**Lab 3: Bayesian leaning and boosting**

1. **BAYESIAN LEARNING**
2. **Assignment 3**

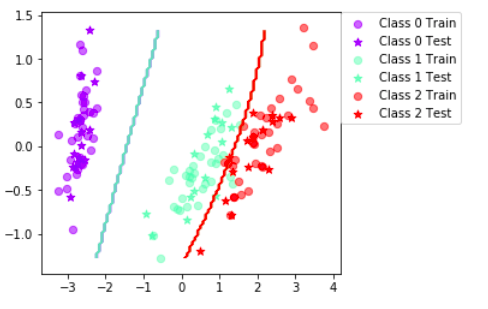
**When can a feature independence assumption be reasonable and when not?**

Feature independence assumptions could be made based on domain knowledge, i.e., nature of the features.

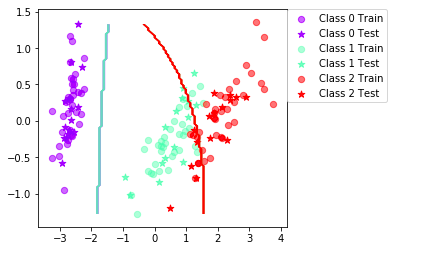
Ex: Height and weight are likely to be dependent

Statistical dependency can be checked by looking at covariance matrix (off-diagonal elements = 0) or scatter plots between features. However, only linear dependency can be checked.

**How does the decision boundary look for the Iris dataset?**

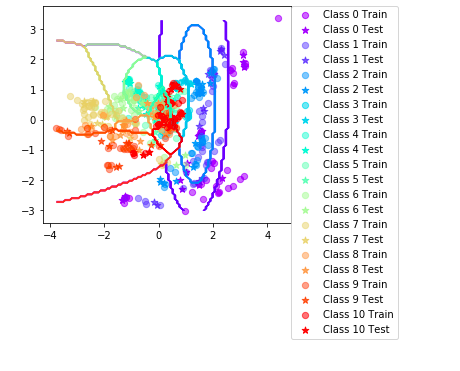


**Figure 1:** Decision boundary using Bayesian classifier **WITHOUT** taking into account independent features assumption (Data set Iris)



**Figure 2:** Decision boundary using Bayesian classifiertaking into account independent features assumption (Data set Iris)

Class 0 is separated from the other class but class 1 and 2 is overlap in this 2D plot.



**Figure 3:** Decision boundary using Bayesian classifier(Data set Vowel)

**How could one improve the classification results for this scenario by changing classifier or, alternatively, manipulating the data?**

As the class 1 and 2 are overlapped, we can try SVM to project the data points into a higher dimensional space.

Manipulating data:

Checking outliers and clean or remove outliers.

Naïve Bayes classifier bases on estimation of posteriori which is estimated from priors, covariance matrix (i.e., independent features) and means. Hence, sampling is important in this method to have accurate estimates of parameters.

Shuffle and resample the training and validation data

1. **BOOSTING**
2. **Assignment 5**

Compute the classification accuracy of the boosted classifier on some data sets using testClassifier from labfuns.py and compare it with those of the basic classifier on the vowels and iris data sets (see Assignment 3):

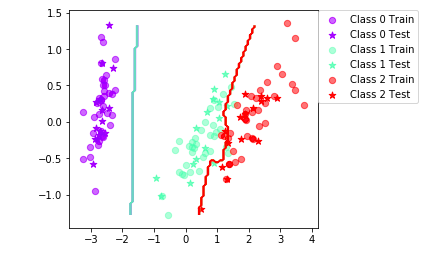
|  |  |  |
| --- | --- | --- |
| **Data** | **Basic classifier** | **Boosted classifier** |
| Iris | 89 (SD 4.16) | 94.1 (SD 6.72) |
| Vowel | 64.7 (SD 4.03) | 80.2 (SD 3.52) |

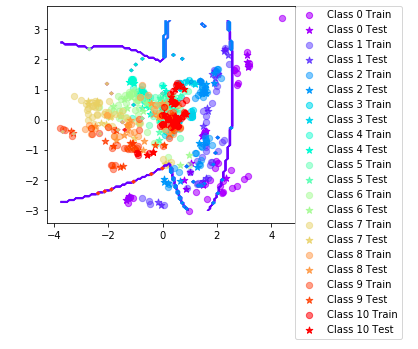
**Is there any improvement in classification accuracy? Why/why not?**

