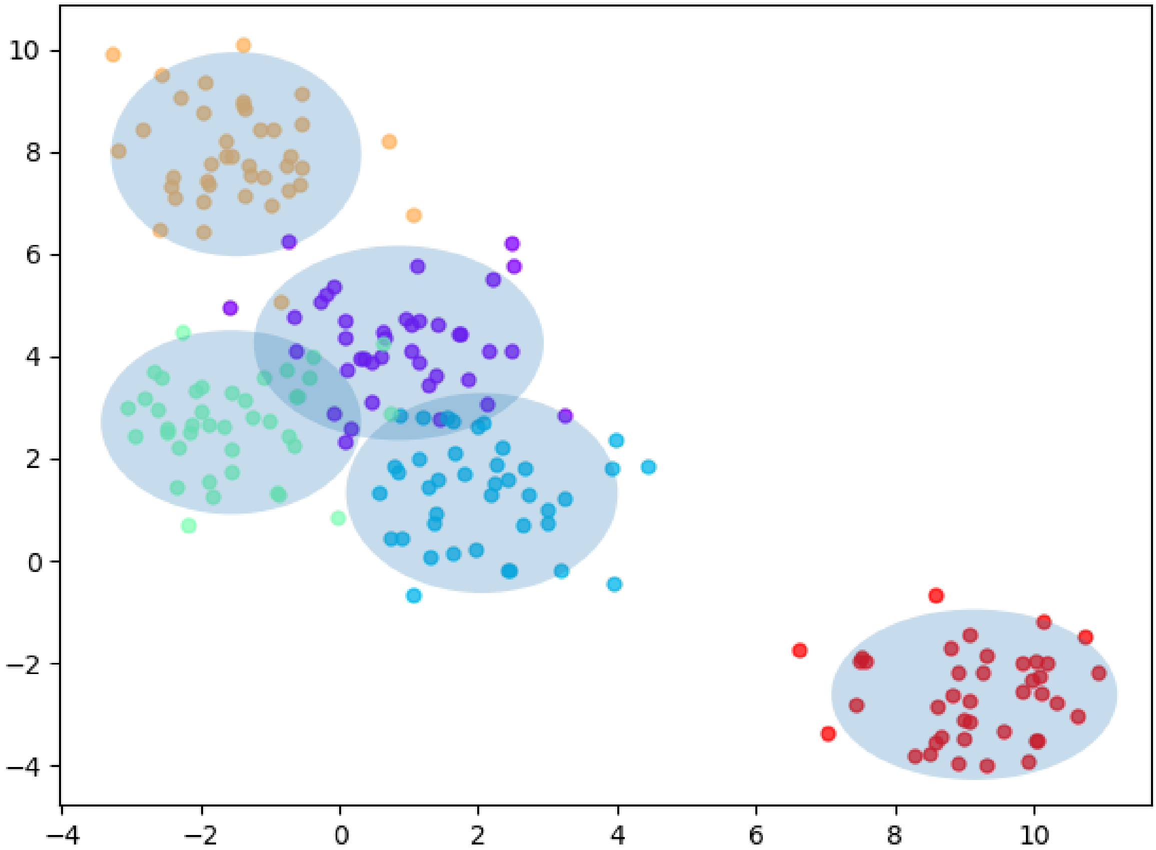
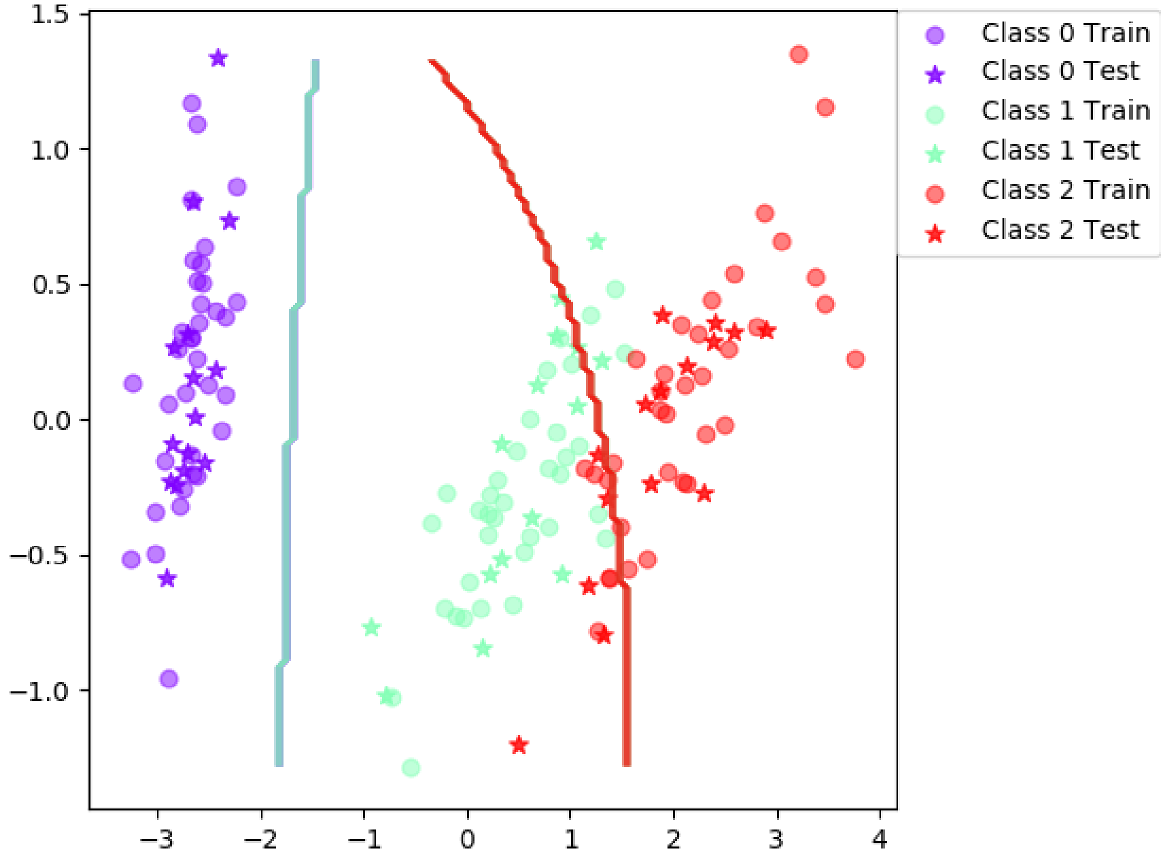
Lab 3

