

Assignment 5

William Richard

November 1, 2012

1 Accuracies and Standard Deviations

| algorithm | A | B | C | ionosphere |
|---------------------------|---------------------|---------------------|---------------------|---------------------|
| generativeSharedCov | 0.952 +/- 0.0175 | 0.850 +/- 0.0548 | 0.830 +/- 0.0458 | 0.855 +/- 0.0592 |
| generativeSeperateCov | 0.515 +/- 0.0376 | 0.550 +/- 0.1225 | 0.515 +/- 0.1566 | 0.641 +/- 0.0697 |
| fishersLinearDiscriminant | 0.054 +/- 0.0138 | 0.150 +/- 0.0806 | 0.160 +/- 0.0663 | 0.128 +/- 0.0514 |
| logisticRegression | 0.485 +/- 0.038 | 0.645 +/- 0.111 | 0.700 +/- 0.114 | 0.661 +/- 0.076 |

Table 1: Mean accuracy +/- 1 standard deviation for all algorithms on all datasets

Surprisingly, the generative model with shared covariance matrices did best in all cases, not only with the highest average accuracy but also the lowest standard deviations.

2 Learning Curves

With all algorithm on all datasets, the training size does not seem to have a profound effect. It does help to have larger training sets, but in most cases the training set size did not drastically effect the preformance of the algorithm.

Interestingly, on dataset A, having a training set of size 1000 with the generative regression with seperate covariance matrices saw a great improvement in performance. This is likely random chance, rather than meaningful improvement. See fig 1.

It is possible that I did not impliment logistic regression correctly - for some datasets, like B and C (see figures 2 and 3) it looks like some variation occured for different sized training sets, while for the A and ionosphere dataset (figures 1 and 4) no difference appeared to occur when training set changed. Also, the logistic training took a very long time - on the order of 8-10 minutes per fold on dataset A.

It appears that though the generative model with shared covariances is one of the simlier algorithms that we explored, in almost all scenarios it performed best, with the highest mean accuracy and smallest standard deviations from that mean.

