School of Computing and Information Systems The University of Melbourne

COMP30027 Machine Learning (Semester 1, 2021)

Tutorial: Week 11

1. Why is a perceptron (which uses a **sigmoid** activation function) equivalent to logistic regression?

2. Consider the following training set:

(x_1, x_2)	y
(0,0)	0
(0,1)	1
(1,1)	1

Consider the initial weight function as $\theta = \{\theta_0, \theta_1, \theta_2\} = \{0.2, -0.4, 0.1\}$ and the activation function of the perceptron as the step function of $f = \begin{cases} 1 & if\ \Sigma > 0 \\ 0 & otherwise \end{cases}$.

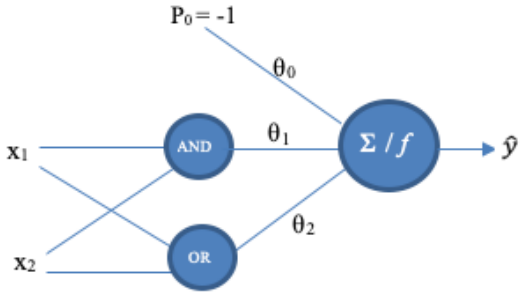
a) Draw the perceptron graph and calculate the accuracy of the perceptron on the training data before training?

b) Using the perceptron *learning rule* and the learning rate of $\eta = 0.2$. Train the perceptron **for one epoch**. What are the weights after the training?

c) What is the accuracy of the perceptron on the training data after training for one epoch? Did the accuracy improve?

3. Consider the two layers deep network illustrated below. It is composed of three perceptrons. The two perceptrons of the first layer implement the AND and OR function, respectively.

Determine the weights θ_1, θ_2 and bias θ_0 such that the network implements the XOR function. The initial weights are set to zero, i.e., $\theta_0 = \theta_1 = \theta_2 = 0$, and the learning rate η (eta) is set to 0.1.



Notes:

- The input function for the perceptron on layer 2 is the weighted sum (Σ) of its input.
- The activation function *f* for the perceptron on layer 2 is a *step function*:

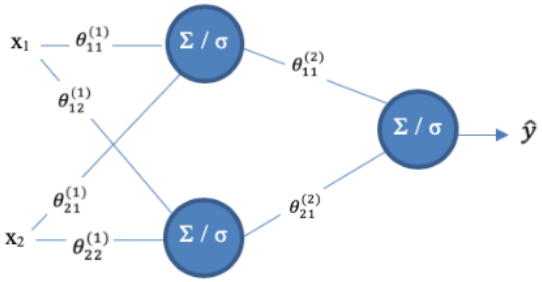
$$f = \begin{cases} 1 & if \sum > 0 \\ 0 & otherwise \end{cases}$$

- Assume that the weights for the perceptrons of the first layer are given.

4. Why is a neural network suitable for deep learning? What is significant about the representation that we attempt to learn?

5. Describe the mathematical formula of a multilayer perceptron with 1 hidden layer. Assume the input size is 1000, the hidden layer size is 100, and the output size is 20. Identify the parameters of the model, and their size.

6. [OPTIONAL] Consider the following multilayer perceptron.



The network should implement the XOR function. Perform **one** epoch of backpropagation as introduced in the lecture on multilayer perceptron.

Notes:

- The activation function *f* for a perceptron is the *sigmoid function*:

$$f(x) = \frac{1}{1 + e^{-x}}$$

- **The bias** nodes **are set to -1**. They are not shown in the network
- Use the following initial parameter values:

$$\theta_{01}^{(1)} = 2 \qquad \theta_{02}^{(1)} = -1 \qquad \theta_{01}^{(2)} = -2$$
$$\theta_{11}^{(1)} = 6 \qquad \theta_{12}^{(1)} = 8 \qquad \theta_{11}^{(2)} = 6$$
$$\theta_{21}^{(1)} = -6 \qquad \theta_{22}^{(1)} = -8 \qquad \theta_{21}^{(2)} = -6$$

- The learning rate is set to $\eta = 0.7$

i. Compute the activations of the hidden and output neurons.

ii. Compute the error of the network.

iii. *Backpropagate* the error to determine $\Delta\theta_{ij}$ for all weights θ_{ij} and updates the weight θ_{ij}.