Vele Tosevski

1002174657

CSC 411

March 8, 2019

**HW4 Write-Up**

1. a)

| Layers | # Units | # Weights | # Connections |
|---|---|---|---|
| Convolution Layer 1 | 290 400 | 34 848 | 105 415 200 |
| Convolution Layer 2 | 186 624 | 307 200 | 223 948 800 |
| Convolution Layer 3 | 64 896 | 884 736 | 149 520 384 |
| Convolution Layer 4 | 64 896 | 663 552 | 112 140 288 |
| Convolution Layer 5 | 43 264 | 442 368 | 74 760 192 |
| Fully Connected Layer 1 | 4 096 | 37 748 736 | 37 748 736 |
| Fully Connected Layer 2 | 4 096 | 16 777 216 | 16 777 216 |
| Output Layer | 1 000 | 4 096 000 | 4 096 000 |

*Table 1: AlexNet Network Size*

b)

i) To reduce the amount of parameters, we can reduce the size of the filters and double the amount of filters at each convolution layer. For example, two 3x3 filters constitute less parameters than one 5x5 filter. Taking for example layer 2 seen below in figure 1, if we change the filter to two 3x3 filters, we get a 3x3 filter and 256 x 2 = 512 kernels. Therefore, no. parameters = (3 x 3 x 48 x 256) x 2 = 221 184 parameters < 307 200 parameters! We could also remove the two fully-connected layers and only have convolution layers. This decreases the total amount of weights drastically.

ii) Similarly, by removing the fully-connected layers, you remove a drastic amount of connections making the test run time much less than before.

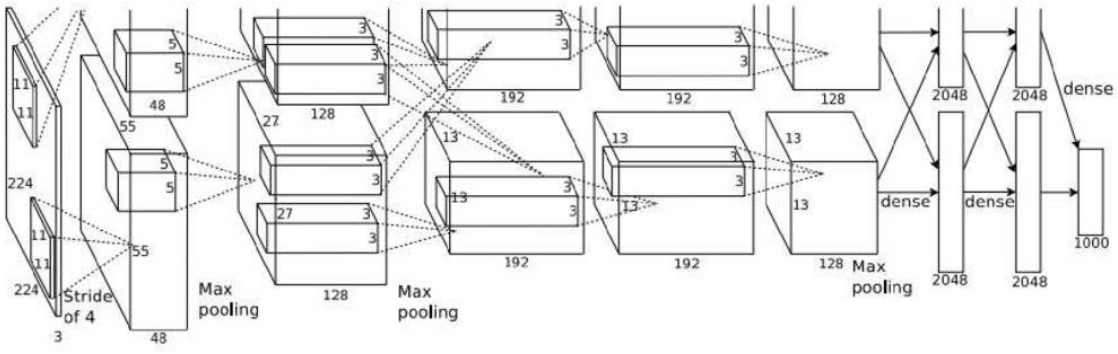


*Figure 1: AlexNet Architecture*

2. a). $p(x|\mu,\sigma) = \dfrac{p(x,\mu,\sigma)}{p(\mu,\sigma)}$

$$= \frac{\sum_K p(y_k, x, \mu, \sigma)}{p(\mu,\sigma)}$$

$$= \frac{\sum_K p(x|y_k,\mu,\sigma)\, p(y_k,\mu,\sigma)}{p(\mu,\sigma)} \quad (1)$$

$p(x,\mu,\sigma) = p(x|\mu,\sigma)\, p(\mu,\sigma)$

$$= \frac{\sum_K p(x|y_k,\mu,\sigma)\, p(y_k,\mu,\sigma)}{p(\mu,\sigma)}\; p(\mu,\sigma)$$

$$= \sum_K p(x|y_k,\mu,\sigma)\, p(y_k,\mu,\sigma) \quad (2)$$

$p(y=k, x, \mu, \sigma) = p(x|y=k,\mu,\sigma)\, p(y=k,\mu,\sigma)$ (3)

$\therefore\ p(y=k|x,\mu,\sigma) = \dfrac{p(y=k,x,\mu,\sigma)}{p(x,\mu,\sigma)}$ ← (3) ← (2)

$$= \frac{p(x|y=k,\mu,\sigma)\, p(y=k,\mu,\sigma)}{\sum_K p(x|y=k,\mu,\sigma)\, p(y=k,\mu,\sigma)} \quad (4)$$

