Assignment Week 3

Due: 2/5/2024

Shimon Greengart

For regression machine learning, two important metrics are Mean Square Error and R-Squared. Mean Square Error (MSE) measures the distance between the predicted target values and the actual values, measuring the variance from statistics. A higher MSE means a much more random prediction, while an MSE of 0 means a perfect prediction. However, MSE does not use the same units as the target values, making it much harder to compare different models. To fix that, one can take the square root of the MSE, getting the Root Mean Square Error, measuring the standard deviation from statistics. It is much more comparable between models because it is much clearer what it is measured in.

R-Squared measures how much of the variance in the target value is explained by the model. An R-Squared value of 1 means a perfect model that fully explains the target value. A lower R-Squared value means that the model explains less about how the target value varies, with a negative R-Squared value being terrible. One flaw with R-Squared is that it always increases when more independent variables are added, even if they don't predict anything. Adjusted R-Squared has been modified to remove this bias.

In classification machine learning, accuracy is generally not the best metric. The reason for this is that in a case of unbalanced classes, the model could perfectly detect one class but not detect the other at all yet still have fairly good accuracy. For example, xkcd has a comic called Is it Christmas?, where it determines if a day is Christmas or not by always saying "No", even when it is Christmas. But the comic reports an accuracy of 99.73%, even though on Christmas, it is always wrong, making it useless as a predictor. As such, given unbalanced classes (and in most cases), accuracy is not the best metric.

Better metrics are recall and precision. Like accuracy, they are probabilities. Recall measures, given a group you are classifying, the probability that your model correctly identifies it. For example, the xkcd model has a recall for Christmas of 0, since it consistently fails to identify Christmas. It is most useful in a case of unbalanced classes where you want to make sure that you catch every positive case. For example, when screening for cancer, you want to make sure that you catch every patient with cancer, even if it means giving a few false positives.

Precision measures the probability that, when your model classifies an instance as being of a certain group, it made that classification correctly. A higher precision means fewer false positives, cases where you identified an object as class X but it was actually class Y. It is most useful in a case where you have big problems if you incorrectly identify the class. For example, if you are the IRS, one accusation on someone politically connected of tax evasion that turns out to have been false could cause you to lose your entire budget, so you would want a high Precision.

In most cases, you want both high recall and precision, perhaps with a greater emphasis on recall. In such a case, you could use F_1 Score, which is the harmonic mean of precision or recall. This means that if either precision or recall is low, F_1 Score is also low, and it is only high if both of the other scores are high.