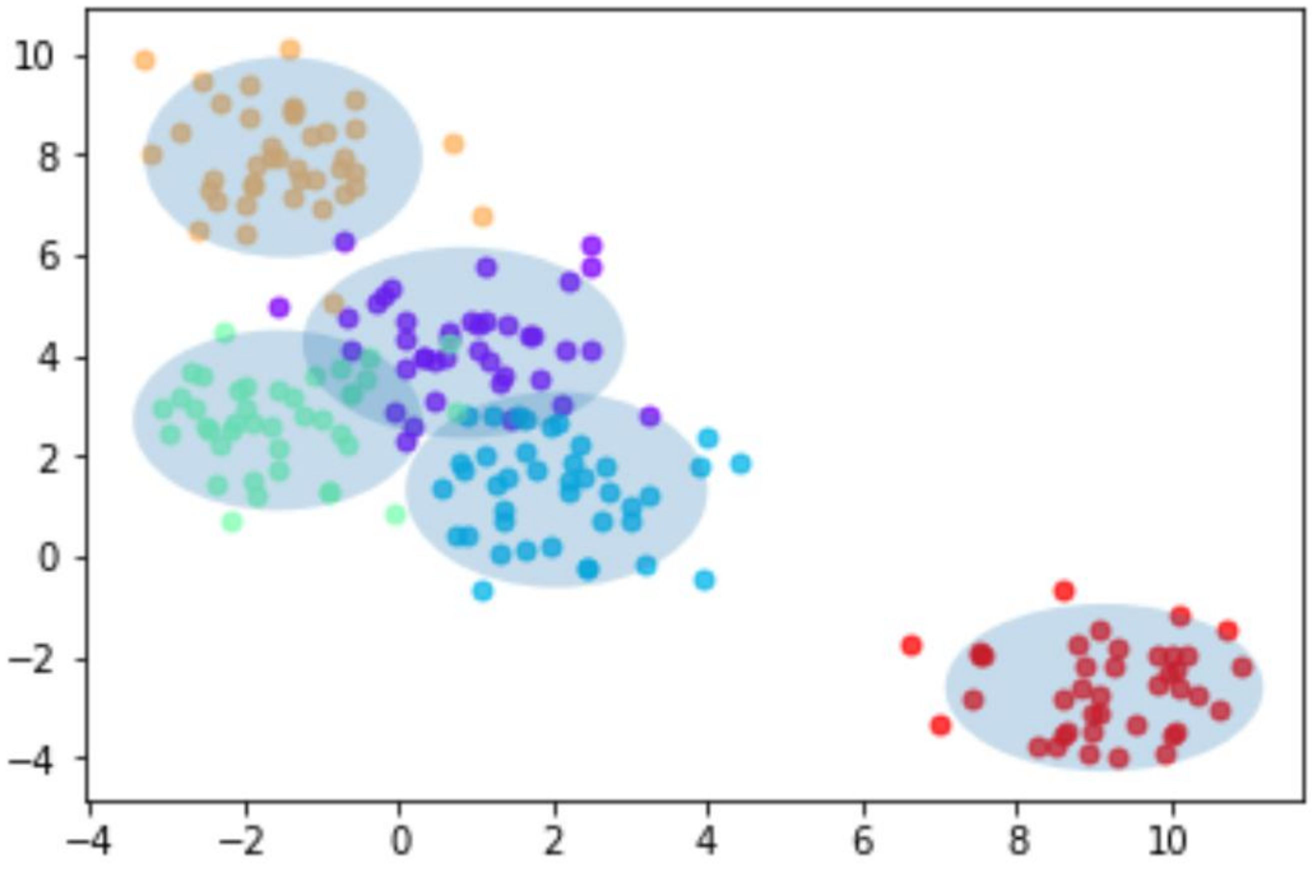
Machine Learning Lab 3

Catherine Weldon + Qiao Ren

Assign 1

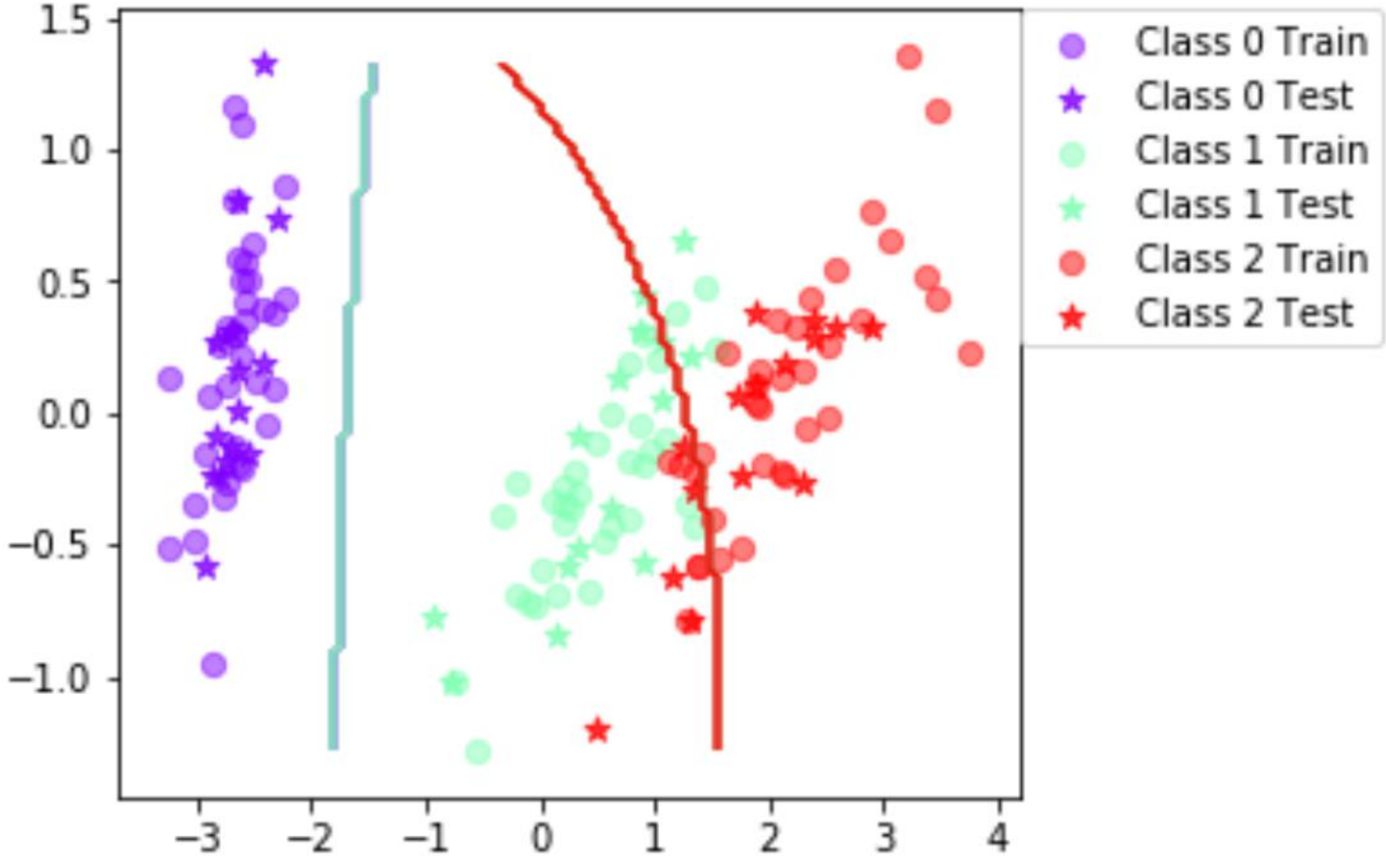


Assign 3

In [353]:
```
testClassifier(BayesClassifier(), dataset='iris', split=0.7)
```
```
Trial: 0 Accuracy 84.4
Trial: 10 Accuracy 95.6
Trial: 20 Accuracy 93.3
Trial: 30 Accuracy 86.7
Trial: 40 Accuracy 88.9
Trial: 50 Accuracy 91.1
Trial: 60 Accuracy 86.7
Trial: 70 Accuracy 91.1
Trial: 80 Accuracy 86.7
Trial: 90 Accuracy 91.1
Final mean classification accuracy  89 with standard deviation 4.16
```

In [354]:
```
testClassifier(BayesClassifier(), dataset='vowel', split=0.7)
```
```
Trial: 0 Accuracy 61
Trial: 10 Accuracy 66.2
Trial: 20 Accuracy 74
Trial: 30 Accuracy 66.9
Trial: 40 Accuracy 59.7
Trial: 50 Accuracy 64.3
Trial: 60 Accuracy 66.9
Trial: 70 Accuracy 63.6
Trial: 80 Accuracy 62.3
Trial: 90 Accuracy 70.8
Final mean classification accuracy  64.7 with standard deviation 4.03
```

Assign 3+4



- **When can a feature independence assumption be reasonable and when not?**
  - It is reasonable to assume independence if there is not a correlation between variables. This can be found computing the covariance matrix and looking for a diagonal
- **How does the decision boundary look for the Iris dataset? How could one improve the classification results for this scenario by changing classifier or, alternatively, manipulating the data?**
  - Boundary isn't great, one could improve the classifier using boosting or bagging
  - Can manipulate by removing or combining dependent features

Assign 5

Iris:

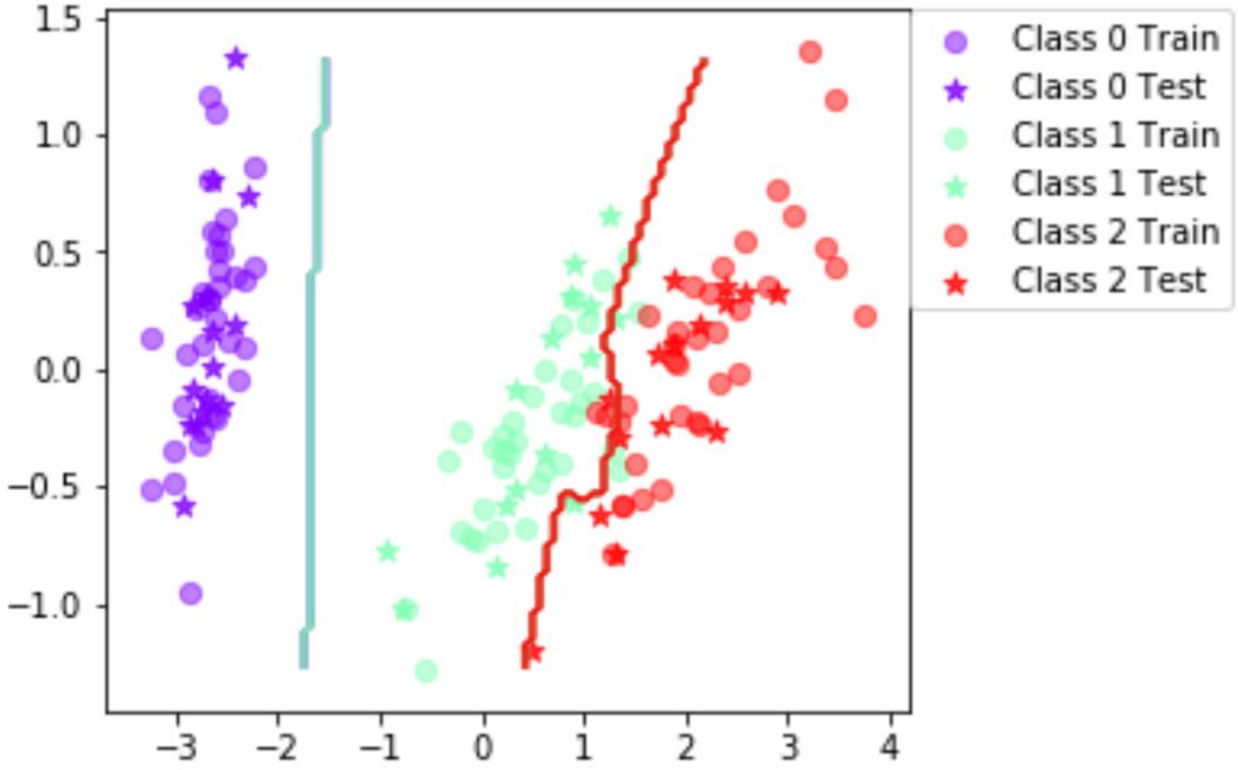
```
Trial: 0 Accuracy 95.6
Trial: 10 Accuracy 100
Trial: 20 Accuracy 93.3
Trial: 30 Accuracy 91.1
Trial: 40 Accuracy 97.8
Trial: 50 Accuracy 93.3
Trial: 60 Accuracy 93.3
Trial: 70 Accuracy 97.8
Trial: 80 Accuracy 95.6
Trial: 90 Accuracy 93.3
Final mean classification accuracy  94.8 with standard deviation 2.82
```