Note: $p(y=k)$ is independent of $p(\mu,\sigma)$ since it is the prior belief of what $p(y=k)$ is and is not dependent on $\mu$ or $\sigma$.

$$p(y=k, \mu, \sigma) = p(y=k)\, p(\mu, \sigma) \quad (5)$$

(5) → (4)

$$= \frac{p(x \mid y=k, \mu, \sigma)\, p(y=k)\, \cancel{p(\mu, \sigma)}}{\sum_K p(x \mid y=k, \mu, \sigma)\, p(y=k)\, \underbrace{p(\mu, \sigma)}_{\text{constant}}}$$

$$= \frac{p(x \mid y=k, \mu, \sigma)\, p(y=k)}{\sum_K p(x \mid y=k, \mu, \sigma)\, p(y=k)}$$

... plugging in known definitions from the handout ...

$$p(y=k \mid x, \mu, \sigma) = \frac{\left(\prod_{d=1}^{D} 2\pi\sigma_d^2\right)^{-1/2} \exp\left(-\sum_{d=1}^{D} \frac{1}{2\sigma_d^2}(x_d - \mu_{kd})^2\right) \alpha_k}{\sum_{i=1}^{K}\left[\left(\prod_{d=1}^{D} 2a\sigma_d^2\right)^{-1/2} \exp\left(-\sum_{d=1}^{D} \frac{1}{2\sigma_d^2}(x_d - \mu_{id})^2\right) \alpha_i\right]}$$

b). $p(y^{(1)}, x^{(1)}, \dots, y^{(N)}, x^{(N)} \mid \theta)$

$$= \frac{p(y^{(1)}, x^{(1)}, \dots, y^{(N)}, x^{(N)}, \theta)}{p(\theta)} \quad \text{... apply bayes rule}$$

$$= \frac{p(\theta)\, p(y^{(1)}, x^{(1)} \mid \theta) p(y^{(2)}, x^{(2)} \mid \theta) \cdots p(y^{(N)}, x^{(N)} \mid \theta)}{p(\theta)}$$

$$= p(y^{(1)}, x^{(1)} \mid \theta) \cdots p(y^{(N)}, x^{(N)} \mid \theta) = \prod_{i=1}^{N} p(y^{(i)}, x^{(i)} \mid \theta)$$

$$\ell(\theta; D) = -\log p(y^{(1)}, x^{(1)}, \dots, y^{(N)}, x^{(N)} \mid \theta)$$

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

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

$$= -\sum_{i=1}^{N} \log \left[ \frac{p(y^{(i)}, x^{(i)}, \theta)}{p(\theta)} \right]$$

$$= -\sum_{i=1}^{N} \log \left[ \frac{p(x^{(i)} \mid y^{(i)}, \theta)\, p(y^{(i)}, \theta)}{p(\theta)} \right]$$

$$= -\sum_{i=1}^{N} \log \left[ p(x^{(i)} \mid y^{(i)}, \theta)\, p(y^{(i)} \mid \theta) \right]$$

$$= \boxed{-\sum_{i=1}^{N} \left[ \log p(x^{(i)} \mid y^{(i)}, \theta) + \log p(y^{(i)} \mid \theta) \right]}$$

... substituting in the given definitions

$$= -\sum_{i=1}^{N}\left[\log\left[\left(\prod_{d=1}^{D}2\pi\sigma_d^2\right)^{-1/2}\exp\left(-\sum_{d=1}^{D}\frac{1}{2\sigma_d^2}(x_d-\mu_{y(i)d})^2\right)\right]+\log\prod_{k=1}^{K}\alpha_k^{\mathbb{I}\{y^{(i)}=k\}}\right]$$

$$= -\sum_{i=1}^{N}\left[\frac{-1}{2}\sum_{d=1}^{D}\log(2\pi\sigma_d^2)-\sum_{d=1}^{D}\frac{1}{2\sigma_d^2}(x_d-\mu_{y(i)d})^2+\sum_{k=1}^{K}\mathbb{I}\{y^{(i)}=k\}\log\alpha_k\right]$$

$$= \boxed{\sum_{i=1}^{N}\left(\frac{1}{2}\sum_{d=1}^{D}\log(2\pi\sigma_d^2)+\sum_{d=1}^{D}\frac{1}{2\sigma_d^2}(x_d-\mu_{y(i)d})^2-\log\alpha_{y(i)}\right)}$$

