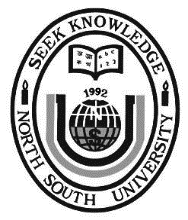
**CSE 445: Machine Learning**

**Faculty - ITN**

Midterm Marks: 45, section:4

Time : 60 Minutes

1. Suppose you have a model which detects if a bank customer is a loan defaulter or not. Here the number of true positives is 54, false positives are 12, the true negative is 47 and false negatives are 5. Calculate the precision, recall, and F1 score. Now if you have a chance to improve either precision or recall then which one will you choose to improve? Explain your choice. [6]
2. Consider the following "dataset" with a single feature x, and corresponding true label values y are shown in the table below [16]

|  |  |
| --- | --- |
| x | y |
| 1 | 4 |
| 3 | 5 |
| 5 | 7 |
| 8 | 11 |
| 10 | 14 |
| 12 | 18 |

1. Plot the data of y against x in a scatterplot [2]
2. Find the updated value of m and c up to the third iteration. [ Assume the avg line and its corresponding m and c as the first iteration value. Assume that the learning rate = 0.01. Use MSE as your cost function here] [8]
3. Find the final predicted value y\_hat for this dataset after 3rd iteration [3]
4. Compute the RMSE between the final predicted values obtained in (b) and true label values, showing all work. [3]
5. If SSfit= 30 and SSmean= 80 then find out the corresponding R2 score. Explain with an example why the higher the R2 score the better it is [4]
6. Explain with examples why logistic regression is a better model for classification tasks than linear regression. Also, write the cost function of logistic regression and explain it graphically [8]
7. In your own words, define and explain with example 2 of the following terms, in four sentences or less: [4]
8. Normalization
9. Hyperparameter
10. Semi-supervised learning
11. Explain One-hot encoding and Label Encoding. Does the dimensionality of the dataset increase or decrease after applying these techniques? [4]
12. What is stratified sampling and this is applicable for what kind of dataset. Explain with example. [3]