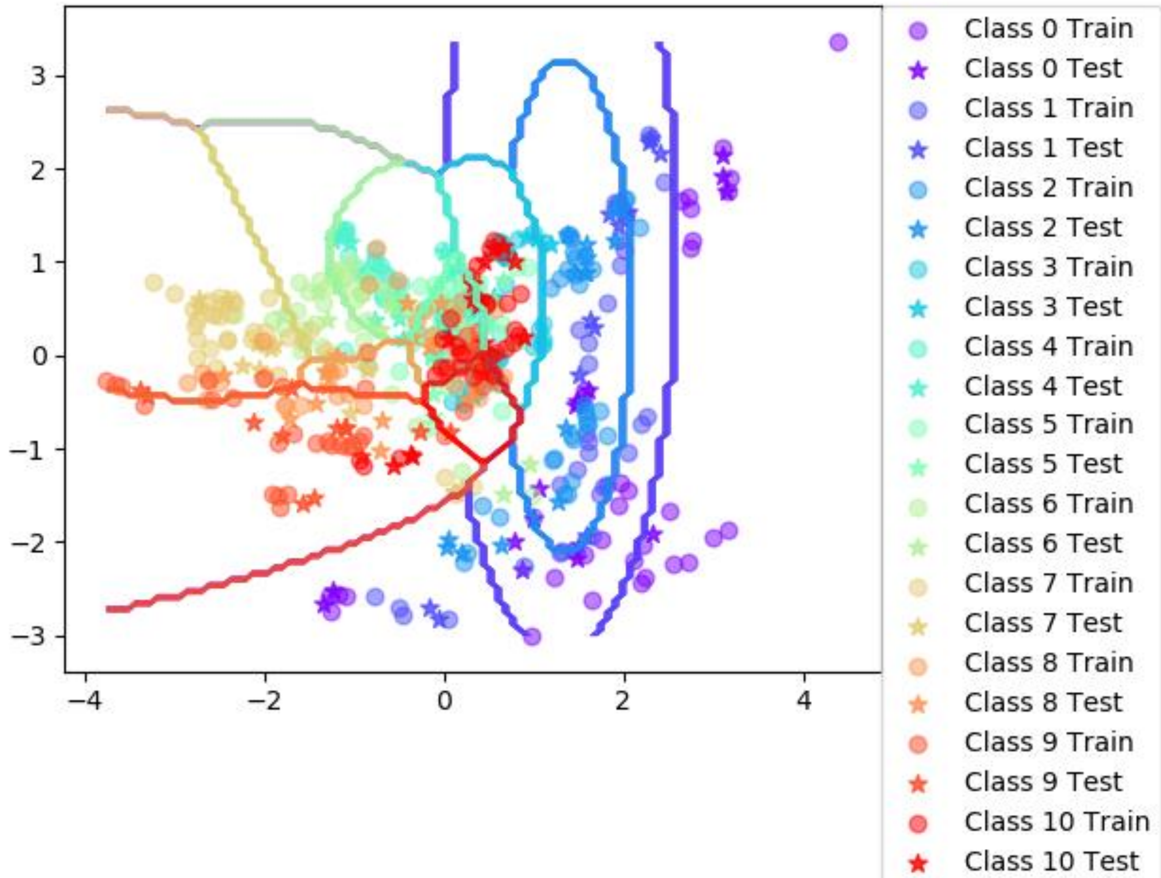


Test dataset X, without boosting (naiva bayes)



Iris dataset without boosting (Naive Bayes)

Final mean classification accuracy 89 with standard deviation 4.16



