Project II - Report

Tarren Engberg (engbergt) - sole member.

WARNING: my q1.py program takes significant time for values of K higher than 3 (@ 1+ minutes)

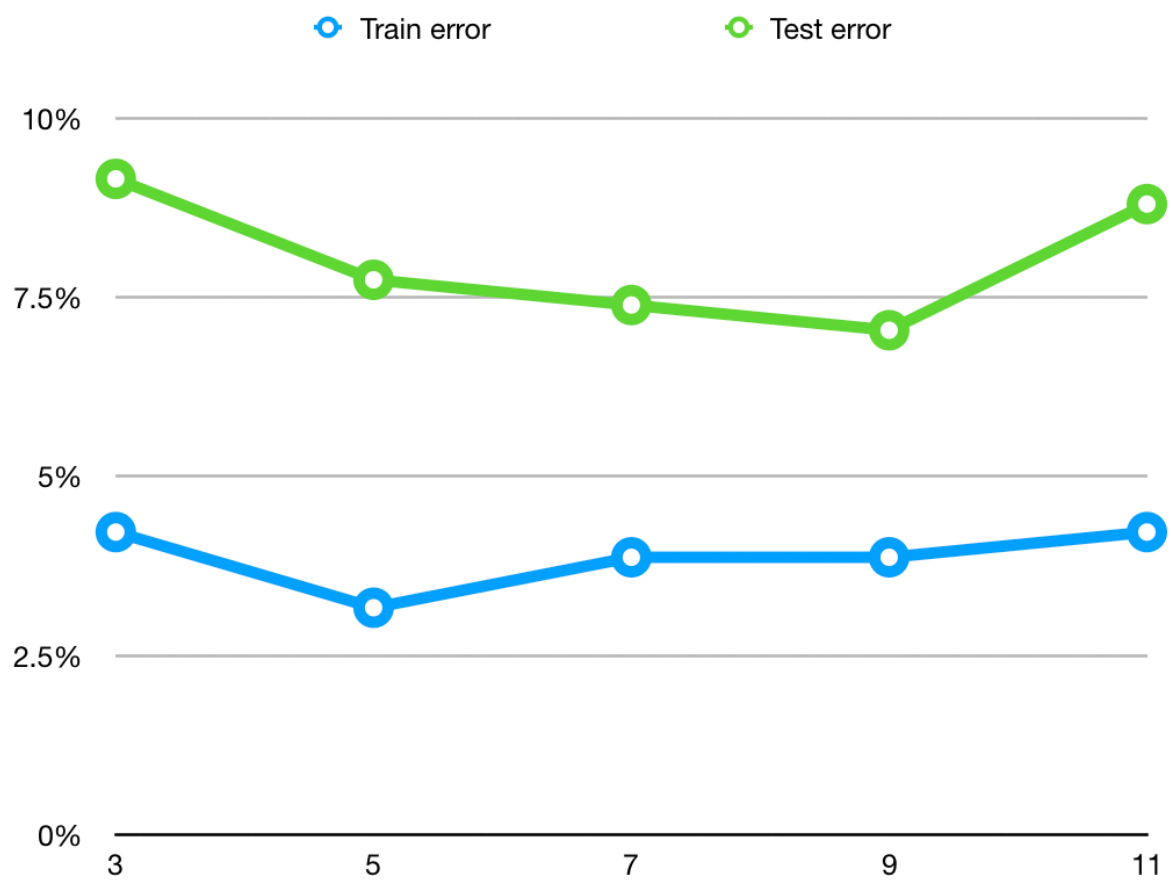
Program Instructions

Place the train and test data in the same directory as the program '*q1.py*' and use the command: "*python3 q1.py rain.csv test.csv 3*"

Part I

I found that my implementation takes a long time to run and thus I have commented out the "*leave-one-out cross-validation error*" test as it uses an extremely large K value. If I had time to refactor I would memoize each point's nth dimensional position for faster computation time. I did add in an optimization, but it wasn't enough to bring it down to a reasonable amount of time. My recommendation, for time sake, is to use a **K value of *3*** for functional verification.

K	3	5	7	9	11
Training error	4.225%	**3.169%**	3.873%	3.873%	4.225%
Testing error	9.155%	7.746%	7.394%	**7.042%**	8.803%



I find that the training error is least with a K value of 5 and the test error is least with a K value of 9.

Part II a
I got pretty far on this section and was unable to get my best test function working due to the weirdness of how numpy handles splitting 2d arrays.

Learned Decision Stump (with labels for each branch)

Computed Information Gain for the selected test

Training & Testing Error Rates

Part II b