CSCI3230 Fundamentals of Artificial Intelligence
Written Assignment 1

Due date: 23:59:59 (GMT +08:00), October 16th, 2020

I. Neural Network

a) For the following single perceptron, write output unit O in terms of input units I_j's and weights w_j's where the activation function is denoted by f. Assume there is a input unit $I_0 = 1$ connecting to the neuron with weight w_0.

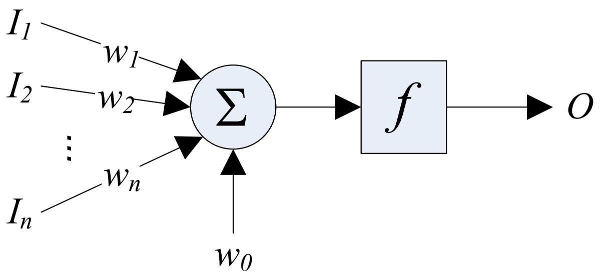

Figure 1. Single perceptron

b) For the following multiple layer perceptron (neural network), error term is $E = \frac{1}{2}\sum_{m=1}^{H_{K+1}}(O_m - T_m)^2$ where T_m is the m-th target desired output. Let $h_{i,k}$ be the output of $N_{i,k}$ and $h_{i,0} = 1$, write down the equation of $h_{i,k}$ and O_m.

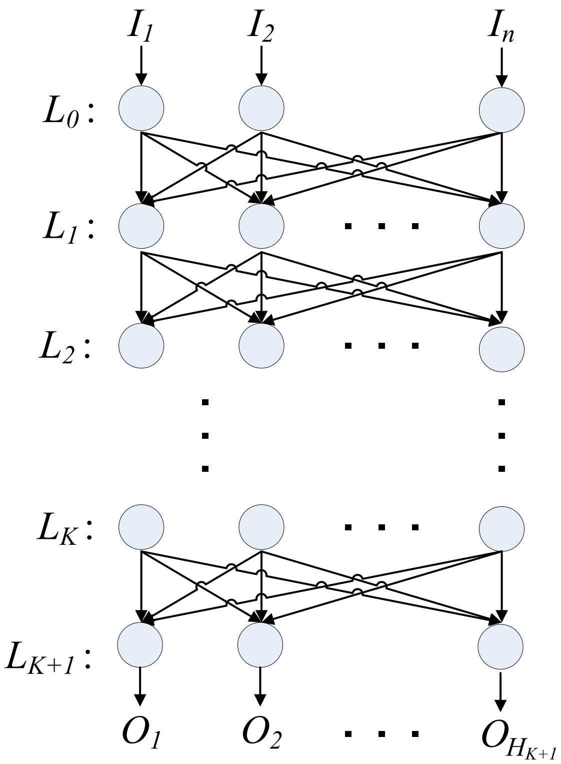

Figure 2. Multiple layer perceptron where L_0 is the Input Layer and L_{K+1} is the Output Layer. For each layer L_i, it has H_i number of neurons. For each node $N_{i,j}$, the j-th node in the layer L_i, is the single perceptron in part a. For each two nodes in adjacent layers, there is a connection with weight, $w_{i,j,k}$, which is the weight connecting $N_{i,j}$ and $N_{i+1,k}$.

c) For the activation function $f(z) = \frac{1}{1+e^{-z}}$, show that $f'(z) = f(z)(1 - f(z))$.

d) One way to iteratively minimize a smooth function $g(z)$, is to use gradient descent The rule is $x_{t+1} = x_t - \alpha g'(x_t)$ until $|x_{t+1} - x_t| < \varepsilon$, where α and ε are the learning rate and the tolerance level respectively. Explain why we need the learning rate α.

e) For fixed input, error term $E = \frac{1}{2}\sum_{m=1}^{H_{K+1}}(O_m - T_m)^2$ is also a function of $w_{i,j,k}$'s. Our goal is to find a set of $w_{i,j,k}$'s such that E is minimized.

i) For figure 3, express $\frac{\partial E}{\partial w_{K,j,k}}$ for the purple-colored output neuron in terms of the symbols in figure 3 and T_k.

