CS446: Machine Learning, Fall 2017, Homework 3

Name: Lanxiao Bai (lbai5)

Worked individually

Problem 1

Solution: Suppose the output of the first hidden layer is $\mathbf{h}$, then

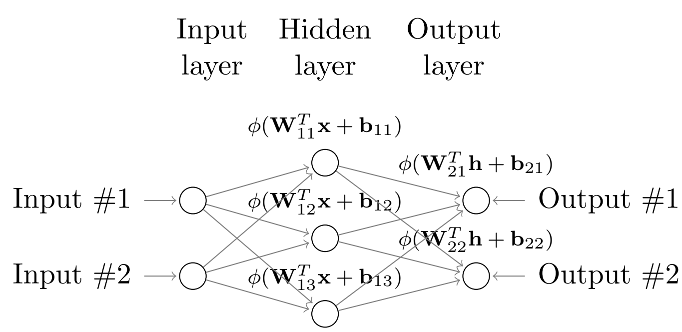

Figure 1: Neural Network when D = 2, H = 3, K = 2

Problem 2

Solution:

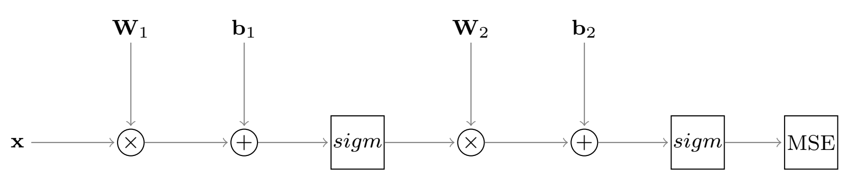

Figure 2: Computational Graph of Neural Network when D = 2, H = 3, K = 2

Problem 3

Solution:

$$f(\mathbf{x_i}) = \phi(\mathbf{W}_2^T\phi(\mathbf{x}_i\mathbf{W}_1 + \mathbf{b}_1) + \mathbf{b}_2)$$

in which $\mathbf{W}_1$ is a $d \times H$ matrix, $\mathbf{W}_2$ is a $H \times K$ matrix, $\mathbf{b}_1$ is a $1 \times H$ matrix, $\mathbf{b}_1$ is a $1 \times K$ matrix and ϕ is an element-wise sigmoid function.

Problem 4

Solution: To make this clear, we first break down the formula into the following

- $l_i = \sum_{j=1}^{d} w_{1_{ji}} x_j + b_{1_i}$
- $z_i = \phi(l_i)$
- $l'_i = \sum_{j=1}^{H} w_{1_{ji}} z_j + b_{1_i}$
- $o_i = \phi(l'_i)$
- $E = \frac{1}{2} \sum_{i=1}^{K} (y_i - o_i)^2$

then we have

$$\begin{aligned}
\frac{\partial E}{\partial w_{hk}} &= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial l'_k} \frac{\partial l'_k}{\partial w_{2_{hk}}} \\
&= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial l'_k} \frac{\partial l'_k}{\partial w_{hk}} \\
&= -(y_k - o_k) o_k (1 - o_k) z_h
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial E}{\partial b_k} &= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial l'_k} \frac{\partial l'_k}{\partial b_{2_k}} \\
&= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial l'_k} \frac{\partial l'_k}{\partial b_k} \\
&= -(y_k - o_k) o_k (1 - o_k)
\end{aligned}$$

in which $o_k = \phi(\sum_{j=1}^{H} w_{2_{jk}} \phi(\sum_{n=1}^{d} w_{1_{nj}} x_n + b_{1_j}) + b_{2_k})$ and $z_i = \phi(\sum_{j=1}^{d} w_{1_{ji}} x_j + b_{1_i})$.

Problem 5

Solution: Similarly,

$$\begin{aligned}
\frac{\partial E}{\partial w_{dh}} &= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial l'_k} \frac{\partial l'_k}{\partial z_h} \frac{\partial z_h}{\partial l_h} \frac{\partial l_h}{\partial w_{1_{dh}}} \\
&= \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial l'_k} \frac{\partial l'_k}{\partial z_h} \frac{\partial z_h}{\partial l_h} \frac{\partial l_h}{\partial w_{dh}} \\
&= -\left[\sum_k (y_k - o_k) o_k (1 - o_k) w_{hk}^{(2)} \right] [z_h (1 - z_h)] x_i
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial E}{\partial b_h} &= \frac{\partial E}{\partial o_k}\frac{\partial o_k}{\partial l'_k}\frac{\partial l'_k}{\partial z_h}\frac{\partial z_h}{\partial l_h}\frac{\partial l_h}{\partial b_{1_h}} \\
&= \frac{\partial E}{\partial o_k}\frac{\partial o_k}{\partial l'_k}\frac{\partial l'_k}{\partial z_h}\frac{\partial z_h}{\partial l_h}\frac{\partial l_h}{\partial b_h} \\
&= -\left[\sum_k (y_k - o_k)o_k(1-o_k)w^{(2)}_{hk}\right][z_h(1-z_h)]
\end{aligned}$$

in which $o_k = \phi(\sum_{j=1}^{H} w_{2_{jk}}\phi(\sum_{n=1}^{d} w_{1_{nj}}x_n + b_{1_j}) + b_{2_k})$ and $z_i = \phi(\sum_{j=1}^{d} w_{1_{ji}}x_j + b_{1_i})$.

Problem 6

Solution: If more hidden layers are added into the network, when updating weights of input-hidden layer, the calculation of gradient should take in more differential terms based on chain rule.

References