

COSC 4335 “Data Mining” Assignment3 Fall 2018
Making Sense of Data—Learning and Comparing Classification Models for a Dataset
Individual Project

Due dates: Nov 9, 2018 at 11:59p

This course assignment is an opportunity for you to investigate different classification approaches; the idea is to apply different classification techniques to a challenging dataset, to compare the results, to potentially enhance the accuracy of the learnt models via selecting better parameters/preprocessing/using kernels/incorporating background knowledge and to summarize your findings in a report. You must choose one dataset from the below list and compare and interpret the results of using different approaches for the dataset.

datasets:

1. Spambase Data Set: <https://archive.ics.uci.edu/ml/datasets/Spambase>

The classification algorithms you must apply the following 4 approaches to obtain classification models for the dataset you chose from the dataset above:

1. Neural Networks
2. Support Vector Machines
3. Random Forest
4. Decision Trees

Other requirements for Assignment3:

- Accuracy of classification algorithms should be measured using 10-fold cross validation.
- In your report after comparing the experimental results, write a paragraph or two trying to explain/speculate why, in your opinion one classification algorithm outperformed the others.
- Include a brief discussion in your report, how you have selected the parameters of particular data mining algorithms (for each classifier, choose at least 3 parameters for testing and discussion).
- Finally, at the end of your report provide a 1-2 paragraphs summary that summarizes the most important findings of Assignment3

Deliverables:

- Please submit both the report and the .ipynb with all codes using blackboard.
- Name the report as <last name>_StudentID_P3.docx (or <last name>_StudentID_P3_.pdf)

Suggested packages and functions:

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import cross_val_score
```