Hw4

2018年10月31日 星期三　　下午5:06

1a)

	# Units	# Weights	# Connections
Convolution Layer 1	290400	34848	105415200
Convolution Layer 2	186624	307200	223948800
Convolution Layer 3	64896	884736	149520384
Convolution Layer 4	64896	663552	112140288
Convolution Layer 5	43264	442368	74760192
Fully Connected Layer 1	4096	177209344	177209344
Fully Connected Layer 2	4096	16777216	16777216
Output Layer	1000	4096000	4096000

b) i) In order to reduce the parameters for the network, we can reduce filter for each layers. Therefore, we reduce total number of weight in this model.

ii) Reduce height and width of an input going to convolutional layer. Since there is total of WHK^2CM connections. Reduce W and H can reduce total connections.

2 a)
$$P(y=k \mid x,\mu,\sigma) = \frac{P(y=k,x,\mu,\sigma)}{P(x,\mu,\sigma)}$$
$$= \frac{P(x \mid y=k,\mu,\sigma)P(y=k,\mu,\sigma)}{P(x \mid \mu,\sigma)\,P(\mu,\sigma)}$$
$$= \frac{P(x \mid y=k,\mu,\sigma)\,P(y=k \mid \mu,\sigma)\,P(\mu,\sigma)}{P(x \mid \mu,\sigma)\,P(\mu,\sigma)}$$
$$= \frac{P(x \mid y=k,\mu,\sigma)\,P(y=k)}{\sum_k P(y=k)\,P(x \mid y=k,\mu,\sigma)}$$
$$= \frac{\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-1/2} \exp\left\{-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2\right\}\alpha_k}{\sum_k\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-1/2} \exp\left\{-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2\right\}\alpha_k}$$

b)
$$l(\theta;D) = -\log P(y^{(1)},x^{(1)} \ldots\ldots y^{(N)},x^{(N)} \mid \theta)$$
$$= -\log P(y^{(1)},x^{(1)} \mid \theta) \ldots - \log P(y^N,x^N \mid \theta)$$
$$= -\log \prod^{N} P(y^i,x^i \mid \theta)$$
$$= -\log \prod^{N} \frac{P(x^i,y^i,\theta)}{P(\theta)}$$
$$= -\log \prod^{N} \frac{P(x^i \mid y^i,\theta)\,P(y^i \mid \theta)\,P(\theta)}{P(\theta)}$$
$$= -\log \prod^{N} P(x^i \mid y^i,\theta)\,P(y^i \mid \theta)$$
$$= -\sum_{i=1}^{N} \log P(x^i \mid y^i,\theta)\cdot P(y^i \mid \theta)$$
$$= -\sum_{i=1}^{N} \log P(x^i \mid y^i,\theta) - \sum_{i=1}^{N} \log P(y^i \mid \theta)$$
$$= -\sum\left[\log\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right)^{-1/2} \exp\left\{-\sum_{i=1}^{D}\mathbb{1}[y^i=k]\frac{1}{2\sigma_i^2}(x_i-\mu_{kj})^2\right\}\right] - \sum \log\alpha_k$$
$$= -\sum\left[-\frac{1}{2}\log\left(\prod_{i=1}^{D} 2\pi\sigma_i^2\right) + \log\left(\exp\left\{-\sum_{i=1}^{D}\mathbb{1}[y^i=k]\frac{1}{2\sigma_i^2}(x_i-\mu_{kj})^2\right\}\right)\right] - \sum \log\alpha_k$$
$$= \sum\left[\frac{1}{2}\sum\log(2\pi) + \frac{1}{2}\sum\log(\sigma_i^2) + \sum\mathbb{1}[y_i=k]\frac{1}{2\sigma_i^2}(x_i-\mu_{kj})^2\right] - \sum\log\alpha_k$$
$$= \sum\left[\frac{1}{2}\sum\log(2\pi) + \frac{1}{2}\sum\log(\sigma_i^2) + \sum\mathbb{1}[y_i=k]\frac{1}{2\sigma_i^2}(x_i-\mu_{kj})^2 - \log\alpha_k\right]$$

c) From question (b).