Vowel:

```
Trial: 0 Accuracy 76.6
Trial: 10 Accuracy 86.4
Trial: 20 Accuracy 83.1
Trial: 30 Accuracy 80.5
Trial: 40 Accuracy 72.7
Trial: 50 Accuracy 76
Trial: 60 Accuracy 81.8
Trial: 70 Accuracy 82.5
Trial: 80 Accuracy 79.9
Trial: 90 Accuracy 83.1
Final mean classification accuracy  80.1 with standard deviation 3.5
```

- Is there any improvement in classification accuracy? Why/why not?
  - Yes, the Adaboost algorithm helped improve accuracy by re-weighting the model to focus on the incorrectly classified points

Assign 5 Cont…



- Plot the decision boundary of the boosted classifier on iris and compare it with that of the basic. What differences do you notice? Is the boundary of the boosted version more complex?
  - The boundary is better: less misclassified points
  - It also appears to be more complex, particularly in the middle
- Can we make up for not using a more advanced model in the basic classifier (e.g. independent features) by using boosting?
  - **{not sure}**

- Testing different parameters:
  - With the Iris, higher split resulted in better accuracy
  - 4 trials seemed good enough to improve results, not worth going higher than 10

Assign 6

- Iris: no boost

```
Trial: 0 Accuracy 95.6
Trial: 10 Accuracy 100
Trial: 20 Accuracy 91.1
Trial: 30 Accuracy 91.1
Trial: 40 Accuracy 93.3
Trial: 50 Accuracy 91.1
Trial: 60 Accuracy 88.9
Trial: 70 Accuracy 88.9
Trial: 80 Accuracy 93.3
Trial: 90 Accuracy 88.9
Final mean classification accuracy  92.4 with standard deviation 3.71
```

- Iris: boost

```
Trial: 0 Accuracy 95.6
Trial: 10 Accuracy 100
Trial: 20 Accuracy 95.6
Trial: 30 Accuracy 93.3
Trial: 40 Accuracy 93.3
Trial: 50 Accuracy 95.6
Trial: 60 Accuracy 88.9
Trial: 70 Accuracy 93.3
Trial: 80 Accuracy 93.3
Trial: 90 Accuracy 93.3
Final mean classification accuracy  94.6 with standard deviation 3.67
```

- Vowel: no boost

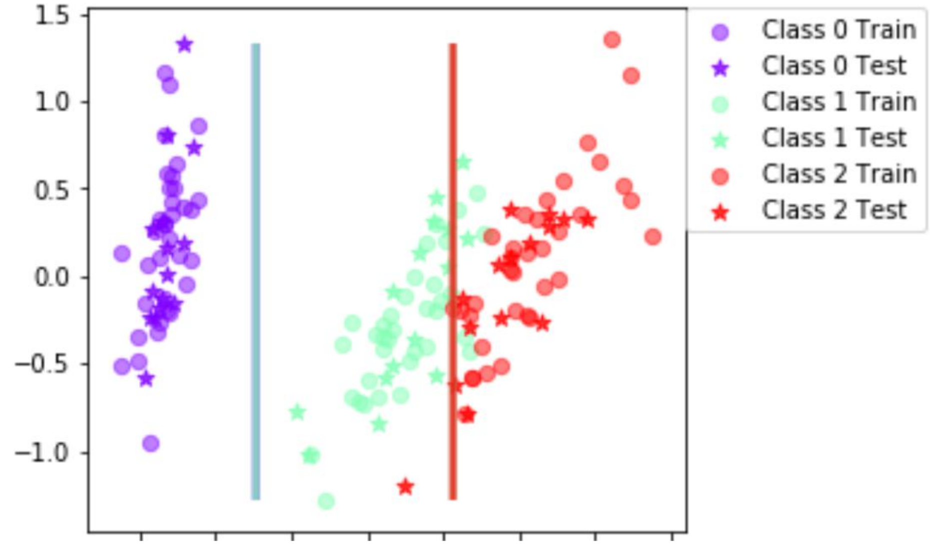
```
Trial: 0 Accuracy 63.6
Trial: 10 Accuracy 68.8
Trial: 20 Accuracy 63.6
Trial: 30 Accuracy 66.9
Trial: 40 Accuracy 59.7
Trial: 50 Accuracy 63
Trial: 60 Accuracy 59.7
Trial: 70 Accuracy 68.8
Trial: 80 Accuracy 59.7
Trial: 90 Accuracy 68.2
Final mean classification accuracy  64.1 with standard deviation 4
```

