Dakota Crowder

dcrowder2@alaska.edu

CSCE A415 Machine Learning

Assignment 5

Write-Up

My findings were pretty interesting, there was a trend upwards for the accuracy vs number of instances sent into the naïve Bayesian classifier, which makes sense as more data would help increase accuracy of probabilities, but with the True Positive rate vs False Positive rate, the ROC curve, it was very strange, the TPR reached respectable numbers of upwards of .9, however instead of a increasing FPR coinciding with the TPR, the FPR kind of went all over the place making the ROC curve look very different from any other ROC curve I have ever seen. I assume this is an okay result, since the FPR never got above .02, meaning there was a very, very small number of false positives, which makes it very accurate. However while I assume this just means this data has some good conditional probability and allowed the Naïve Bayesian Classifier to do a great job, I do worry my calculations for the FPR and TPR are incorrect in some way, I did have some order of operations issues with the conditional probabilities that caused me some terrible bugs that took a while to fix. I hope the same simple issues that cause huge differences isn't happening with the FPR and TPR calculations, but other than that I think this is a success.

Here are the final graphs for one run, all the runs I did looked like these, so I thought I would just show on pair to illustrate my results: