COMP 341 Homework #5 Report / Barış Aracı 62595

Answer Q1: We normalize the histogram by dividing each value to total number of values. It is suggested because each image has different details and sizes.

Answer Q2: At early stage, there is over-fitting until k equals 5. After k is 5, both error rates of testing and training are increasing so there is under-fitting. Best is around 7-9 since both testing and training should be low.

Answer Q3: This time there is under-fitting at early stage. After c (I assume c is exponential number of 10) equals 2, over-fitting behavior was seen in the histogram. Best is when c is around 2 and 2.5 since both error rates are low.

Answer Q4: I would choose logistic regression because it has a better cross-validation score which is 0,13.

Answer Q5: Test data has higher error rate than training data before lambda equals 0,1. So, there is over-fitting. After 0.1, values of test data are still higher than training data even though it decreases. I think after 0.1, it is not clearly observable since all values are very close to each other. Average error for variation1 is 0,204 and standard deviation is 0,144 for linear regression. Average error for variation1 is 0,111 and standard deviation is 0,022 for ridge-regression. So, ridge-regression has lower values which is better.

Answer Q6: Both training and test data error rates are increasing linearly from the beginning. There is under-fitting as lambda is closer to 4 since there is obvious high rates. But before that, there is no clear over-fitting values. Average error for variation2 is 0,12169742333678646 and standard deviation is 0.008957960228846967 for linear regression. Average error for variation2 is 0.12169742333678206 and standard deviation is 0.008957960228849744 for ridge-regression. So, ridge-regression has slightly better average error and linear regression has slightly better standard deviation since those values are lower.

Answer Q7: In this histogram, there is always over-fitting since error rates of training data are low and error rates of test data are high. Average error for airfoil is 0.2777234258443021 and standard deviation is 0.014553491594176845 for linear regression. Average error for airfoil is 0.27771722491099593 and standard deviation is 0.0144902279354525 for ridge-regression. So, ridge-regression has slightly lower values which is better.