Eda Gür

50488

Comp 341 - Introduction to Artificial Intelligence HW5

1-) Because each picture has a different number of pixels, but we want all pictures to have equal weights. If picture A has 10 pixels picture B has, let’s say, 100 pixels, the saturation histogram of B will have a higher contibution in the training. If both have 1/10 of their pixels in the first bin, A will have 1 in the first bin and B will have 10 in the first bin. So, when we are using their features, B will have higher weight. We don't want this. So, we normalize for each picture to contribute equally to the algorithms. Otherwise, pictures with higher resolution would have higher importance.

Note for the following questions 2: My results for the KNN algorithm differs from the given results. I will make my comments based on the graph that the professor provided.

2-) For k values up to (not including) 9, there is overfitting, because the error is small for the training set but greater for the test set. This shows that the algorithm works well for the training set but not so well for the test case. The same is true for k = 11 and k = 15. We must instead be able to generalize from the training set to the test set. There is no overfitting/underfitting for k = 9 as test error and training error are the same. Conversely, for k = 13, k=17 and k=21, the training error is more than the test error and the training error is larger compared to the other k values. So there is underfitting. This all goes along with what we have learnt in class. Too small values for k can result in overfitting and too large values for k can result in underfitting (as it approaches the prior probabilities). Finally, k = 17 gives the smallest error for the test set. So, it is my selected k value.

3-) C values greater than 10 results in overfitting, because the error for the training is small (and getting smaller as C is increasing) compared to the test set. C=1 is slightly overfitting since the training set gives slightly less error than test set. C=10^-0.5 results in slight overfitting as the training error is slightly less than the test error. C=1 and C=10 results in underfitting as the training error is at its max and is greater than the test set. The training set and the tet set gives the same error for C=10^0.5. So we can’t say overfitting or underfitting. C = 10 gives the smallest test set error and is thus the best C value.

4-) C=10 gives an average error of 0.12 and K=17 gives an average error of 0.13. So I would pick logistic regression over Kn nearest neighbor.

5-)All the lambda values cause overfitting, because all the training data have smaller error compared to test data for all lambda values. Also, the standard deviation is smaller (almost 0) for the training data too as can be seen from the graph. This is further proof of the overfitting. I would choose ridgereg over linreg, because it has a smaller average error value of 0.11104423519368649 compared to 0.20590100255729174 of linreg. Also, ridgereg has a smaller standard deviation than linreg, 0.022490244530031994 compared to 0.14634624985398925.

6-) All lambda values cause overfitting as training data has smaller error and standard deviation than test data for all lambda values. When it comes to the comparing the logreg with ridgereg2, ridgereg2 has selected parameter 0 and as it is stated in the console output, it is equivalent to linear regression. So, for this case of lambda = 0 they are equivalent. The average error is 0.12169742333678366 and the standard deviation is 0.008957960228842706 for both. However, as the lambda values increase, the errors for both training and test increase and so do the standard deviations for the ridgereg. So I would choose linreg over ridgereg2 for this data variation overall, but for the selected lambda (0) for ridgereg2, they are equivalent.

7-) All lambda values cause overfitting as training data has smaller error and standard deviation than test data for all lambda values. When it comes to comparing the linreg and ridgereg_airfoil, they have the same average error of 0.27771722491099593 and standard deviation of 0.014553491594176923. We know that ridgereg with lambda 0 is equal to linear regression. We can also see from the graph that the errors and the standard deviations are the same for all lambdas. We also know that the parameter doesn’t matter for the linear regression. So we can say that linear regression and logistic regression are equivalent for this dataset.