Vowel dataset without boosting (Naive Bayes)

Final mean classification accuracy 64.7 with standard deviation 4.03

Questions

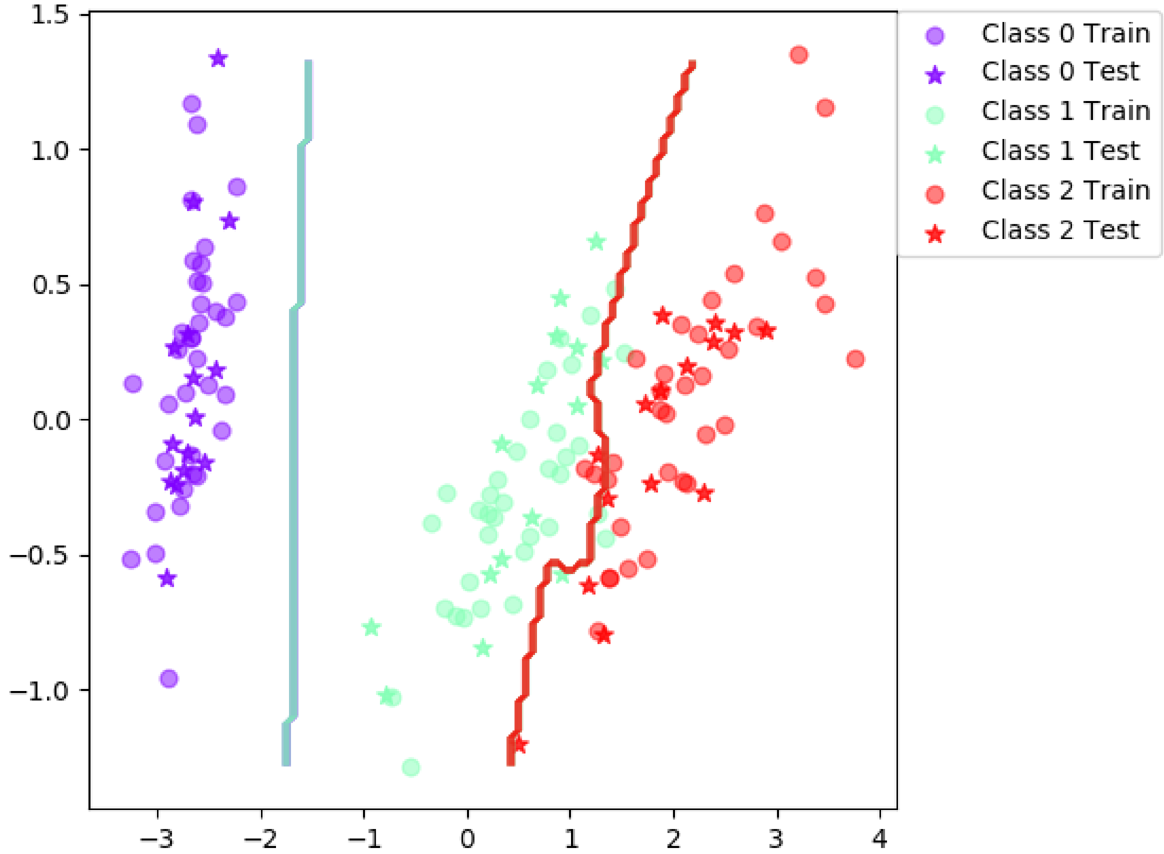
When can a feature independence assumption be reasonable and when not?

