Ian Cross

Machine Learning A1

2.1

(Basic line graph from excel)

We can conclude a couple of things about the algorithm sensitivity

1. For the J48 classifier, the number of the features doesn’t really affect the overall accuracy of the algorithm. This means that those extra features are irrelevant (and actually make it slightly worse).
2. However, for the IBk classifier, the extra features actually make the accuracy much worse, and using the initial 14 features has the highest percent accuracy.

Problem 2.2 – Learning Curves

|  |  |  |
| --- | --- | --- |
|  | Mean | Standard Deviation |
| J48 on 14Features | 70.67 | 5.11 |
| J48 on 54Features | 61.28 | 7.97 |
| Ibk on 14Features | 75.5 | 8.31 |
| Lbk on 54Features | 52.33 | 1.639 |

I would say that from these curves, we can say that generally the larger the training set size, the higher the accuracy of the algorithm (although there were cases that the lines remained somewhat constant).