Assignment 5: KNNs

CSCI 297: Introduction to Machine Learning

Due: Saturday October 11th @ 11:59 PM

Professor Watson

In this assignment you will be creating a KNN or K-Nearest Neighbors classifier to determine whether the growth found within a human body is pancreatic cancer or not. There are eight different attributes and a diagnosis of whether the cancer was malignant (M) or benign (B). The KNN will group these 100 examples into two groups.

I am not concerned with how you code the KNN. You should perform some exploratory data analysis to provide evidence for feature selection, regularization, scaling, manipulation, or any other thing you do to your dataset.

Once you have a good KNN classifier, you should perform a series of tests to determine the best hyper parameterization. I would like to see some evidence of this testing; it can be a graph of accuracy given changes in hyperparameters or any other clever way you can demonstrate your testing.

You have complete freedom for this assignment in terms of how you code the assignment (no right answers per say). I am only interested in the results in terms of accuracy and precision. You may use any code from canvas or our book, but you may not use the internet.

This will be a paired programming assignment. You must work with someone else in the course. Pairs should not discuss their assignment with other pairs, please just consult one another. A large part of getting good accuracy on this assignment is manipulating the hyperparameters. I am not going to give you a marker for accuracy, you should get the best accuracy you can. Your explanations, process, steps, and analysis should all be documented and turned in along with your code. This well help determine if you took enough steps to increase your accuracy.

Good Luck!