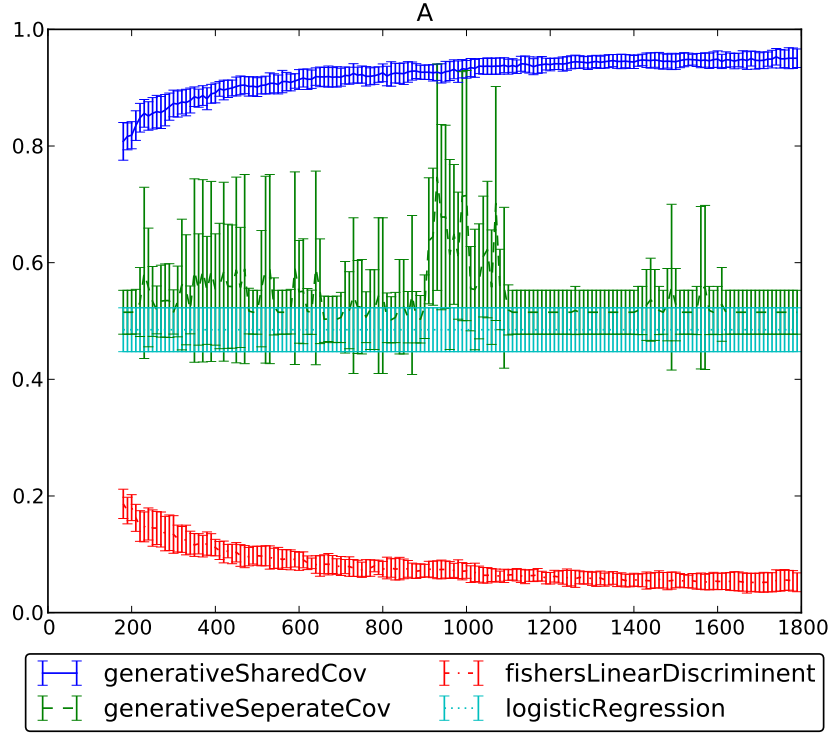


Figure 1: Learning Curves on the A dataset

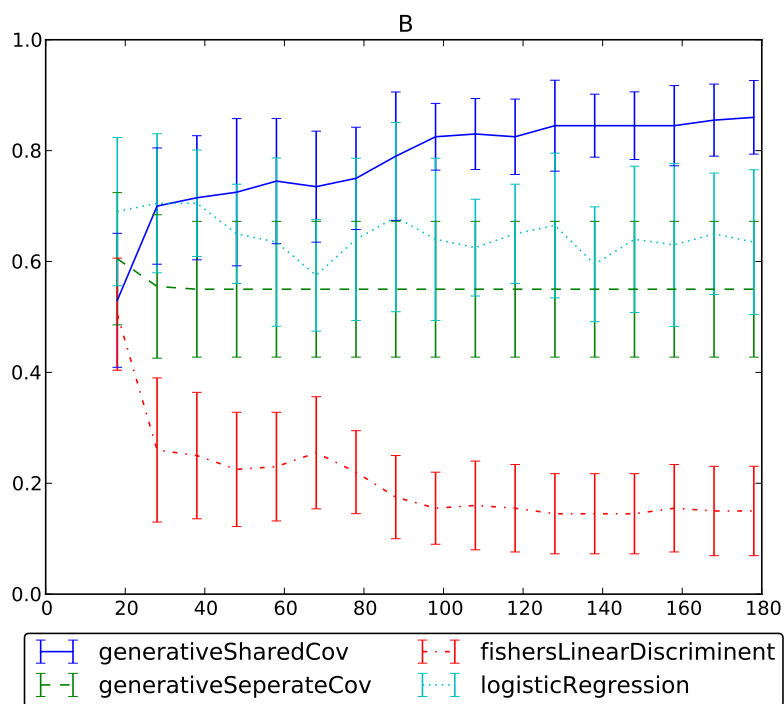


Figure 2: Learning Curves on the B dataset

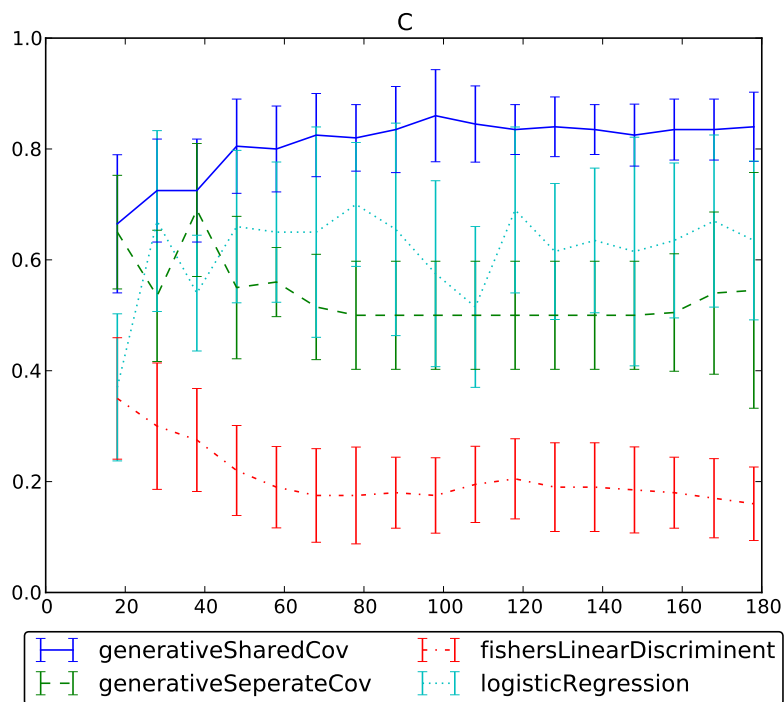


Figure 3: Learning Curves on the C dataset

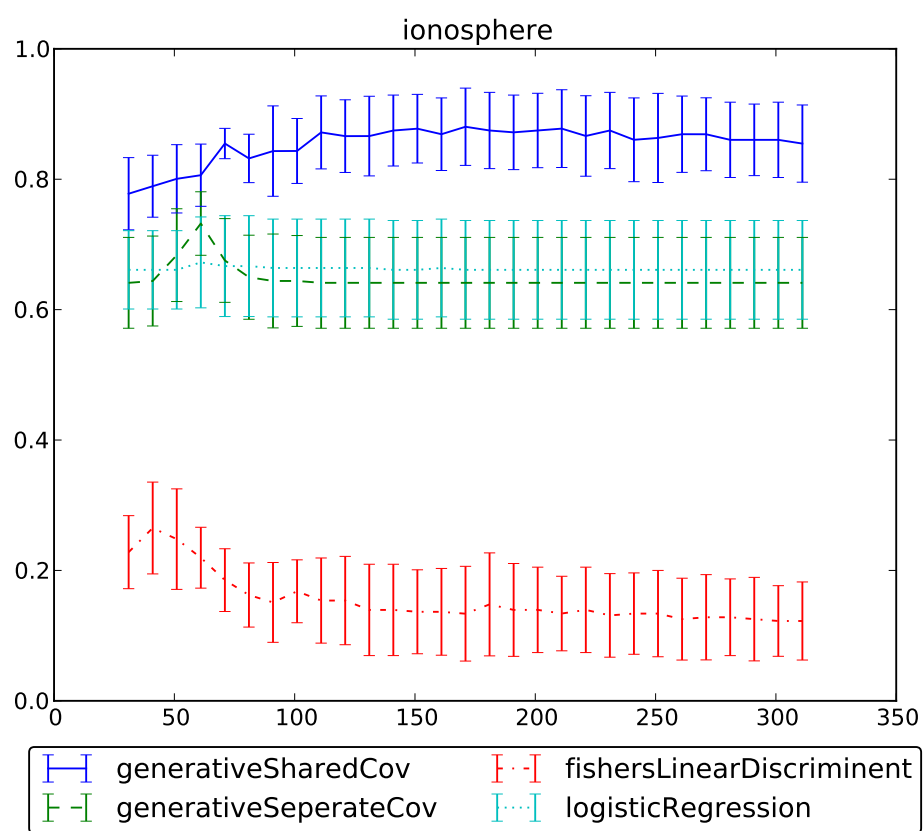


Figure 4: Learning Curves on the ionosphere dataset