Computational learning Final

Authors: Ahmad Barq, Anas Abu Gharbieh

Data set:

We have a dataset for predicting weather types sunny cloudy rainy snowy, we basically have 4 classes to classify

Problem:

We wanted to a train a model that can classify the weather type given 10 features that we have.

We also had to change the categorial columns into numbers so for example cloudy was given the number 0 rainy was given the number 2, sunny was given the number 2, snowy was given the number 3

First model

For the first approach we used decision trees with cross validation, we used the predefined decision tree model by sklearn library and then made it run 5 times on the training data set and afterwards calculated the mean cross validation score which turned out to be 87.00%

Second try

For the second try we used bagging for cross validation, basically by defining a loop that trains model of a decision try 5 times resulting in 5 models with different but close model accuracy, the mean accuracy across the 5 models was 90.61% which is better than what we have got with the first model that we have built

Third try

We implemented an adaboost algorithm in hopes of getting better results, since we have 4 classes we ended up assigning the 0, 2 and 3 classes to be the value -1, and then the 1 value stayed as 1, in the end we did get better results as the results we got 91.5%

The results were indeed better but were not so significant since the data set we are using is not as complicated as other data sets.

Conclusion:

We concluded that adaboost did indeed give us the best results out of the three tries that we have made but at the same time we realized that our data set is indeed not as complicated as others meaning that the more depth we gave into the adaboost function the more of an overfitting problem it became clear to us that we shouldn’t use more than a max\_depth of 2 to get the best results out of the adaboost, there were some tries were adaboost even gave worse results than the 2 models prior to it so we decided at 59 estimators and a max depth of 2 at last.