# Deep Learning Assignment -2

1. B – A function must be differentiable to perform gradient descent. Accuracy is non-differentiable, meaning it cannot be used to optimize our machine learning models.

2. Metrics can be accuracy, precision recall, f1 score. And also F1.5 and F0.5 score. The realization of F1 score is the ROC AUC curve. One will favour precision and other will favour recall. Basically, improving on False positives or False negatives helps. However, I will choose Confusion matrix and accuracy.

Confusion matrix- It provides insight into the errors that are made by the classifier and also the type of errors. It is also used to visualize important predictive analysis like precision, recall, accuracy etc. It also gives direct comparisons of true positives, true negatives, false positives and false negatives.

Accuracy- Overall measure of how much the model is correctly predicting on the entire set of data. Since I am interested in predicting the highest number of individuals in the right class accuracy is a good measure and it is simple to understand and I am used to it.

3. K-fold cross validation. Cross validated error is the best guess for the average error we get with our classification model on new data. We may also use other methods like Akaike Information Criterion- it is a measure of the goodness fit of an estimated model or Bayesian Information Criterion.

6. Back propagation is the method of computing gradients backwards in the network through recursive application of chain rule. We compute the gradient with respect to the inputs so we can chain it backwards through the layers. For sequentially computing all the gradients in SGD, a backward sweep is applied. It precisely makes use of dl/dx. We usually compute the gradient with respect to W, the linear transform (or parameter) matrix (tensor) for layer i, so that we can use it to perform a parameter update. Gradient on x can be useful for visualization, interpreting what the Neural Network might be doing and we can determine the sensitivity of the loss function with respect to these inputs. Also, the parameters of the network like inputs, weights and bias are inputs of the cost function. If we want to minimize the loss that is the output of the cost function we need to find all the parameters of the network. By back propagating through chain rule, we find the partial derivatives of the loss function with respect to its input parameters in every layer and change these parameters in the direction of their gradients.

7.A) In the figure, the first graph has AUC 1. That means it can classify perfectly between positive and negative classes. The second graph’s AUC is 0.8 – it has some predictive power or it can classify upto some extent. The third figure’s AUC is 0.5- meaning it can randomly guess (like tossing a coin) or a bad classifier. Hence, the higher the AUC the better the performance of a model.

B) Sensitivity is the proportion of positive class that is correctly classified ( True Positive Rate). Specificity refers to the False positive rate. The high the TPR the better the model performs. Similarly, the lower the FPR the better the model performs. For example, the third figure has 50% TPR and 50% FPR. It is a bad classifier and has the least predictive power. The first figure has 100% TPR and 0% FPR. That means it can the actual positives will 100% test positives. Hence, it is an exxellent classifier.