CSL7020: Machine Learning - 1

End Semester Exam
IIT Jodhpur
Maximum Points: 50
Duration: 120 minutes

June 5, 2021

Note:

1. This question paper has total 7 problems. **Problem 1 is to be finished over `https://quizizz.com/` by 7:30 PM.** You have to solve and submit scanned copy handwritten answers for **3 problems by 9:10 PM. Late submission will be penalized.**
2. Please clearly write your roll number and name in answer sheet and submit via Google Classroom.
3. **Academic honesty is of highest importance at IIT-J. Any case of copying answer directly from the web or from each other may lead to "F" Grade in the course and surely 0 points in this exam.**

Problem 1: MCQ [20 points]

Link: `joinmyquiz.com`
Code: 46789178 (To be completed by 7:30 PM.)

Problem 2: Perceptron [10 points]

I Can the **complement** of the following function be represented using perceptron. If yes, what will be the decision boundary. If no, explain why. **According to last digit of your roll number, please solve correct problem.**
[For odd roll numbers]:

$$y = \overline{x_1}.\overline{x_2}.x_3 + x_1.\overline{x_2}.x_3 + x_1.x_2.\overline{x_3} + \overline{x_1}.x_2.x_3, \tag{1}$$

[For even roll numbers]:

$$y = \overline{x_1}.x_2.\overline{x_3} + x_1.\overline{x_2}.x_3 + x_1.x_2.\overline{x_3} + \overline{x_1}.x_2.\overline{x_3}, \tag{2}$$

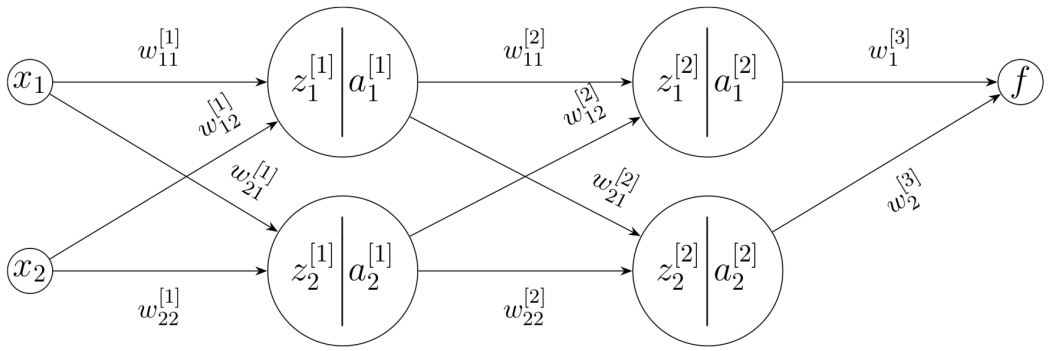
here $\cdot$ and $+$ represents OR and AND respectively. Further, $\overline{z}$ represents complement of z. *(7 points)*

II Write down three difference between MP-Neuron and perceptron.

Problem 3: Gradient Descent and Backprop [10 points]

I What is a difference between Stochastic Gradient Descent and Mini Batch Gradient Descent. *(2 points)*

II Consider a 3-layer network shown in Figure 1:



$$Z^{[1]} = \begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \end{bmatrix} = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad , \quad A^{[1]} = \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} = \begin{bmatrix} \sigma(z_1^{[1]}) \\ \sigma(z_2^{[1]}) \end{bmatrix}$$

$$Z^{[2]} = \begin{bmatrix} z_1^{[2]} \\ z_2^{[2]} \end{bmatrix} = \begin{bmatrix} w_{11}^{[2]} & w_{12}^{[2]} \\ w_{21}^{[2]} & w_{22}^{[2]} \end{bmatrix} \begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \end{bmatrix} \quad , \quad A^{[2]} = \begin{bmatrix} a_1^{[2]} \\ a_2^{[2]} \end{bmatrix} = \begin{bmatrix} \sigma(z_1^{[2]}) \\ \sigma(z_2^{[2]}) \end{bmatrix}$$

Figure 1: Three Layer Network.

Given that $f = w_1^{[3]} a_1^{[2]} + w_2^{[3]} a_2^{[2]}$. Compute following derivatives: $\frac{\delta f}{\delta z_1^{[2]}}$, $\frac{\delta f}{\delta Z^{[2]}}$, $\frac{\delta f}{\delta Z^{[1]}}$, $\frac{\delta f}{\delta w_{11}}$.*(8 points)*

Problem 4: SVM: Hyperplanes, linear-separability

I Suppose two lines passing through (10,0) and (11,0) respectively are parallel to each other and they make 45 degree from X-axis. Find out the width between these two lines. (4 Points)

II Suppose a dataset contain 100 positive and negative samples each. Further, suppose the positive samples are at radius 2, and negative samples are at radius 4 from the origin. Find out the transformation such that they become linearly separable. (4 points)

III Find out hyperplane separating positive and negative samples in above case. Write down its equation. How many support vectors will be there in this example. (2 points)

————————————————————-END OF QP————————————————————-