CSC411 A4

Q1.

a.

$N_t = WHM, P_t = K^2CM, U_t = WHK^2CM$ for convolution layer

$N_t = WHM, P_t = W^2H^2CM, U_t = W^2H^2CM$ for connected layer

Layer 2,4,5 should be computed separately and multiply by 2, because they are not connected to layer between GPUs

Layer	# of units	# weights	# connection
Convolution Layer 1	290400	34848	105415200
Convolution Layer 2	186624	307200	111974400
Convolution Layer 3	64896	884736	149520384
Convolution Layer 4	64896	663552	112140288
Convolution Layer 5	43264	442368	74760192
Fully Convolution Layer 1	4096	37748736	37748736
Fully Convolution Layer 2	4096	16777216	16777216
Output Layer	1000	4096000	4096000

b.

i

in order to reduce the number of parameters, we can reduce the kernel size of convolution layer. Because $U_t = WHK^2CM$

ii

in order to reduce the number of connections, we can reduce the kernel size to reduce number of connections.

Q2

a.

(a) $P(y=k|\vec{x},\vec{u},\vec{\sigma})$

by Bayes Rule

$$= \frac{P(\vec{x}|y=k,\vec{u},\vec{\sigma})P(y=k)}{P(\vec{x},\vec{u},\vec{\sigma})}$$

By law of total probability.

$$P(\vec{x},\vec{u},\vec{\sigma}) = \sum_k P(y=k)\cdot P(\vec{x}|y=k,u,\sigma).$$

so

$$\frac{P(\vec{x}|y=k,\vec{u},\vec{\sigma})P(y=k)}{P(\vec{x},\vec{u},\vec{\sigma})}$$

$$= \frac{P(\vec{x}|y=k,u,\sigma)P(y=k)}{\sum_k P(y=k)\cdot P(\vec{x}|y=k,u,\sigma)}$$

$$= \frac{\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left\{-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}(x_i-u_{ki})^2\right\}\alpha_k}{\sum_k\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left\{-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}(x_i-u_{ki})^2\right\}\alpha_k}$$

b.

(b) $-\log p(y^{(1)}, x^{(1)} \cdots y^{(N)}, x^{(N)} | \theta)$

$$= -\log \prod_{i=1}^{N} p(y^i, x^i | \theta)$$

$$= -\left(\sum_{i=1}^{N} \log p(y^i) + \sum_{i=1}^{N} \log p(x^i | y^i, \theta)\right) \quad \text{By Naive Bayes}$$

$$= -\left(\sum_{i=1}^{N} \log \alpha_k \, \mathbb{I}(k=y) + \sum_{i=1}^{N} \log\left(\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left\{-\sum_{i=1}^{D} \frac{1}{2\sigma_i^2}(x_i - u_{ki})^2\right\}\right)\right)$$

$$= -\sum_{i=1}^{N}\left[-\frac{1}{2}\sum_{j=1}^{D}(\log 2\pi + \log \sigma_j^2) - \sum_{j=1}^{D}\frac{1}{2\sigma^2}(x_j - u_{ij}^2)\right] - \sum_{i=1}^{N} \log \alpha_k$$

c.

(c) $$\frac{\partial \ell(\theta, D)}{\partial u_{kj}} = \frac{1}{2}\sum_{i=1}^{N} \frac{1}{\sigma^2} \frac{\partial((x_j^i - u_{ij})^2)}{\partial u_{kj}}$$

$$= \frac{1}{2} \cdot \sum_{i=1}^{N} \frac{1}{\sigma^2} \cdot (-2)(x_j^{ik} - u_{kj})$$

$$= -\sum_{i=1}^{N} \frac{1}{\sigma^2}(x_j^i - u_{kj})$$

$$= -\sum_{i=1}^{N} \mathbb{1}(y^i = k)(x_{ij} - u_{kj}) \frac{1}{\sigma_j^2},$$

$$\frac{\partial l(\theta, D)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left(\sum\sum -\frac{1}{2}\log 2\pi\sigma_i^2 + \frac{1}{2}\sum\sum \frac{1}{\sigma^2}(x_j^i - u_{yj})^2 \right)$$

$$= -\frac{N}{2\sigma^2} + \sum_{i=1}^{N}(x_{ij} - u_{kj})^2 \cdot \frac{1}{2\sigma_j^4}$$

Let $\frac{\partial l(\theta, D)}{\partial u_{kj}} = 0 \qquad \frac{\partial l(\theta, D)}{\partial \sigma^2} = 0.$

Then $\sigma^2 = \frac{1}{N}\sum (X^i - u_{y(i)})^2$

$$0 = \sum_{i=1}^{N} \mathbb{1}(y^i = k)(x_{ij} - u_{kj})\frac{1}{\sigma_j^2},$$

Then $\sum_{i_k=1}^{N_k}(x_j^{ik} - u_{kj}) = 0$

$N \cdot u_{kj} = \sum_{i=1}^{N} x_j^{ik}$

Then $u_{kj} = \frac{1}{N_k}\sum_{i_k=1}^{N_k} x_j^{ik}$

Then $u = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(y^{(i)} = k)\, x^{(i)}$

d.

(d) $L(\alpha_{y_1}, \alpha_{y_2} \cdots \alpha_{y_N}, \lambda)$

$$= \sum_{i=1}^{N} \log \alpha_{y_i} - \lambda \left(\sum_k P(y=k) - 1 \right)$$

$$= \sum_{i=1}^{N} \log \alpha_{y_i} - \lambda \left(\sum_k \alpha_k - 1 \right)$$

$$\frac{\partial L}{\partial \alpha_j} = \sum_{i=1}^{N} \left(\frac{\partial \log \alpha_{y_i}}{\alpha_j} \right) - \lambda \frac{\partial \sum_k \alpha_k}{\alpha_j}$$

$$= 0$$

Then $\sum_{i=1}^{N} \left(\frac{\partial \log \alpha_{y_i}}{\alpha_j} \right) = \lambda = \frac{1}{\alpha_j} \sum_{i=1}^{N} \mathbb{1}(y^i = k)$

Then $\frac{1}{\alpha_j} \sum_{i=1}^{N} \mathbb{1}(y^{(i)} = j) = \lambda$

Then $\alpha_j = \frac{1}{\lambda} \sum_{i=1}^{N} \mathbb{1}(y^i = j)$

Then $\sum_{j=1}^{K} \alpha_j = \frac{1}{\lambda} \sum_{j=1}^{K} \sum_{i=1}^{N} \mathbb{1}(y^{(i)} = j) = 1$

Then $\lambda = \sum_{j=1}^{K} \sum_{i=1}^{N} \mathbb{1}(y^{(i)} = j) = N$

So $\frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(y^{(i)} = k)$ is the MLE for α_k