Machine Learning - UNIT 2: Neural Networks & Multi-layer Perceptron

UNIT 2: Neural Networks

1. With examples and diagrams explain what do you understand by the following:

- **a)** McCulloch and Pitts Neurons
- **b)** Perceptron
- **c)** Perceptron Convergence Theorem
- **d)** Linear Separability

2. Write and describe the Perceptron Algorithm. What is the difference between the 'Recall' and 'Training' phase?(**Solution**:Refer 02_Perceptron_MKN.pdf & Stephen Marsland Section 3.3.3)

3. What is a perceptron? Illustrate a perceptron for AND function with a diagram.

4. Design a Perceptron that produces output of logical OR of their inputs.(**Solution**: Refer 02_Perceptron_MKN.pdf for short explanation. If asked for more marks, refer Stephen Marsland Section 3.3.4 and 3.3.5. Concept expected - **not** the Python Code)

5. The values of x and their corresponding values of y are shown in the table below

x	0	1	2	3	4
y	2	3	5	4	6

- **a)** Find the least square regression line $y = a x + b$.
- **b)** Estimate the value of y when $x = 10$.

(**Solution**: Refer https://www.analyzemath.com/statistics/linear_regression.html)

6. For a year, five randomly selected students took a math aptitude test before they began their statistics course. The Statistics Department has three questions.

- **a)** What linear regression equation best predicts statistics performance, based on math aptitude scores?
- **b)** If a student made an 80 on the aptitude test, what grade would we expect her to make in statistics?
- **c)** How well does the regression equation fit the data?

Student	x_i	y_i
1	95	85
2	85	95
3	80	70
4	70	65
5	60	70

(**Solution**: Refer https://stattrek.com/regression/regression-example.aspx)

7. State Hebb's rule. Explain how Hebb's rule improves the performance of McCulloch and Pitts Neurons.(**Solution**: Refer Stephen Marsland Section 3.1.1)

8. Consider the training examples in figure below with the attributes x1 and x2 and their target class labels.

Point	X	Y	Class
P1	1	1	+
P2	-2	-1	-
P3	2	-2	-
P4	-1	-1.5	-
P5	-2	1	+
P6	1.5	-1	+

Initial weight vector=(0, 1, 0.5) and learning rate η=0.2. Show stepwise computation of weight vector using perceptron training rule. Compute the final weight vector.

9. What is Linear Regression? List the critical assumptions of Linear Regression. Mention few applications of Linear Regression.(**Solution**: Refer 03_Perceptrons - Putting it all together_MKN.pdf).

10. Consider two perceptrons defined by the threshold expression **wo + w1x1 + w2x2 > 0.** Perceptron **A** has weight values

$w_0 = 1, \; w_1 = 2, \; w_2 = 1$

and perceptron B has the weight values

$w_0 = 0, \; w_1 = 2, \; w_2 = 1$

*Perceptron **A*** is more-general~than *Perceptron **B*** - True or False? Justify your answer.

11. What do you understand by the following terms w.r.t Perceptron?

- **a)** Learning Rate
- **b)** Bias
- **c)** Perceptron Convergence Theorem (**<u>Solution</u>**:Refer 02_Perceptron_MKN.pdf)

12. Discuss perceptron training rule and its significance in training perceptron.

13. What do you understand by the term 'Linear Separability'? Illustrate a few graphs which depicts linear and Non-Linear Separability. (**<u>Solution</u>**:Refer 02_Perceptron_MKN.pdf and Stephen Marsland Section 3.4)

14. What is the problem pointed out by Minsky and Papert regarding 'Linear Separability' and XOR Function? Give a brief overview of how it can be solved. (**<u>Solution</u>**:Refer 02_Perceptron_MKN.pdf and Stephen Marsland Section 3.4.2)

15. Design a two-input perceptron that implements the boolean function A $\wedge\neg$ B.

16. " 'Least Square Error' is a Cost Function used by the Perceptron to minimise the error". Justify this statement with an appropriate example. (**<u>Solution</u>**:Refer 02_Perceptron_MKN.pdf)