Bayes classifiers assume that the value of a feature is independent of the value of any other feature, given the class variable. In reality (example in NLP) features tend to be dependent but we still assume them as independent hence the name "Naive". Surprisingly NB models perform well despite the conditional independence assumption. In the case of the iris dataset, for example, we can suppose that `sepal length` and `sepal width` are positively correlated as well as `petal length` and `petal width`. When using NB you win in time complexity but loses in accuracy.

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?

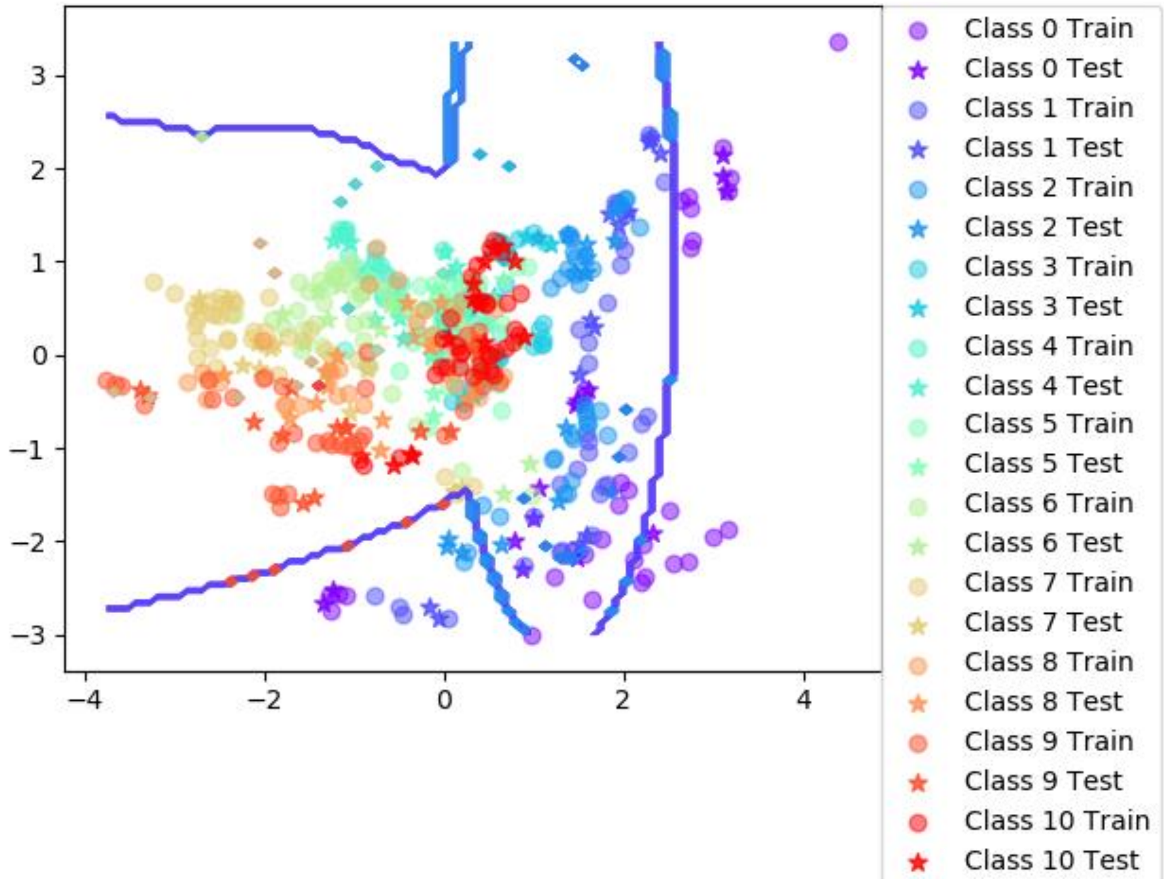
The decision boundary between class 0 and 1 is well defined as expectable for easily separable classes. However, classes 1 and 2 are noisier and often overlap. This leads to an unclear and counterintuitive boundary leaning on the left. This is clearly a result of the weak nature of the classifier. Probably SVM with slack variables or random forests would have worked better as classifiers. Non-linear transformations to the dataset may also be a solution.

