EE526 Homework 2

Xingche Guo

October 9, 2019

Problem 1

$$\because \; f(z) = \log(1 + e^z)$$
$$\therefore \; f^{'}(z) = \frac{1}{1 + e^{-z}}$$
$$\therefore \; f^{''}(z) = \frac{e^{-z}}{(1 + e^{-z})^2} > 0 \quad \forall z \in \mathbb{R}$$

Problem 2

$$\frac{dp_j}{dz_i} = \begin{cases} \frac{\sum_{k \neq i} e^{z_k + z_i}}{(\sum_{k=1}^{n} e^{z_k})^2}, & for \; i = j \\ \frac{-e^{z_j + z_i}}{(\sum_{k=1}^{n} e^{z_k})^2}, & for \; i \neq j \end{cases}$$

Problem 3

Note that

$$\frac{dp_j}{dz_i} = \frac{\sum_{k=1}^{n} e^{z_k + z_i} I(i = j) - e^{z_j + z_i}}{(\sum_{k=1}^{n} e^{z_k})^2} = p_i I(i = j) - p_i p_j.$$

Therefore:

$$\frac{\partial J}{\partial z_i} = \sum_{j=1}^{n} \frac{\partial J}{\partial p_j} \frac{\partial p_j}{\partial z_i} = \sum_{j=1}^{n} -\frac{y_j}{p_j} \Big\{ p_i I(i = j) - p_i p_j \Big\}$$
$$= -y_i + p_i (\sum_{j=1}^{n} y_j) = p_i - y_i$$

Problem 4

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, W_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \boldsymbol{b}_0 = (0,0)^T.$$

iter 1:

$$Z_0 = W_0 X + \boldsymbol{b}_0 \mathbf{1}^T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$P_0 = \mathrm{softmax}(Z_0) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix},$$

$$\frac{\partial J}{\partial Z_0} = \frac{1}{2}(P_0 - Y) = \begin{pmatrix} -1/4 & 1/4 \\ 1/4 & -1/4 \end{pmatrix},$$

$$\frac{\partial J}{\partial W_0} = \frac{\partial J}{\partial Z_0}\frac{\partial Z_0}{\partial W_0} = \frac{1}{2}(P_0 - Y)X^T = \begin{pmatrix} -1/4 & 1/4 \\ 1/4 & -1/4 \end{pmatrix},$$

$$\frac{\partial J}{\partial \boldsymbol{b}_0} = \frac{\partial J}{\partial Z_0}\frac{\partial Z_0}{\partial \boldsymbol{b}_0} = \frac{1}{2}(P_0 - Y)\mathbf{1}^T = (0,0)^T$$

$$W_1 = W_0 - \tau \frac{\partial J}{\partial W_0} = \begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 1/4 \end{pmatrix}; \quad \boldsymbol{b}_1 = \boldsymbol{b}_0 - \tau \frac{\partial J}{\partial \boldsymbol{b}_0} = (0,0)^T$$

iter 2:

$$Z_1 = W_1 X + \boldsymbol{b}_1 \mathbf{1}^T = \begin{pmatrix} 1/4 & -1/4 \\ -1/4 & 1/4 \end{pmatrix},$$

$$P_1 = \mathrm{softmax}(Z_1) = \begin{pmatrix} 0.622 & 0.378 \\ 0.378 & 0.622 \end{pmatrix},$$

$$\frac{\partial J}{\partial Z_1} = \frac{1}{2}(P_1 - Y) = \begin{pmatrix} -0.189 & 0.189 \\ 0.189 & -0.189 \end{pmatrix},$$

$$\frac{\partial J}{\partial W_1} = \frac{1}{2}(P_1 - Y)X^T = \begin{pmatrix} -0.189 & 0.189 \\ 0.189 & -0.189 \end{pmatrix},$$

$$\frac{\partial J}{\partial \boldsymbol{b}_1} = \frac{1}{2}(P_1 - Y)\mathbf{1}^T = (0,0)^T$$

$$W_2 = W_1 - \tau \frac{\partial J}{\partial W_1} = \begin{pmatrix} 0.439 & -0.439 \\ -0.439 & 0.439 \end{pmatrix}; \quad \boldsymbol{b}_2 = \boldsymbol{b}_1 - \tau \frac{\partial J}{\partial \boldsymbol{b}_1} = (0,0)^T$$

Problem 5

Let nIter = 10000, the two hidden layers have 50 and 25 neurons, then:

Learning Rate	10^{-1}	10^{-2}	10^{-3}
Training Error Rate (normalized)	0.0049	0.0297	0.1536
Testing Error Rate (normalized)	0.1322	0.1101	0.1642
Total Time (Seconds)	41	38	39

Let nIter = 10000, the two hidden layers have 25 and 10 neurons, then:

Learning Rate	10^{-1}	10^{-2}	10^{-3}
Training Error Rate (normalized)	0.0042	0.0369	0.2834
Testing Error Rate (normalized)	0.1368	0.1192	0.2541
Total Time (Seconds)	22	22	22

red number means obtain the maximum iteration number/the algorithm fail to converge.

Problem 6

X is standardized, batch size = 256, then:

Models (nIter = 2×10^4)	Learning Rate	Training Error	Testing Error
Input $\to$ (10, Linear)	0.1	6.13%	7.52%
Input $\to$ (50, ReLU) $\to$ (10, Linear)	0.25	0.00%	3.03%
Input $\to$ (100, ReLU) $\to$ (10, Linear)	0.25	0.00%	2.60%
Input $\to$ (140, ReLU) $\to$ (10, Linear)	0.25	0.00%	2.93%
Input $\to$ (50, ReLU) $\to$ (50, ReLU) $\to$ (10, Linear)	0.1	0.13%	3.84%
Input $\to$ (100, ReLU) $\to$ (20, ReLU) $\to$ (10, Linear)	0.1	0.23%	4.05%