

Lab 3

This is a lab on classification lab. There are some datasets given below:

- 1) <https://www.kaggle.com/uciml/iris>
- 2) <https://www.kaggle.com/oddrational/mnist-in-csv>
- 3) <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
- 4) <https://www.kaggle.com/rajyellow46/wine-quality>

First, you need to apply the data preprocessing steps on the dataset. You need to choose what types of data preprocessing you want to apply. See below:

<https://scikit-learn.org/stable/modules/preprocessing.html>

You need to choose one of the datasets and apply at least three classification algorithms on it. You need to write a discussion on why you have chosen these algorithms. Then you need to write a discussion on comparing their performance. In our class, we discussed Logistic Regression, SVM, Decision Tree and K-NN as our classification algorithms. You need to choose from those algorithms.

For performance measures, please see below for suggestions, you can apply other performance measures of your choice as well:

- 1) Confusion Matrix
- 2) ROC
- 3) Precision, Recall, F-measures

See below for more info:

https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics

Optional:

You can try out all 4 datasets and apply 4 classification algorithms on them to observe how it turns out.

Help:

- 1) Textbook GitHub page: <https://github.com/Akramz/Hands-on-Machine-Learning-with-Scikit-Learn-Keras-and-TensorFlow>
- 2) Example codes in modules
- 3) Learn about DataSet Loading Utilities
<https://scikit-learn.org/stable/>