BOOSTING



Iris dataset with boosting (naive Bayes)

Final mean classification accuracy 94.1 with standard deviation 6.72



Vowel dataset with boosting (Naive Bayes)

Final mean classification accuracy 80.2 with standard deviation 3.52

Questions

Is there any improvement in classification accuracy? Why/why not?

Yes, there is an improvement of the classification in both the *iris* and *vowel* datasets. This is expectable because the boosting technique allows to concentrate on the misclassified samples and build a more accurate model increasing the variance (NB generally have high bias and low variance).

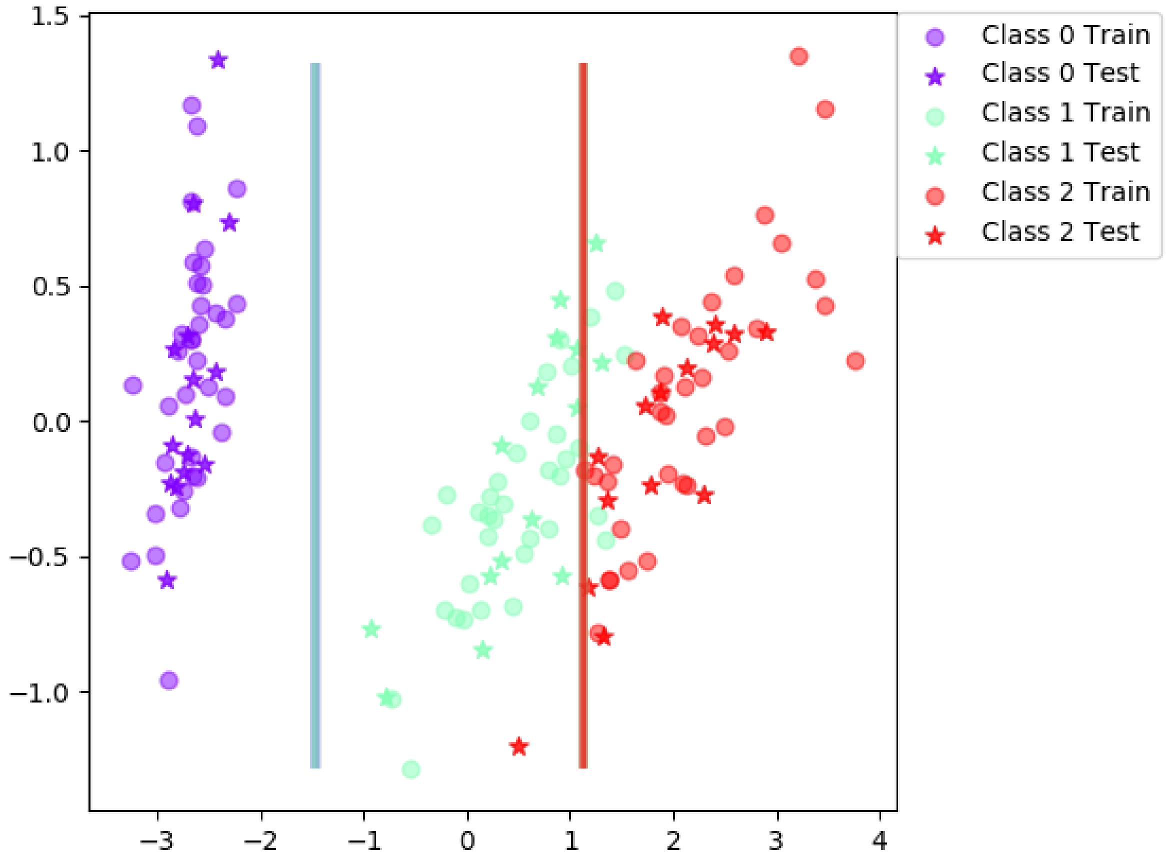
Plot the decision boundary of the boosted classifier on *iris* and compare it with that of the basic. What differences do you notice? Is the boundary of the boosted version more complex?

