

Exercise 7: Analytics I: Introduction to Machine Learning

Please submit your files in a zipped file. The zipped file should be named as follows: "EntryNum_Name_Ex6.zip". The zipped file should contain your Python script, named "EntryNum_Name_Ex6.py" and the data file for the machine learning problem.

All submissions will be checked for evidence of plagiarism. Students whose submissions are found to have evidence of plagiarism will be subject to, at minimum, losing all marks for the exercise.

1. (10 marks) For the dataset provided, use ML algorithms from Python's Sklearn package to determine their classification accuracies for the dataset. Apply only the following algorithms: gradient boosting classifier, feedforward neural network, logistic regression, and support vector machine classifier. Use an 80:20 train:test split, with a random_state = 10. Set a random number seed of 1234 at the beginning of your code.

The hyperparameters for each of the above ML algorithms are given below.

Gradient boosting classifier: number of estimators = 250, random_state = 10.

Neural net classifier (SKlearn's MLPclassifier): hidden layer sizes = (150,100,30), 'relu' activation function, random_state = 10.

Logistic regression: random_state = 10.

Support vector machine classifier: gamma = 'auto', kernel = 'rbf', random_state = 10.

Report the following classification accuracy measures: the area under the receiver operating characteristic curve score, the classification accuracy, the F1 score, and the Sklearn classification report. The sample output format for two of the above algorithms (with the actual numbers erased) is shown below. Follow the output format exactly in your assignment (including placement of the dashes, spaces, newlines, etc.).

2. (5 marks) In the dataset provided, increase the number of 1s (automate this) to 750, 1250, 1750, 2250 and 3000 (change the 0s to 1s randomly). For each case, generate the AUC score, the F1 score and the classification accuracy and plot them in a single graph. Comment on how the scores change with respect to each as the number of 1s increase.

Gradient boosting classifier accuracy measures (AUC, F1 Score):

Gradient boosting classifier classification report:

	precision	recall	f1-score	support
0.0				
1.0				
accuracy				
macro avg				
weighted avg				

Neural net accuracy measures (AUC, F1 Score):

Neural net classification report:

	precision	recall	f1-score	support
0.0				
1.0				
accuracy				
macro avg				
weighted avg				