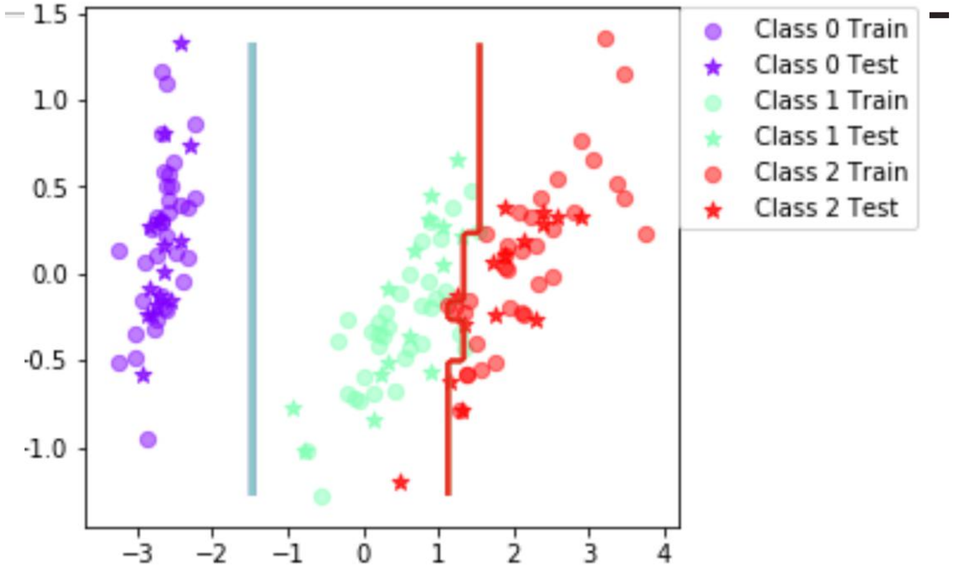
- Vowel: boost

```
Trial: 0 Accuracy 86.4
Trial: 10 Accuracy 89.6
Trial: 20 Accuracy 87
Trial: 30 Accuracy 90.9
Trial: 40 Accuracy 84.4
Trial: 50 Accuracy 79.9
Trial: 60 Accuracy 87
Trial: 70 Accuracy 88.3
Trial: 80 Accuracy 83.1
Trial: 90 Accuracy 87
Final mean classification accuracy  86.9 with standard deviation 2.95
```

Assign 6 cont…

- 1. no boost, 2. boost





- Is there any improvement in classification accuracy? Why/why not?
  - Yes, the Adaboost algorithm helped improve accuracy by re-weighting the model to focus on the incorrectly classified points
- Plot the decision boundary of the boosted classifier on iris and compare it with that of the basic. What differences do you notice? Is the boundary of the boosted version more complex?
  - The boundary is better: less misclassified points
  - Boundary is definitely more complex
- Can we make up for not using a more advanced model in the basic classifier (e.g. independent features) by using boosting?
  - The re-weighting should help with this

Assign 7

- If you had to pick a classifier, naive Bayes or a decision tree or the boosted versions of these, which one would you pick? Motivate from the following criteria:
- Outliers:
    - Would avoid boosting as it would try to fit outliers that we don't necessarily want to model
- Irrelevant inputs: part of the feature space is irrelevant
    - If there is any feature redundancies or irrelevance it might be better to go with a decision tree as linear independence is not required and bayes will be oversensitive to irrelevant data, that said, naïve bayes still works on somewhat dependent features
- Predictive power:
    - Both classifiers work well and boosting improves the models
        - Decision tree is high variance low bias – tends to overfit, needs pruning
        - Bayes is higher bias and low variance
- Mixed types of data: binary, categorical or continuous features, etc.
    - both decision tree and bayes could work on all the above types of features
    - binary: decision tree is better
    - categorical : decision tree is better
    - continuous: bayes is better
- Scalability: the dimension of the data, D, is large or the number of instances, N, is large, or both.