The boundary is now more complex and fits better the underlying data (see fig above). The tendency to lean on the left has been removed and because boosting allowed the classification algorithm to focus on the misclassified points.

Can we make up for not using a more advanced model in the basic classifier (e.g. independent features) by using boosting?

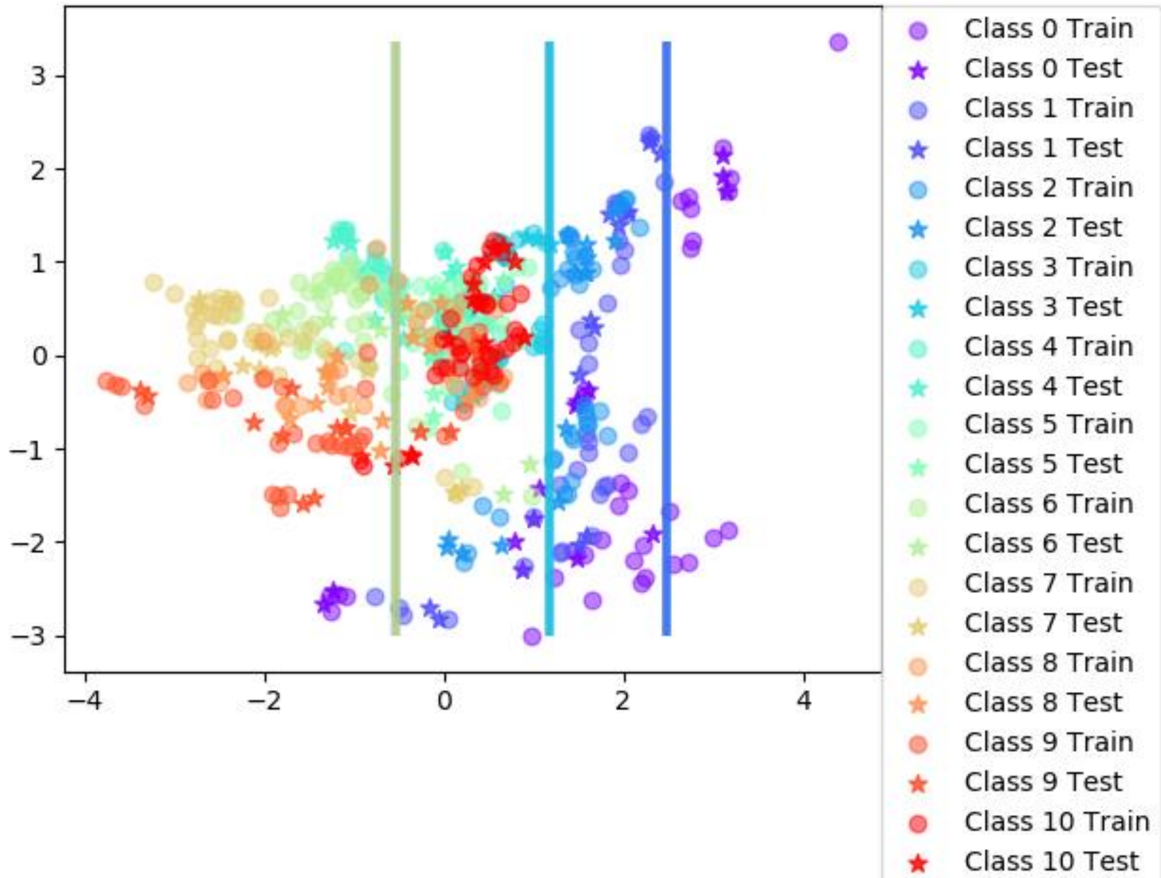
It is indeed possible to increase the accuracy of a weak classifier by means of boosting. However, we should be able to run the weak classifier on partitions of the datasets with different distributions.

