United International University
Department of Computer Science and Engineering
CSE 489 Machine Learning,
Mid Exam, Spring 2019
Total Marks: 30, Time: 1:45 minutes

Answer all the questions.

1. (a) You are given a task where you have to estimate the monthly income of some person in Dhaka city given their occupation, age, educational qualification, and number of siblings. Monthly income falls somewhere in the range [10,100000]. Choose a supervised learning algorithm of your choice for the task, and explain your reasons behind that choice. [2]

   (b) You have built a machine learning model given some dataset. After some analysis on the predictions made by the model, you have come to understand that the models generalization capability may increase if more training observations are provided, but collecting more samples would be extremely inconvenient due to time and resource constraints. Can you think of a workaround for this situation? [2]

   (c) You want to create an ensemble of neural networks for a problem at hand, but the model will be deployed for predicting user activity on a website in real time, and feeding observations through the ensemble will cause a delay that degrades user-experience. Propose an alternative for this situation. [2]

2. (a) You are training a neural network with stochastic gradient descent without momentum having $learning\_rate = 0.5$. As training iterations go, training loss increases rapidly instead of decreasing and becomes too large too soon. You recheck your gradient descent implementation, but do not find any visible bug. Briefly describe a potential reason. [2]

   (b) You are training a neural network having 200 fully connected hidden layers with *tanh* activation function. After 50 epochs, you check how the weights of different layers are changing for a sanity check. Something funny strikes you. The weights connected to the layers closer to the loss function are being updated, but the weights connected to the earlier layers (layer 1, layer 2, layer 20, etc.) are not changing. You do not want this to happen. Could you propose a solution to this problem? [2]

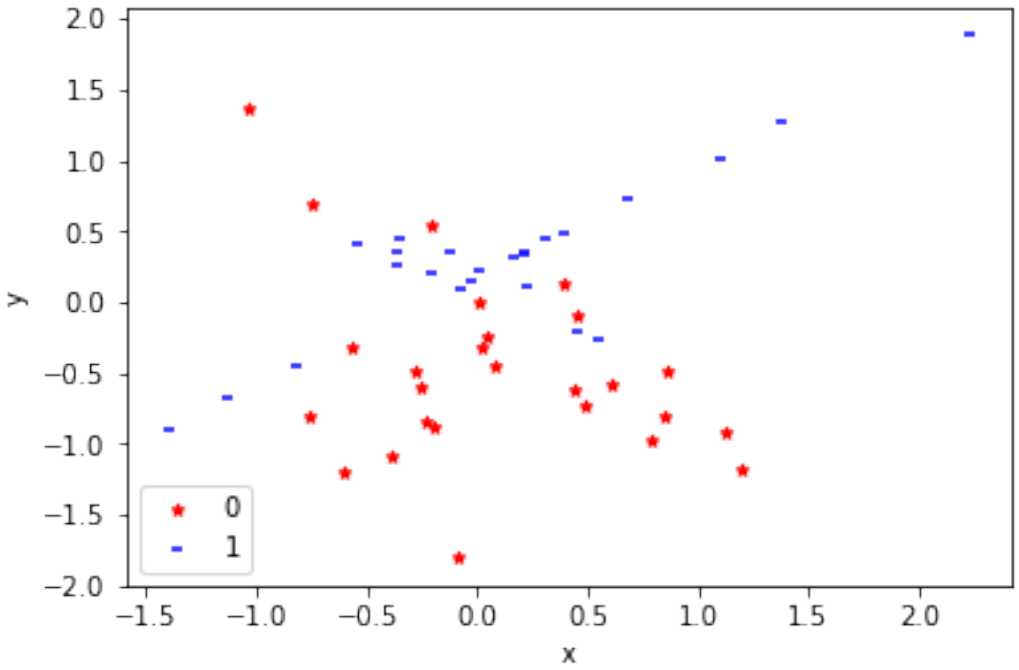
   (c) "Stochastic gradient descent has the potential of avoiding shallow local minima that batch gradient descent could have never avoided" ... Do you agree with this statement? Support your answer with adequate explanation. [2]

3. (a) "Stochastic gradient descent uses the same learning rate(step size) for all the weights which is the best way of ensuring a faster decrease in training error " . . . Do you agree with the statement? Could you propose a better solution? [2]