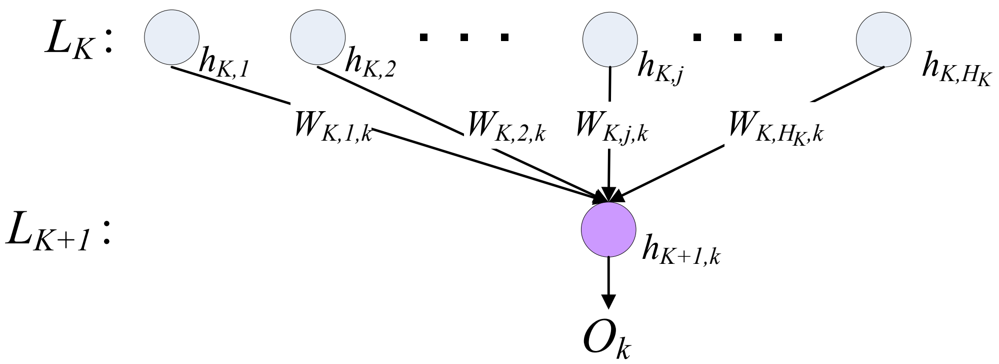

Figure 3. The bottom part of the multiple layer perceptron, where $h_{i,j}$ is the output value of Node $N_{i,j}$. The purple-colored node indicates each of the H_{K+1} nodes in the output layer L_{K+1}.

ii) Referring to the yellow-colored hidden neuron and the symbols in figure 4, show that $\frac{\partial E}{\partial h_{i+1,k}} = \sum_{m=1}^{H_{i+2}} \Delta_{i+2,\hat{m}} \cdot w_{i+1,k,\hat{m}}$, where $\Delta_{i,k} = \frac{\partial E}{\partial h_{i,k}} \cdot h_{i,k} \cdot (1 - h_{i,k})$.

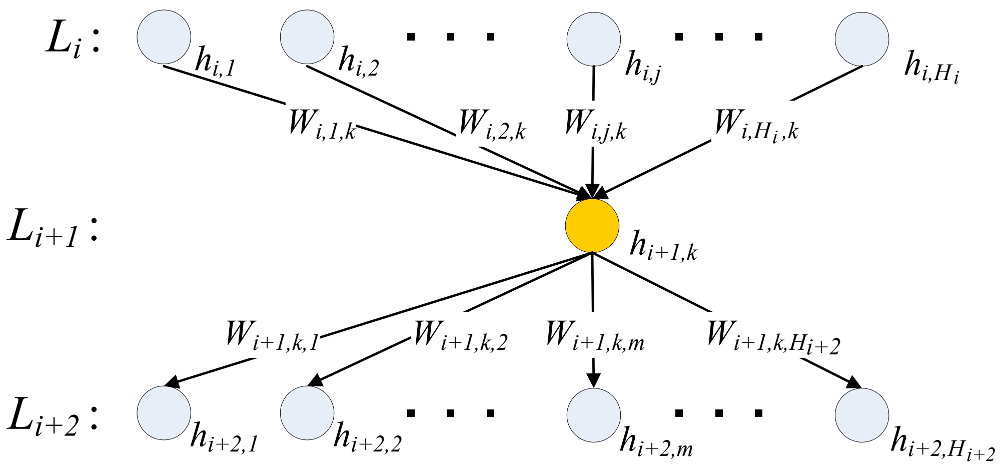

Figure 4. Intermediate three layers extracted from the multiple layer perceptron, where $h_{i,j}$ is the output value of Node $N_{i,j}$. The yellow-colored node indicates each of the H_{i+1} nodes in the output layer L_{i+1}.

iii) For figure 4, express $\frac{\partial E}{\partial w_{i,j,k}}$ for the yellow-colored hidden neuron in terms of the symbols in figure 4 and $\Delta_{v,u}$.

iv) Hence, write down the backward propagation algorithm using e(i), e(ii) and e(iii).

f) Please illustrate the problem that occur when the neural network has a large number of layers.

g) Please describe what overfitting is and how it can be avoided.

II. Assignment Submission

You **MUST** complete this assignment by using any one of the computer text editors (e.g. MS Word, WordPad, iWork Pages… etc.) and then save the document to PDF format with A4 printable page size. Scan version of the hand written work is **NOT** accepted. Please limit the file size of the PDF file less than 1MB.

You **MUST** submit the PDF file to the submission system on our course homepage (within CUHK network), otherwise, we will **NOT** mark your assignment.

III. Important Points

You MUST STRICTLY follow these points:

a. You MUST strictly follow the submission guidelines.
b. Remember to type your FULL NAME, STUDENT ID on the assignment.
c. Late submission will NOT be entertained according to our submission system settings.
d. Plagiarism will be seriously punished.

IV. Late Submission

According to the course homepage, late submission will lead to marks deduction.

No. of Days Late	Marks Deduction
1	10%
2	30%
3	60%
4 or above	100%