Decision Trees



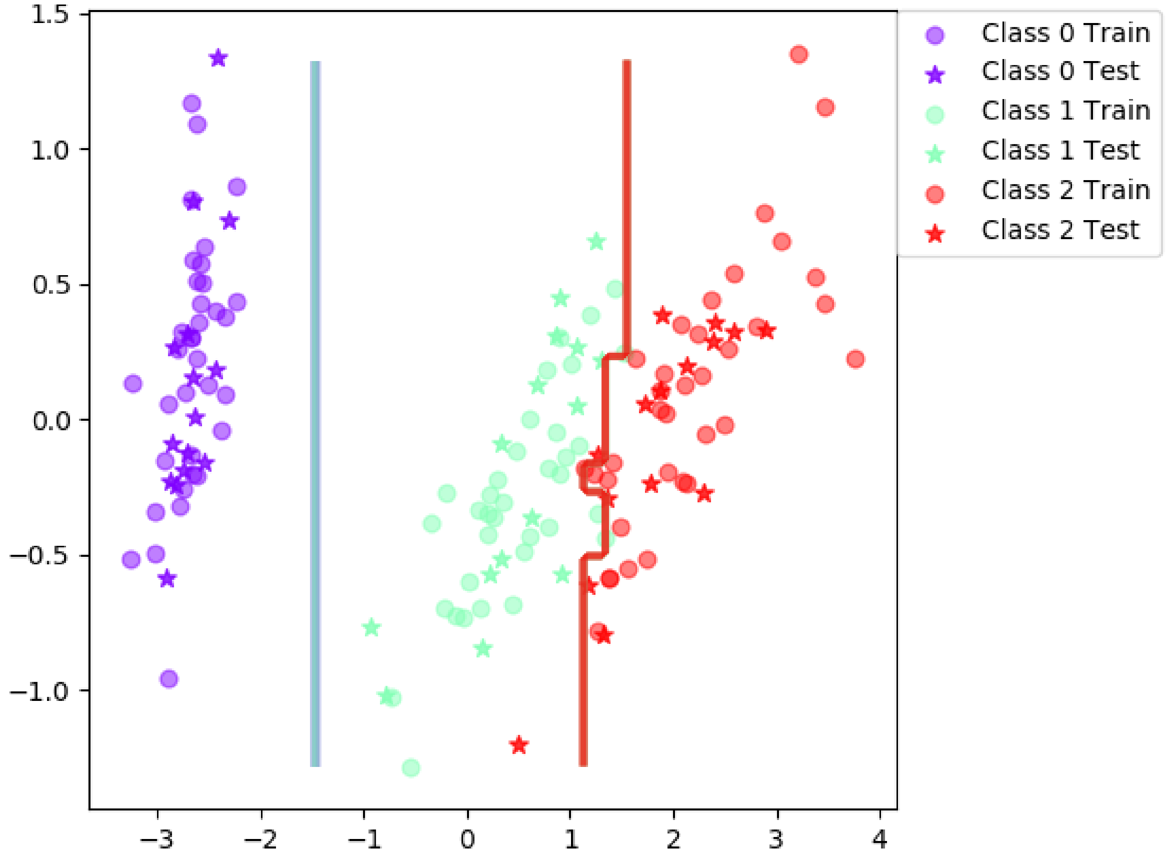
Iris dataset Decision Tree without boosting