<u>UNIT 2: Multi-layer Perceptron</u>

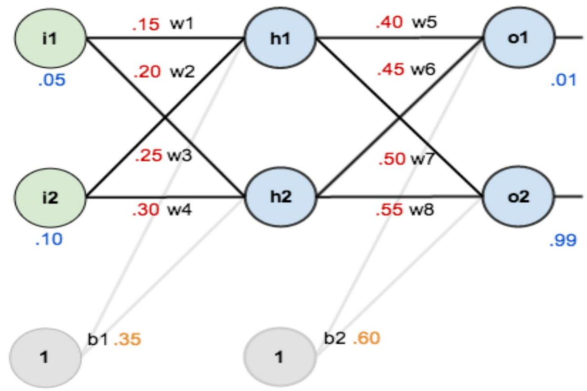
17. What do you understand by the term 'MLP(Multi-layer Perceptron'? What were the limitations of Single Layer Perceptron which led to the evolution of MLP?

18. "Multi-layer Perceptrons are the solution to the linear separability problem pointed out by Minsky and Papert in 1969". Justify this statement with proper illustrations and examples.

19. Illustrate the working of a MLP for the figure given below:

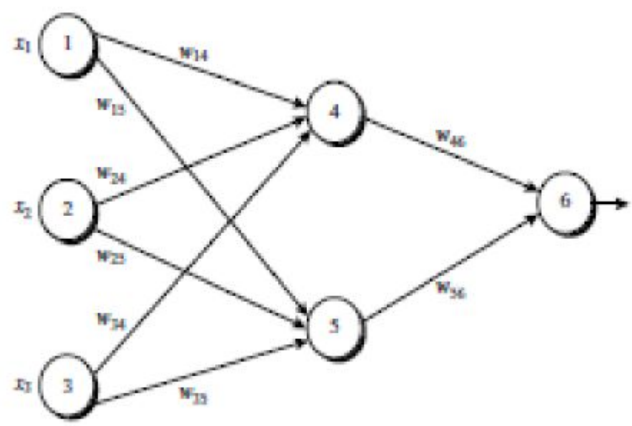
$$\frac{n\sum xy - \sum x \cdot \sum y}{n\sum x^2 - (\sum x)^2}$$

$$\bar{y} - a\bar{x}$$



Ensure that you show one ‘epoch’ (one forward pass and one backward pass). Also clearly depict the error in output and the calculate the modified weight of $\mathbf{w_5}$. Write, illustrate and explain the ‘Forward Pass Algorithm’ used in MLP.

20. Consider the multilayer feed-forward neural network shown in fig below. Let the learning rate be 0.9. The initial weight and bias values of the network are given in Table 9.1, along with the first training tuple,***X***<.1, 0, 1>, with a class label of 1.



Initial Input, Weight, and Bias Values

x_1	x_2	x_3	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	w_{46}	w_{56}	θ_4	θ_5	θ_6
1	0	1	0.2	−0.3	0.4	0.1	−0.5	0.2	−0.3	−0.2	−0.4	0.2	0.1

(a) Compute the net input and output at each node namely 4,5 and 6.

(b) Compute the error at each node namely 4,5 and 6

(c) Compute the weight and bias updates.

21. What do you understand by ‘Back Propagation of error’ in MLP? Why is called ‘Gradient Descend’ in MLP?

22. Write, illustrate and explain the ‘Backpropagation Algorithm’ used in MLP.

23. Explain the two phases in the working of an MLP for one epoch - that is Forward Pass and Back Propagation of error.

24. What do you understand by the following terms w.r.t MLP:

(i) Gradient Descend

(ii) Local Optima

(iii) Global Minima

25. What is the difference between linearly separable and non-linearly separable data? Explain with clear illustrations how MLP handles non-linearity.

26. Write the Backpropagation Algorithm in MLP. Explain with illustrations, the ‘Chain Rule’ in calculus in finding out the partial derivative of the total error (Etotal) w.r.t weights connecting the input and hidden layer, example w1. (**<u>Tip</u>**: For illustration, take a three layer MLP having two neurons at each layer. Also, the mathematical derivation is **not** expected).