

University of Waterloo
ECE 657A: Data and Knowledge Modeling and Analysis
Winter 2019

Homework 4: Classification I
Due: February 28th, 2019 5:00pm

Overview

Collaboration: Do your work and report individually. You can collaborate on the right tools to use and setting up your programming environment.

Hand in: One report per person, via the LEARN dropbox. Also submit the code / scripts needed to reproduce your work. Report as a PDF or a python notebook.

General Objective: To practice performing classification on a small dataset.

Specific Objectives:

- Establish your software stack to carry out data analysis homeworks, assignments and the project for the rest of the course.
- Load a simple dataset and perform some basic data preprocessing, take out only the attributes which are continuous or have distances defined.
- Divide data into train and test portions, justify your split decision.
- Perform classification and plot the results along some meaningful dimensions, analyse the results.

Tools: You can use libraries available in Python or R available to you. You need to mention which libraries you are using, any blogs or papers you used to figure out how to set carry out your calculations.

Data sets

For this homework you will use the Bank Marketing dataset:

- <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
- Download from Data Folder link, read data set description.

Tasks

In the class we talked about how to use k-Nearest Neighbours and SVM algorithms to classify supervised data.

- Load the dataset and store it as a DataFrame.
- Formulate the problem as kNN and SVM (linear) problem and run using standard libraries in your language or choice.

- Create a few plots of your model on the test data two of the data dimensions indicating the predicted elements of each class using different colours or shapes. You may need to try plotting various pairs of dimensions to see which provide some interesting result. Be sure to label your axis and legend.
- Provide a short explanation of the result you have shown and what it means.
- **For fun :** do the same for SVM with RBF kernel and compare the results for the same dimensions.