Homework 4

Q1

(a)

	# Units	# Weights	# Connections
Convolution Layer 1	290 400	34 848	105415 200
Convolution Layer 2	186 624	614 400	447 897 600
Convolution Layer 3	64 896	884 736	149 520 384
Convolution Layer 4	64 896	1327 104	224 280 576
Convolution Layer 5	43 264	884 736	149 520 384
Fully Connected Layer 1	4096	177 209 344	177 209 344
Fully Connected Layer 2	4096	16 777 216	16 777 216
Output Layer	1000	4096 000	4096 000

(b)

i. we can decrease the number of units in Fully Connected Layers.

ii. we can redue the number of kernels.

Q2

(a)

$$P(y=k \mid x, \mu, \sigma) = \frac{P(y=k) \cdot P(x \mid y=k, \mu, \sigma)}{P(x \mid \mu, \sigma)}$$

$$= \frac{P(y=k) \cdot P(x \mid y=k, \mu, \sigma)}{\sum_{j=1}^{k} P(x, y=j \mid \mu, \sigma)}$$

$$= \frac{\alpha_k \cdot \left(\prod_{d=1}^{D} 2\pi \sigma_d^2\right)^{-1/2} \exp\left(-\sum_{d=1}^{D} \frac{1}{2\sigma_d^2}(x_d - \mu_{kd})^2\right)}{\sum_{j=1}^{k} \alpha_j \left(\prod_{d=1}^{D} 2\pi \sigma_d^2\right)^{-1/2} \exp\left(-\sum_{d=1}^{D} \frac{1}{2\sigma_d^2}(x_d - \mu_{jd})^2\right)}$$

(b) $$\ell(A, D) = -\log \prod_{n=1}^{N} P(x^{(n)} \mid y^{(n)}, A) \cdot P(y^{(n)})$$

$$= -\log \prod_{j=1}^{K} \alpha_j^{\sum_{n=1}^{N} \mathbb{I}(y^{(n)} = j)} \left(\prod_{d=1}^{D} 2\pi \sigma_d^2\right)^{-N/2} \exp\left(-\sum_{d=1}^{D} \frac{1}{2\sigma_d^2} \sum_{n=1}^{N} (x_d^{(n)} - \mu_{y^{(n)}d})^2\right)$$

$$= -\sum_{j=1}^{K} \sum_{n=1}^{N} \mathbb{I}(y^{(n)} = j) \log \alpha_j + \frac{N}{2} \sum_{d=1}^{D} \log 2\pi \sigma_d^2 + \sum_{d=1}^{D} \frac{1}{2\sigma_d^2} \sum_{n=1}^{N} (x_d^{(n)} - \mu_{y^{(n)}d})^2$$

(c)

$$\frac{\partial l}{\partial \mu_{kd}} = -\frac{1}{\sigma_d^2}\sum_{n|y^{(n)}=k}(x_d^{(n)} - \mu_{kd})$$

$$= -\frac{1}{\sigma_d^2}\left(\sum_{n|y^{(n)}=k} x_d^{(n)} - N_k \mu_{kd}\right)$$

Let $\frac{\partial l}{\partial \mu_{kd}} = 0$

$$\Rightarrow \text{MLE } \mu_{kd} = \frac{\sum_{n|y^{(n)}=k} x_d^{(n)}}{N_k}$$

$$\frac{\partial l}{\partial \sigma_d^2} = \frac{N}{2\sigma_d^2} - \frac{1}{2(\sigma_d^2)^2}\cdot\sum_{n=1}^{k}\left(x_d^{(n)} - \mu_{y^{(n)}d}\right)^2$$

Let $\frac{\partial l}{\partial \sigma_d^2} = 0$

$$\Rightarrow \text{MLE } \sigma_d^2 = \frac{\sum_{n=1}^{N}\left(x_d^{(n)} - \text{MLE } \mu_{y^{(n)}d}\right)^2}{N}$$

(d) $\sum_{j=1}^{k} \alpha_j = 1 \Rightarrow \sum_{j=1}^{k} \alpha_j - 1 = 0$ (constraint)

$$\Rightarrow f(A) = L - \lambda\left(\sum_{j=1}^{k} \alpha_j - 1\right)$$

$$\Rightarrow \frac{\partial f}{\partial \alpha_k} = -\frac{\sum_{n=1}^{N} I(y^{(n)}=k)}{\alpha_k} - \lambda$$

let $\frac{\partial f}{\partial \alpha_k} = 0$

$$\Rightarrow \text{MLE } \alpha_k = -\frac{\sum_{n=1}^{N} I(y^{(n)}=k)}{\lambda}$$

$$= -\frac{N_k}{\lambda}$$

Considering constraint, we get $\lambda = -N$

$$\Rightarrow \text{MLE } \alpha_k = \frac{N_k}{N}$$