c). $$\frac{\partial l}{\partial \mu_{kd}} = \sum_{i=1}^{N}\frac{\partial}{\partial\mu_{kd}}\left(\sum_{d=1}^{D}\frac{1}{2\sigma_d^2}(x_d^{(i)}-\mu_{y(i)d})^2\right)$$

$$= \sum_{i=1}^{N}\sum_{d=1}^{D}\frac{1}{2\sigma_d^2}\cdot 2(x_d^{(i)}-\mu_{y(i)d})\cdot(-1)$$

$$= -\sum_{i=1}^{N}\sum_{d=1}^{D}\frac{1}{\sigma_d^2}(x_d^{(i)}-\mu_{y(i)d})$$

$$= -\sum_{i=1}^{N}\frac{1}{\sigma_d^2}(x_d^{(i)}-\mu_{kd})\,\mathbb{I}\{y^{(i)}=k\} = 0$$

$$\sum_{i=1}^{N}(\mu_{kd}\,\mathbb{I}\{y^{(i)}=k\}-x_d^{(i)}\mathbb{I}\{y^{(i)}=k\}) = 0$$

$$\sum_{i=1}^{N}\mu_{kd}\mathbb{I}\{y^{(i)}=k\} = \sum_{i=1}^{N}x_d^{(i)}\mathbb{I}\{y^{(i)}=k\}$$

$$\mu_{kd} = \frac{\sum_{i=1}^{N}x_d^{(i)}\mathbb{I}\{y^{(i)}=k\}}{\sum_{i=1}^{N}\mathbb{I}\{y^{(i)}=k\}}$$

$$= \boxed{\frac{\sum_{i=1}^{N}x_d^{(i)}\mathbb{I}\{y^{(i)}=k\}}{N_k}}$$

$$\frac{\partial \ell}{\partial(\sigma_d^2)} = \sum_{i=1}^{N}\left(\frac{1}{2}\sum_{d=1}^{D}\frac{1}{2\sigma_d^2}\cdot 2\sigma_d^2 + \sum_{d=1}^{D}\frac{1}{2\sigma_d^4}(x_d - \mu_{y(i)d})^2\right)$$

$$= \sum_{i=1}^{N}\left(\frac{1}{2}\sum_{d=1}^{D}\frac{1}{\sigma_d^2}\cdot 1 - \frac{1}{2}\sum_{d=1}^{D}\frac{1}{\sigma_d^4}(x_d - \mu_{y(i)d})^2\right)$$

← For specific $d$

$$= \frac{1}{2}\sum_{i=1}^{N}\left(\frac{1}{\sigma_d^2} - \frac{1}{\sigma_d^4}(x_d - \mu_{y(i)d})^2\right) = 0$$

$$\frac{N}{\sigma_d^2} = \frac{1}{\sigma_d^4}\sum_{i=1}^{N}(x_d - \mu_{y(i)d})^2$$

$$\boxed{\sigma_d^2 = \frac{1}{N}\sum_{i=1}^{N}(x_d - \mu_{y(i)d})^2}$$

d). $g(a_k) = \sum_{k=1}^{K} a_k = 1$

$f(a_k) = \ell - \lambda(g(a_k) - 1)$

$$\frac{\partial f}{\partial a_k} = \frac{\partial \ell}{\partial a_k} - \lambda\frac{\partial g}{\partial a_k}$$

$$= \left(\sum_{i=1}^{N}\frac{1}{a_k}\mathbb{I}\{y^{(i)} = k\}\right) - \lambda = 0$$

$$\frac{\partial \ell}{\partial a_k} = \sum_{i=1}^{N}\frac{\partial}{\partial a_k}\log a_{y(i)}$$

$$= \sum_{i=1}^{N}\frac{1}{a_k}\mathbb{I}\{y^{(i)} = k\}$$

$$\lambda = \sum_{i=1}^{N}\frac{1}{a_k}\mathbb{I}\{y^{(i)} = k\} \rightarrow \lambda a_k = \sum_{i=1}^{N}\mathbb{I}\{y^{(i)} = k\}$$

$$\sum_{k=1}^{K} a_k\lambda = \sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{I}\{y^{(i)} = k\}$$

$$1\cdot\lambda = \sum_{i=1}^{N} 1 = N$$

$$\therefore \quad \lambda = N$$

So... $a_k = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\{y^{(i)} = k\}$

$$= \frac{N_k}{N}$$ as desired.