CSC411/2515 Homework 4 Solutions

Fall 2018

November 11, 2018

Question 1 a)

Single GPU

Layer	#Units	#Weights	#Connections
Conv1	290,400	34,848	105,415,200 (91,998,720 with typo)
Conv2	186,624	614,400	447,897,600
Conv3	64,896	884,736	149,520,384
Conv4	64,896	1,327,104	224,280,576
Conv5	43,264	884,736	149,520,384
FC1	4,096	177,209,344	177,209,344
FC2	4,096	16,777,216	16,777,216
FC3	1,000	4,096,000	4,096,000

2 GPUs

Layer	#Units	#Weights	#Connections
Conv1	290,400	34,848	105,415,200 (91,998,720 with typo)
Conv2	186,624	307,200	223,948,800 (111,974,400 with max pooling)
Conv3	64,896	884,736	149,520,384
Conv4	64,896	663,552	112,140,288
Conv5	43,264	442,368	74,760,192
FC1	4,096	177,209,344 (37,748,736 with max pooling)	177,209,344 (37,748,736 with max pooling)
FC2	4,096	16,777,216	16,777,216
FC3	1,000	4,096,000	4,096,000

(1.5 marks for getting all correct # weights)
(1 mark for getting most of the correct # weights)
(0.5 marks for getting at least one correct # weights)
(1.5 marks for getting all correct # connections)
(1 mark for getting most of the correct # connections)
(0.5 marks for getting at least one correct # connections)

b)

i) Most parameters are in the FC layers (eg. reduce # units in FC layers). (0.5 marks)
ii) Lots of ways to reduce the number of connections (eg. reduce number of kernels or increase stride) (0.5 marks).

Question 2 a)

$$p(y=k|x,\mu,\sigma) = \frac{p(x|y=k,\mu,\sigma)p(y=k)}{\sum_{j=1}^{K} p(x,y=j|\mu,\sigma)}$$

(0.5 marks)

$$= \frac{\alpha_k(\prod_{i=1}^{D} 2\pi\sigma_i^2)^{-1/2}\exp(-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}(x_i-\mu_{ki})^2)}{\sum_{j=1}^{K}\alpha_j(\prod_{i=1}^{D} 2\pi\sigma_i^2)^{-1/2}\exp(-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}(x_i-\mu_{ji})^2)}$$

(0.5 marks)

b)

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

(0.5 marks)

$$= -\log\prod_{j=1}^{K}\alpha_j^{\sum_{n=1}^{N} I(y^{(n)}=j)}(\prod_{i=1}^{D} 2\pi\sigma_i^2)^{-N/2}\exp(-\sum_{i=1}^{D}\frac{1}{2\sigma_i^2}\sum_{n=1}^{N}(x_i^{(n)}-\mu_{y^{(n)}i})^2)$$

$$= -\sum_{j=1}^{K}\sum_{n=1}^{N} I(y^{(n)}=j)\log\alpha_j + \frac{N}{2}\sum_{i=1}^{D}\log 2\pi\sigma_i^2 + \sum_{i=1}^{D}\frac{1}{2\sigma_i^2}\sum_{n=1}^{N}(x_i^{(n)}-\mu_{y^{(n)}i})^2$$

(0.5 marks for expanding the probabilities in some way)

c)

$$\frac{\partial\ell}{\partial\mu_{ki}} = -\frac{1}{\sigma_i^2}\sum_{n|y^{(n)}=k}(x_i^{(n)}-\mu_{ki}) = -\frac{1}{\sigma_i^2}(\sum_{n|y^{(n)}=k} x_i^{(n)} - n_k\mu_{ki})$$

(0.5 marks) where n_k is the number of data points in class k. You can also do this with an indicator function $I(y^{(n)}=k)$.

$$\mu_{kiMLE} = \frac{\sum_{n|y^{(n)}=k} x_i^{(n)}}{n_k}$$

(0.5 marks)

$$\frac{\partial\ell}{\partial\sigma_i^2} = \frac{N}{2\sigma_i^2} - \frac{1}{2(\sigma_i^2)^2}\sum_{n=1}^{N}(x_i^{(n)}-\mu_{y^{(n)}i})^2$$

(0.5 marks)

$$\sigma_{iMLE}^2 = \frac{\sum_{n=1}^{N}(x_i^{(n)}-\mu_{y^{(n)}iMLE})^2}{N}$$

(0.5 marks)

d)

Constraint: $\sum_{j=1}^{K}\alpha_j - 1 = 0$ so the objective becomes $f(\theta) = \ell - \lambda(\sum_{j=1}^{K}\alpha_j - 1)$ (0.5 marks for knowing the constraint)

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

$$\alpha_{kMLE} = \frac{\sum_{n=1}^{N} I(y^{(n)}=k)}{\lambda}$$

Considering the constraint, we get $\lambda = N$ so $\alpha_{kMLE} = \frac{\sum_{n=1}^{N} I(y^{(n)}=k)}{N}$ (0.5 marks)