Final mean classification accuracy 92.4 with standard deviation 3.71



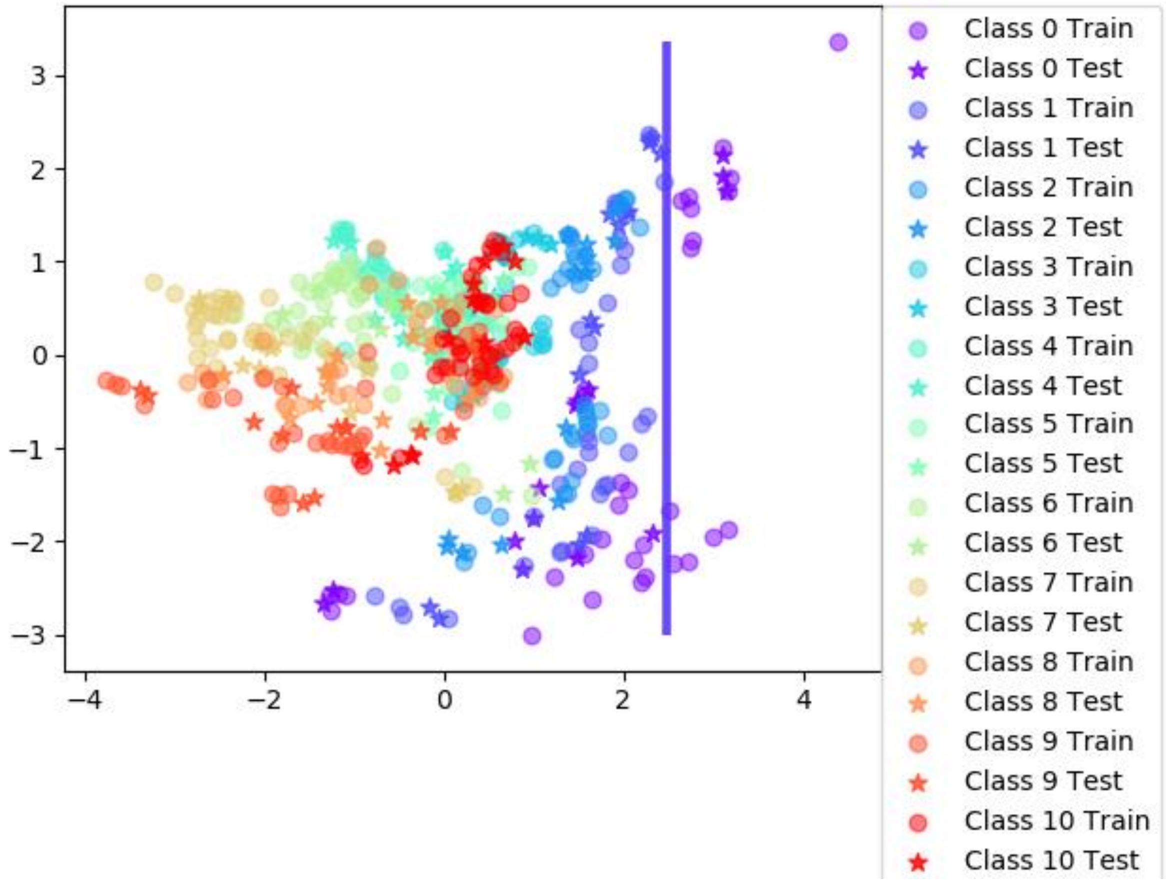
Vowel dataset Decision Tree without boosting

Final mean classification accuracy 64.1 with standard deviation 4



Iris dataset Decision Tree with Boosting

Final mean classification accuracy 94.6 with standard deviation 3.65



Vowel dataset Decision Tree with Boosting

Final mean classification accuracy 86.7 with standard deviation 2.9

Questions

Is there any improvement in classification accuracy? Why/why not?
Yes, the boosted version of the algorithm yields better results than the normal decision trees. A decision tree is a weak classifier and we get the biggest increase in accuracy on the Vowels dataset where data points are more mixed (DT generally have low bias and high variance).

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 indeed more "edgy" in the boosted version.

Can we make up for not using a more advanced model in the basic classifier (e.g. independent features) by using boosting?
Yes, in certain cases (see above).

7.

- **Outliers:** Naïve Bayes without boosting. Decision trees would tend to overfit the data and also a boosted Bayes classifier would give too much weight to the outliers.
- **Irrelevant inputs: part of the feature space is irrelevant:** Decision Trees. They would tend to split ignore the irrelevant part of the feature space concentrating only on attributes with high information gain.
- **Predictive power:** Naïve Bayes with boosting. It tends to yield the best performance on prediction.
- **Mixed types of data: binary, categorical or continuous features, etc.:** Decision Trees are more flexible and work well both with quantitative and qualitative data while Bayes works better with continuous data. Probably using boosting would increase the accurancy.
- **Scalability: the dimension of the data, D, is large or the number of instances, N, is large, or both:** Decision Trees. Bayes works well even with small datasets while decision trees gain in performance when the dataset is large.