Yes. Because we used boosting to combine many weak classifier to reduce 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 decision boundary is more complex and more accurate

**  
Figure 3:** Decision boundary using Boosting Bayesian classifier(Data set Iris)

****

**Figure 4:** Decision boundary using Boosting Bayesian classifier(Data set Vowel)

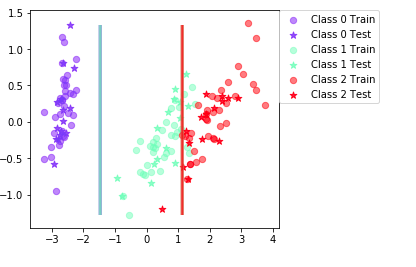
**Can we make up for not using a more advanced model in the basic classifier (e.g. independent features) by using boosting?**

Yes. By boosting we reduce variance to increase testing accuracy.

You may use the function plotBoundary provided in labfuns.py to plot the decision boundary for different datasets and parameters.

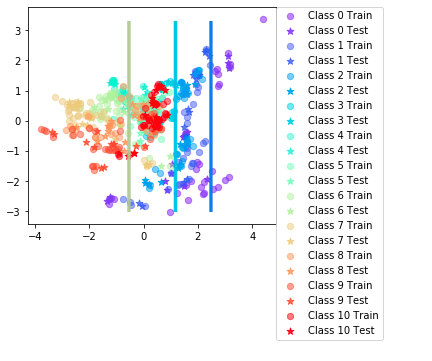
1. **Assignment 6**

Using DecisionTreeClassifier to classify Iris, mean classification accuracy 92.4 (SD 3.71)



**Figure 5**: Decision boundary of Decision tree classifier (Data Iris)

Using DecisionTreeClassifier to classify Vowels, mean classification accuracy 64.1 (SD 4)



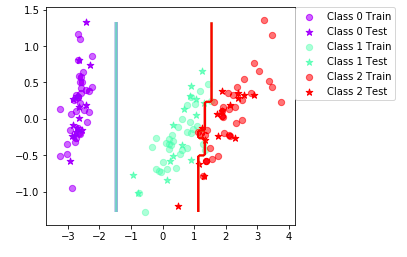
**Figure 6**: Decision boundary of Decision tree classifier (Data Vowel)

Test the decision tree classifier on the vowels and iris data sets.

Repeat but now by passing it as an argument to the BoostClassifier object. Answer questions 1-3 in assignment 5 for the decision tree.

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

**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?**

****

**Figure 6**: Decision boundary of Boosted Decision tree classifier (Data Iris)

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

Yes, because we can reduce bias and variance

1. **Assignment 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:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criteria** | **Naïve Bayes** | | **Decision tree** | |
| Outliers | Sensitive to outliers because posteriori is estimated from mean and variance | | Less sensitive to outliers | |
| Irrelevant inputs: part of the feature space is irrelevant | Less sensitive to irrelevant inputs because the covariance is assumed 0, probability of Y given relevant Xs is not affected by Y given irrelevant Xs | | Sensitive to irrelevant inputs | |
| Predictive power | Basic | Boosted | Basic tree | Boosted tree |
| Iris | 89 (SD 4.16) | 94.1 (SD 6.72) | 92.4 (SD 3.71) | 94.6 (SD 3.67) |
| Vowel | 64.7 (SD 4.03) | 80.2 (SD 3.52) | 64.1 (4) | 87.1 (SD 2.67) |
| Mixed types of data: binary, categorical or continuous features, etc. | Hard to estimate parameter with mixed data | | Simpler with decision tree, only cut-off point for each node needed | |
| Scalability:  Number of data dimension  Number of instances, N  Both | Complex when dimension space is large. Overfitting  Good when N is large  Less suffer from curse dimensionality | | Complex when dimension is large. Distance between data point is large when there are many features. Overfitting  Good when N is large  Suffer from curse of dimensionality, i.e., ratio D/N is large. | |