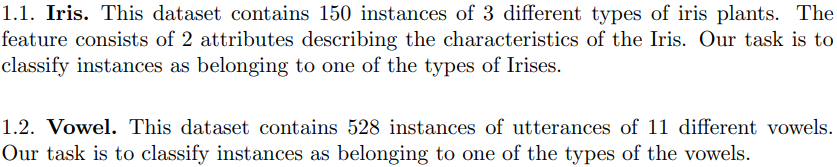
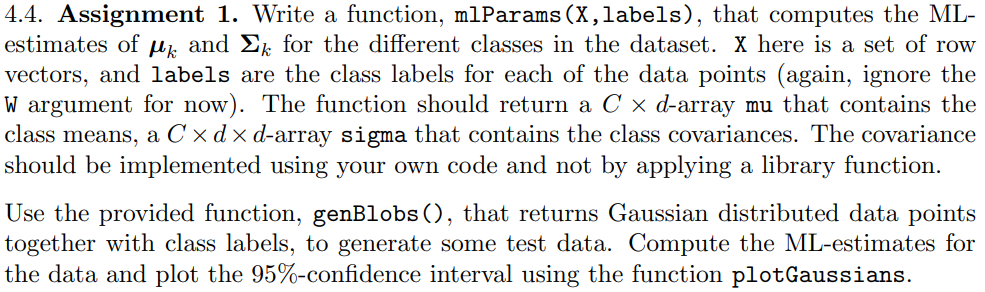
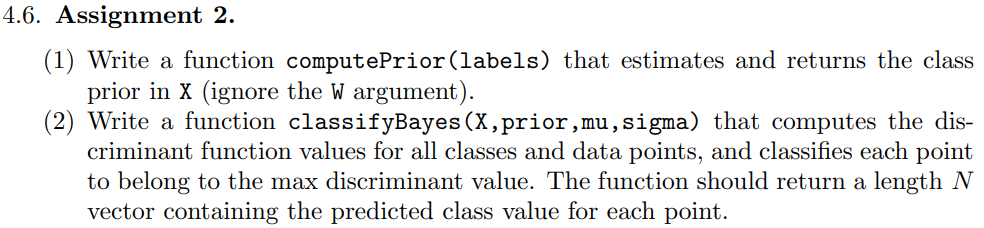
Lab 3: Bayesian Learning and Boosting

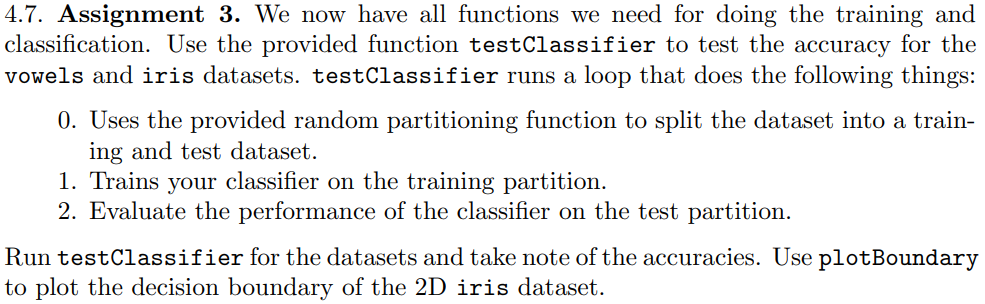








*completed*

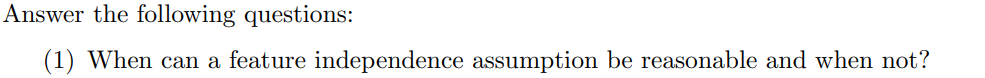


**Bayes classifier – iris**

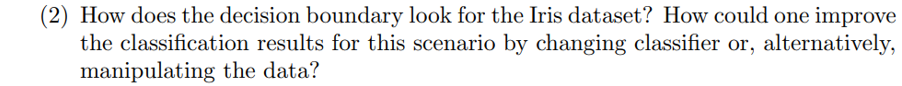
|  |
| --- |
| Trial: 0 Accuracy 84.4 |
| Trial: 10 Accuracy 95.6 |
| Trial: 20 Accuracy 93.3 |
| Trial: 30 Accuracy 86.7 |
| Trial: 40 Accuracy 88.9 |
| Trial: 50 Accuracy 91.1 |
| Trial: 60 Accuracy 86.7 |
| Trial: 70 Accuracy 91.1 |
| Trial: 80 Accuracy 86.7 |
| Trial: 90 Accuracy 91.1 |
| Final mean classification accuracy 89 with standard deviation 4.16 |
|  |

