Landon Curtis

Algorithms of Machine Learning

Model Accuracy

Looking at my model, my original data was in consecutive order meaning when it was sampled to train the model it was taken out of order. I would think this actually does no harm to the model, but it is possible it could. I remember hearing everyone in class saying they added a feature such as a month of year or season of the year to help their model but I thought I would start with another route.

First I want to get some better statistics about my model besides an accuracy score so I will get the ROCAUC as we discussed in class. Just calculated it and it was 0.53 so that is not good at all. I will first start with part A. I have 7 features and will remove each one and see how it affects the accuracy score and the ROCAUC .

| | default | tempMax | tempMin | temp | dew | visibility | uvindex | Sealevel pressure |
|---|---|---|---|---|---|---|---|---|
| Accuracy | 0.83 | 0.83 | 0.84 | 0.84 | 0.83 | 0.82 | 0.83 | 0.83 |
| ROCAUC | 0.53 | 0.53 | 0.54 | 0.54 | 0.5 | 0.49 | 0.53 | 0.53 |

Given that removing each variable barely changed to overall accuracy, that means that attribute has little to no weight. It is actually interesting to see the ROCAUC change and have the accuracy stay the same when changing these variables. This makes me curious about how the model is doing with choosing what attributes it uses to weigh most I find the weights later.

I then added data from Portland, OR. I got it from the same data source as my original data so it was in the same order and had the same attributes. This allowed me to reuse my code that classified the if it rained or not, and add it to the data. Here I trained the model and it

added about a one percent increase in accuracy. I then used .coef_ to find the weights of the attributes. It came out as: 0.03459518,  0.11160856, -0.1413044 ,  0.13837626,  0.09418061, -0.00452307, -0.0034593 ,  1.37209367]]). I am unsure of if this is in the same order as the column of the data it was trained on, but I would assume so. With this assumption, it looks like it now depends on the portland data more than anything else (the last one), and everything else is adding no to little help.

Overall, I feel like I learned a lot from this project. I tried to use all the different ways we learned in class to help my model but could not seem to get it to work much. I attempted to use all of the different methods of training our models, such as k-fold, but all of these strategies seemed only to decrease my accuracy. Using all other suggestions and methods, I got a 1.68% increase which is not great. I feel I would be able to make a better model if I was able to start from scratch.