KalaivaniKalyanAssignment10.2

Kalaivani Kalyanasundaram

8/15/2021

Introduction to Machine Learning

These assignments are here to provide you with an introduction to the "Data Science" use for these tools. This is your future. It may seem confusing and weird right now but it hopefully seems far less so than earlier in the semester. Attempt these homework assignments. You will not be graded on your answer but on your approach. This should be a, "Where am I on learning this stuff" check. If you can't get it done, please explain why.

Include all of your answers in a R Markdown report.

Plot the data from each dataset using a scatter plot.

Use ggplot to plot the scatterplot after reading the data using read.csv so that we can see the data.

Fit a k nearest neighbors' model for each dataset for k=3, k=5, k=10, k=15, k=20, and k=25. Compute the accuracy of the resulting models for each value of k. Plot the results in a graph where the x-axis is the different values of k and the y-axis is the accuracy of the model.

I could not get the accuracy because it gave me a runtime error, I tried solving it but could not figure it out.

Looking back at the plots of the data, do you think a linear classifier would work well on these datasets?

Yes because a linear classifier will split the data very well and we can see the clear split of data in this case so that we can classify easily.

How does the accuracy of your logistic regression classifier from last week compare? Why is the accuracy different between these two methods?

They are somewhat similar but thre is a slight difference in linear classifiers and what we did last week, whichh results in the different accuracy.

Calculate this average distance from the center of each cluster for each value of k and plot it as a line chart where k is the x-axis and the average distance is the y-axis.

This will be a straight line because of the clustering and classification

One way of determining the "right" number of clusters is to look at the graph of k versus average distance and finding the "elbow point". Looking at the graph you generated in the previous example, what is the elbow point for this dataset?

I am not sure on this one. This assignment is reallly hard for me, I can't seem to run the code properly.