**Bayes classifier – vowel**

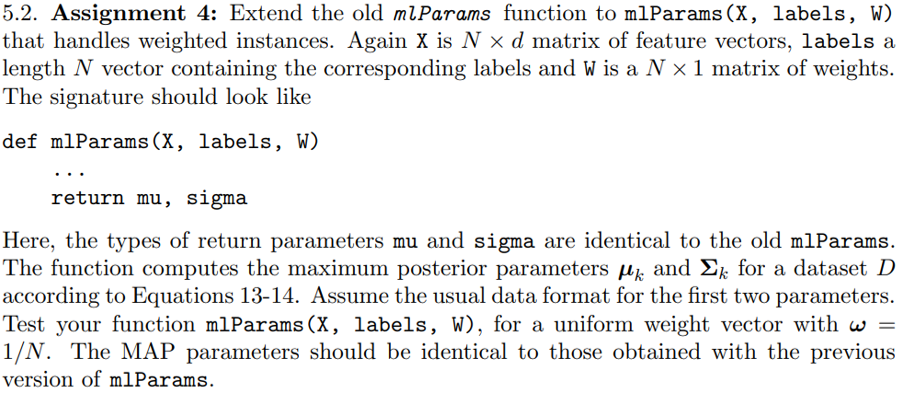
|  |
| --- |
| Trial: 0 Accuracy 61 |
| Trial: 10 Accuracy 66.2 |
| Trial: 20 Accuracy 74 |
| Trial: 30 Accuracy 66.9 |
| Trial: 40 Accuracy 59.7 |
| Trial: 50 Accuracy 64.3 |
| Trial: 60 Accuracy 66.9 |
| Trial: 70 Accuracy 63.6 |
| Trial: 80 Accuracy 62.3 |
| Trial: 90 Accuracy 70.8 |
| Final mean classification accuracy 64.7 with standard deviation 4.03 |



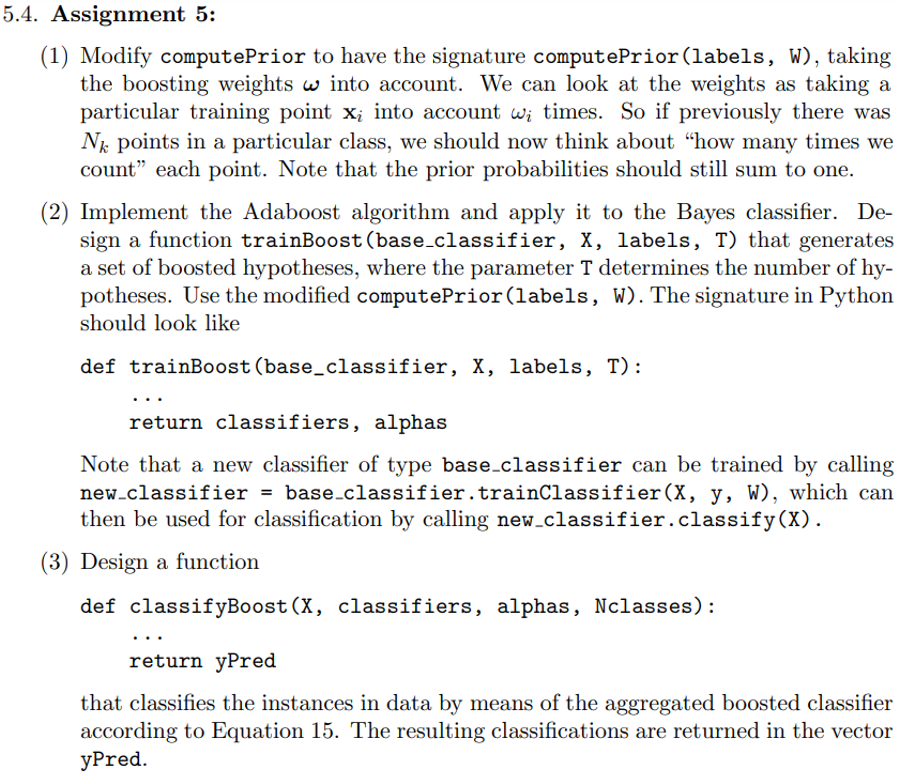
It’s reasonable when features are conditionally independent, or mostly independent. When it’s applied on data with strongly dependent features a lot of information will be lost. Therefore, it would be wiser to choose another model in such an instance instead.



