Machine Learning – Big Data Analytics

Assignment 2 – 25% of Module

Based on the sample ipython notebook files that you have been using in class you are required to complete the following exercises.

Using any datasets of your choice (but not the same ones used in the examples) you should create your own ipython notebook files for each of the following Machine Learning techniques:

a) Linear Regression
b) Polynomial Regression
c) Naive Bayes Classifier
d) KFold Cross Validation
e) Decision Trees
f) Support Vector Machine
g) K-Means Clustering
h) K-Nearest Neighbour
i) Principal Component Analysis

The ipython notebook files should document what you are doing and explain the results. The purpose of this exercise is to show that you can understand, edit and run pre-existing code samples applying ML techniques. Each file is expected to run and should explain the technique.

Each file (a) – (i) carries equal marks. Marks will be awarded for originality and for clean, readable code with good explanations.

This assignment should be submitted on **Thursday the 7th Nov 2019** by midnight. A zip file should be emailed to eoghan.furey@lyit.ie or kevin.meehan@lyit.ie (if the file is too big then a link to cloud storage may be used). Failure to submit by the deadline without a valid medical certificate will result in in a loss of 30% of the marks for this assignment. It is unfair on your class mates to expect that you can have more time when they didn't! Treat it like your job and get it in on time!

NOTE: The samples must be different from those supplied to you in blackboard. You may be asked to explain any of your code so if "re-using" snippets from other sources make sure you understand it.

Also, this must be your own work and if the same files are submitted by two students the both will receive a mark of zero!

Good luck