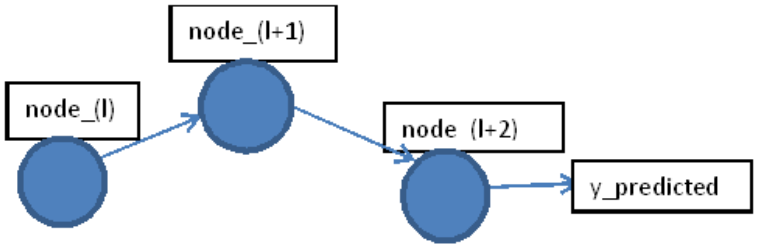
   (b) You are training a neural network on some specific dataset using batch gradient descent, and training error goes down to 0.003 after 100 epochs. Now, you choose to ignore random neurons at different epochs, and retrain the network for 100 epochs. Will training error be greater than 0.003 or less than 0.003? Explain your answer. [2]

   (c) A ML enthusiast has modified the architecture of an existing neural network by replacing ReLU everywhere with $f(x) = 2.x$. Will this modified architecture still have classification capabilities similar to that of the original architecture? Please support your answer with proper justification. [2]

4. (a) Will Logistic Regression model be a good choice for separating the 2 classes of the dataset below? Why? [2]

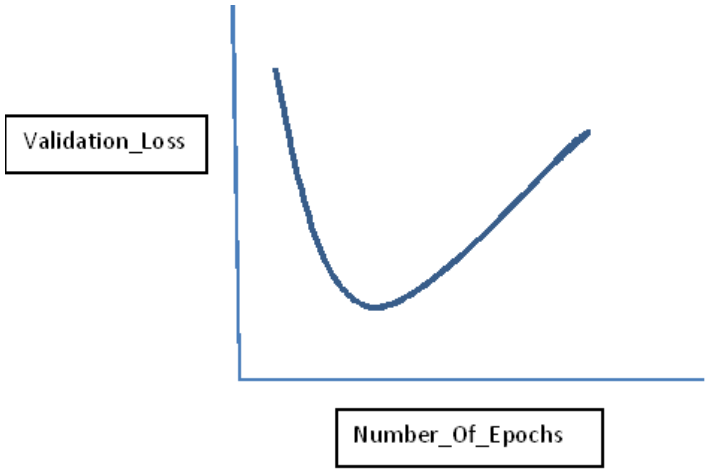


   (b) The image below shows some portions of a neural network. Suppose you are training the model using back propagation. $z^{[l]}$ is the weighted sum for node l while $z^{[l+2]}$ is the weighted sum for node $l+2$. You have to calculate gradients of the loss function with respect to both of these weighted sums. In which order would you calculate them, and why? [2]



(c) You have designed a neural network architecture for a specific problem. Early stopping mechanism has been implemented for increasing the generalization capability of the model.

Now re-draw the image below in your exam script, and specify a point on the curve where you expect the neural network to end its training due to early stopping. [2]

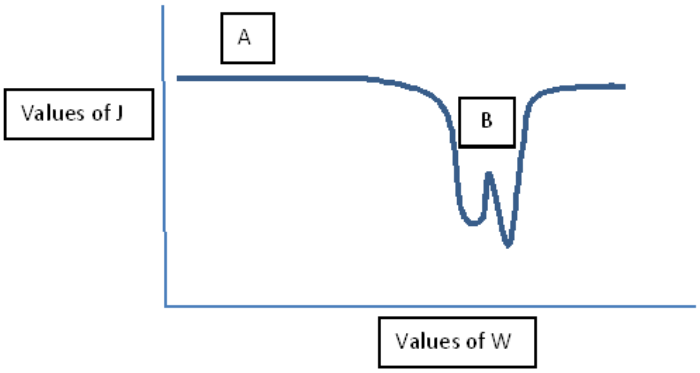


5. (a) Suppose you are using stochastic gradient descent with momentum for training a neural network.The momentum term will either accelerate weight updates, or will slow them down according to the values of some previous gradients of the weight.

In the image below, you can see a graph depicting the values of the cost function J for different values of one of the weights W. [3]

i. Will momentum accelerate or slow down weight update at region A?

ii. Will momentum accelerate or slow down weight update at region B?



(b) Keeping the weights of a neural network close to zero, in other words, keeping them as small as possible reduces the complexity of the final decision boundary, thus reducing over-fitting. How can you incorporate this fact into model training procedure? [3]