There seems to be a lot of bias and low variance due to the boundary having a hard time following the datapoints. The decision boundary could be improved by switching classifier, or by implementing one of the ensemble learning methods (bagging or boosting). Though, there could probably be some improvements by manipulating the data to decrease the bias seen in the plot.



*completed*

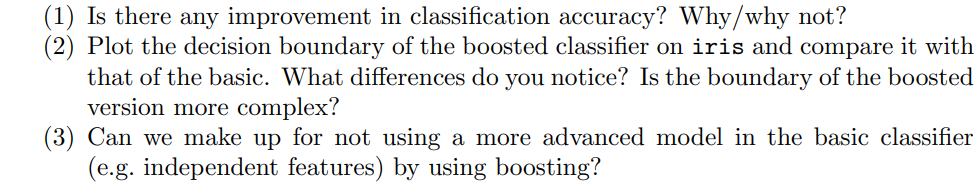


**Bayes classifier – iris (boost)***One of the individual classifiers predicted everything correctly, therefore epsilon == 0 which caused an error when np.log(epsilon) was called. Alpha was set to 2 as that is roughly the value that the best classifiers obtained. Occurred due to the small and simple dataset of 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 94.7 with standard deviation 2.82 |
|  |

**Bayes classifier – vowel (boost)**

|  |
| --- |
| 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 80.2 with standard deviation 3.52 |



Iris dataset:

Final mean classification accuracy **89** with standard deviation **4.16**

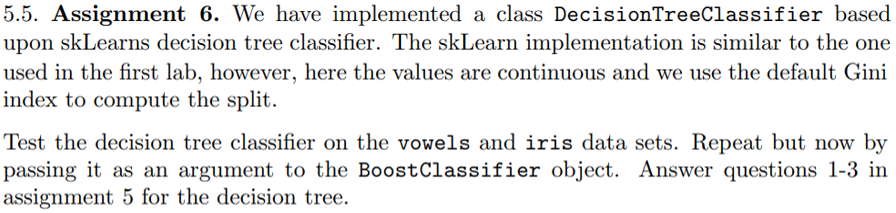
Final mean classification accuracy **94.7** with standard deviation **2.82** (boost)

Vowel dataset:

Final mean classification accuracy **64.7** with standard deviation **4.03**Final mean classification accuracy **80.2** with standard deviation **3.52** (boost)

In both datasets the mean classification accuracy improved when utilising boosting. This is due to the prediction after boosting is a weighted average of the predictions from the multiple classifiers. This improves the accuracy of basic model and can make up for some of the information lost (e.g. dependent features) when not utilising a more advanced model.

Boosting reduces bias. This creates a more complex boundary which can follow the datapoints better and yields an improved result.



**Decision tree – iris**

|  |
| --- |
| 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 |
|  |

**Decision tree – iris (boost)**

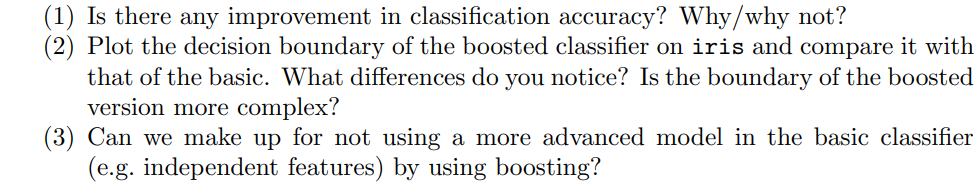
|  |
| --- |
| Trial: 0 Accuracy 95.6 |
| Trial: 10 Accuracy 100 |
| Trial: 20 Accuracy 95.6 |
| Trial: 30 Accuracy 93.3 |
| Trial: 40 Accuracy 93.3 |
| Trial: 50 Accuracy 95.6 |
| Trial: 60 Accuracy 88.9 |
| Trial: 70 Accuracy 93.3 |
| Trial: 80 Accuracy 93.3 |
| Trial: 90 Accuracy 93.3 |
| Final mean classification accuracy 94.6 with standard deviation 3.65 |
|  |