We know that $l(\theta;D) = \sum\left[\frac{1}{2}\sum\log(2\pi) + \frac{1}{2}\sum\log(\sigma_i^2) + \sum\mathbb{1}[y_i=k]\frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2 - \log\alpha_k\right]$

Therefore:
$$\frac{\partial l(\theta;D)}{\partial\mu_{kj}} = \sum\left[\frac{1}{2}\sum\frac{\partial l}{\partial\mu_{kj}}\log(2\pi) + \frac{1}{2}\sum\frac{\partial l}{\partial\mu_{kj}}(\sigma_i^2) + \sum\mathbb{1}[y_i=k]\frac{\partial l}{\partial\mu_{kj}}\frac{1}{2\sigma_i^2}(x_i-\mu_{kj})^2 - \frac{\partial l}{\partial\mu_{kj}}\log\alpha_k\right]$$
$$= \sum\left[0 + 0 - \mathbb{1}[y_i=k]\frac{1}{\sigma_j^2}(x_{ij}-\mu_{kj}) - 0\right]$$
$$= -\sum\mathbb{1}[y_i=k]\frac{1}{\sigma_j^2}(x_{ij}-\mu_{kj}) \qquad ①$$

$$\frac{\partial l(\theta;D)}{\partial\sigma_j^2} = \frac{\partial}{\partial\sigma_j^2}\left[\sum\left[\frac{1}{2}\sum\log(2\pi) + \frac{1}{2}\sum\log(\sigma_i^2) + \sum\mathbb{1}[y_i=k]\frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2 - \log\alpha_k\right]\right]$$
$$= \sum_{i=1}^{N}(x_{ij}-\mu_{kj})^2\frac{1}{2\sigma_j^4} - \frac{N}{2\sigma_j^2}$$

From ①, setting the $\frac{\partial l(\theta;D)}{\partial\mu_{kj}} = 0$, we have
$$-\sum\mathbb{1}[y_i=k]\frac{1}{\sigma_j^2}(x_{ij}-\mu_{kj}) = 0$$
$$\sum\mathbb{1}[y_i=k](x_{ij}-\mu_{kj}) = 0$$
$$\therefore \hat{\mu}_{ML} = \frac{1}{N}\sum_{i=1}^{N}x_{ij}$$

From ②, setting the $\frac{\partial l(\theta;D)}{\partial\sigma_j^2} = 0$, we have
$$\sum_{i=1}^{N}(x_{ij}-\mu_{kj})^2\frac{1}{2\sigma_j^4} - \frac{N}{2\sigma_j^2} = 0$$
$$-\frac{N}{2}\sigma_j^2 + \frac{1}{2}\sum_{i=1}^{N}(x_{ij}-\mu_{kj})^2 = 0$$
$$\therefore \hat{\sigma}_{ML} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ij}-\mu_{kj})^2}$$

d) In this case, we have $f(\alpha_k) = \sum\log\alpha_k$ and constrain of $g(\alpha_k) = \sum\alpha_k - 1$, since $\sum\alpha_k = 1$.

Apply Lagrange multipliers. we have.
$$\therefore \nabla f(\alpha_k) = \lambda\nabla g(\alpha_k)$$
$$\therefore \nabla f(\alpha_k) - \lambda\nabla g(\alpha_k) = 0$$
$$\therefore \frac{\partial}{\partial\alpha_k}\left(\sum\log\alpha_k\right) - \lambda\frac{\partial}{\partial\alpha_k}\left(\sum\alpha_k - 1\right) = 0$$
$$\therefore \sum\frac{1}{\alpha_k} - \lambda = 0$$
$$\therefore \alpha_k = \frac{\sum\mathbb{1}[y^i=k]}{\lambda}$$

take it back to constraint, we have
$$\sum_{i=1}^{K}\frac{\sum_{i=1}^{N}\mathbb{1}[y^i=k]}{\lambda} = 1$$
$$\therefore \frac{1}{\lambda}\sum_{i=1}^{K}\sum_{i=1}^{N}\mathbb{1}[y^i=k] = 1$$
$$\therefore \lambda = \sum_{i=1}^{K}\sum_{i=1}^{N}\mathbb{1}[y^i=k]$$
$$\therefore \lambda = N$$

Substitu λ back to α_k. we have the final form
$$\alpha_k = \frac{1}{N}\sum\mathbb{1}[y^i=k]$$