CS487 - HW5

Cyrus Baker
April 8, 2020

Notes

This assignment uses two datasets, the iris dataset and the MNIST dataset. Because the MNIST dataset is very large I used only a portion of the entire dataset. The last column of the iris dataset contains the class label for each piece of data. The iris dataset is processed by importing it from the directory and the MNIST dataset is loaded with SKlearn.

Accuracy - total correctly classified over total instances
Precision - total correct positives over total predicted positives
Recall - total correct positives over total actual positives
F1 - 2 * (recall * precision) / (recall + precision)

I first ran the decision tree on the datasets without dimension reduction and then compared the results to what happened after applying dimension reduction.

Results

Iris				
	Decision Tree	**PCA**	**LDA**	**Kernel_PCA**
Accuracy	0.9556	0.9778	0.9778	0.8444
Precision	0.9556	0.9778	0.9778	0.8444
Recall	0.9556	0.9778	0.9778	0.8444
F1	0.9556	0.9778	0.9778	0.8444

MNIST				
	Decision Tree	**PCA**	**LDA**	**Kernel_PCA**
Accuracy	0.6427	0.4582	0.6242	0.4239
Precision	0.6427	0.4582	0.6242	0.4239
Recall	0.6427	0.4582	0.6242	0.4239
F1	0.6427	0.4582	0.6242	0.4239

Analysis

It is clear from my results that I have implemented something incorrectly but I ran out of time to figure out exactly what that was. Perhaps my parameters are not set to appropriate values, or maybe I implemented precision, recall, and f1 incorrectly because I shouldn't get the same value as the accuracy score.