Benjamin Katz

Week 6

Feb 27, 2023

Part 2

<https://colab.research.google.com/drive/1bFBXY5bRTzXa5VlVVvoQbrVmr0NK3vkp?usp=sharing>

<https://colab.research.google.com/drive/1G688gcz2_dI2qAL6Xh4LPK13ucZBa5S8?usp=sharing>

Evaluating a regression Model: Root mean square error

In order to evaluate the effectiveness of a regression model, you want to check how close your predictions are to the actual true data. One effective way of doing so is by finding the root of the mean squared error. The mean squared error is the average squared difference between the predicted and the actual. Squaring this number is important because it ensures that it is positive. If one would not ensure that the number was positive, you would incorrectly assume that you have a perfect model if you predict 1000 above the actual for one sample, and 1000 below the actual in another sample. Ensuring the error is positive allows for a true evaluation of the model. Additionally, the root mean squared error gives more weight to larger errors. If the data set contains outliers, the RMSE will be so large that it may not be a good indicator of how the model is actually performing. Because of this, outliers should be removed from the dataset in the preprocessing stage or a different evaluation function should be used. Because we are trying to predict actual housing prices from the Boston data set, we build a regression model. RMSE is an effective evaluation metric for evaluating the goodness of our model.

Evaluating a regression Model: Mean absolute error

Mean absolute error is another way of evaluating the goodness of a regression model. Unlike RMSE, mean absolute error takes the absolute value of the difference between the predicted and actual. In this way, MAE also does not have the same issue of thinking that the model is good just because some predictions were way too high and some were way too low. The major advantage of this approach is that outliers do not disproportionately affect the evaluation like they do in RMSE. A major disadvantage of this metric is that it is not differentiable(Because of the absolute value). It is therefore impossible to take the derivative in an effort to minimize error.

Evaluating a classifier Model: Precision and Recall:

Precision: When looking at a classification model, especially a binary classifier, different metrics can tell you different things about your results. The different evaluations are application specific and in some cases you want to optimize for one over the other. Precision is the amount of true positives over all the predicted positives. Aka what percentage of the predicted positives did the model “get right”. Optimizing for this parameter incentivizes the system to only predict positive when it is very confident. When in doubt, make it negative. Precision might be an important metric for a small company with many applicants looking to hire new employees. They may lose out by not hiring someone who is very talented, but it is much more important to them that they don’t hire people that can not do the job.

Recall: recall is the correctly predicted positives over all the real positives. This gives a percentage of how many of the total actual positives, did the model predict. Optimizing for this metric incentivizes the model to classify almost everything as positive. Recall might be an important metric for a college struggling to hit admissions numbers. They will accept everyone that might have a shot at succeeding in the college. Sure, some will fail, but the admissions office can not afford to risk having people that might have been qualified rejected.

F-score: The F-score is a harmonic mean of precision and recall. Since it is likely that we want our model to have a high Precision as well as a high recall, optimizing the F-score optimizes the balance between the two. The basic F one score is two times precision times recall over precision plus recall. This gives a score between zero and one where zero is the worst possible precision and recall where one is perfect precision and perfect recall.