$$g_n = (\delta_{n-1} - \delta_{n-1}) \frac{\rho}{1+\rho^2}$$