**Decision tree – vowel**

|  |
| --- |
| 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 |

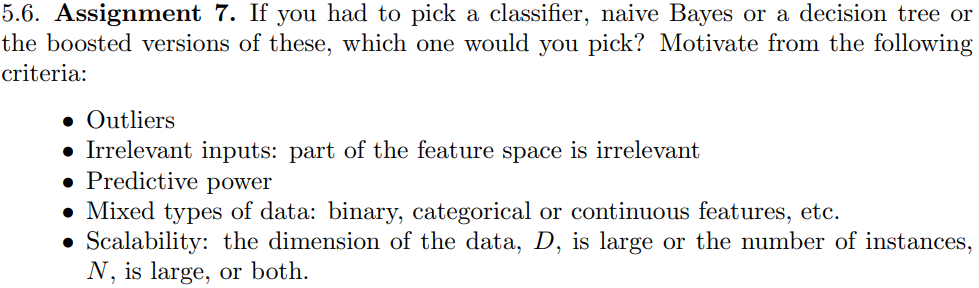
**Decision tree - vowel (boost)**

|  |
| --- |
| Trial: 0 Accuracy 85.7 |
| Trial: 10 Accuracy 90.3 |
| Trial: 20 Accuracy 88.3 |
| Trial: 30 Accuracy 90.9 |
| Trial: 40 Accuracy 84.4 |
| Trial: 50 Accuracy 81.2 |
| Trial: 60 Accuracy 87.7 |
| Trial: 70 Accuracy 86.4 |
| Trial: 80 Accuracy 87 |
| Trial: 90 Accuracy 90.3 |
| Final mean classification accuracy 86.7 with standard deviation 2.7 |



Iris dataset:  
Final mean classification accuracy **92.4** with standard deviation **3.71**Final mean classification accuracy **94.6** with standard deviation **3.65** (boost)Vowel dataset:  
Final mean classification accuracy **64.1** with standard deviation **4**  
Final mean classification accuracy **86.7** with standard deviation **2.7** (boost)

As previously an improvement can be seen when boosting is utilised. There is a decrease in bias and the boundary is more complex and follows the datapoints better. By boosting we can improve the more basic classifier.



* **Outliers:**

Decision trees are more prone to overfitting than Bayes but with good pruning it should handle outliers better, and not let them affect the classifiers as much.

Boosting will increase the weight of misclassified samples and generate a more complex boundary, which in this case, would probably lead to overfitting. Could be desired to include outliers, e.g. royal flush in poker.

* **Irrelevant inputs: part of the feature space is irrelevant:**Decision trees use information gain when constructing its tree and handles irrelevant inputs better than Bayes.

Boosting increases the significance of misclassified samples would reduce the performance of the classifier.

* **Predictive power:**  
  There was a clear advantage in utilising boosting for both Bayes and decision trees. Though, the decision tree seems to have performed slightly better, especially for the vowel dataset.

|  |  |  |
| --- | --- | --- |
| **Iris** |  |  |
| - Bayes | 94.7 | 2.82 |
| - Decision tree | 94.6 | 3.65 |
| **vowel** |  |  |
| - Bayes | 80.2 | 3.52 |
| - Decision tree | 86.7 | 2.7 |

* **N Mixed types of data: binary, categorical or continuous features, etc.:**Both decision trees and Bayes can be used for all data types given some pre-processing. Bayes easier with continuous than decision tree. Decision tree gives a step boundary, Bayes give smooth.
* **Scalability: the dimension of the data, *D*, is large or the number of instances, N is large, or both:**  
  Boosting makes the training of a classifier a lot slower which is a problem when there is a large dataset. In this lab, when boosting was used, a total of 1000 classifiers were created instead of only 100 (when running 100 trials). If computational power is a restrain than it might be preferable to not implement boosting and lose some accuracy.  
    
  In general decision trees perform better with a large dataset compared to Bayes, and the opposite is true for a small dataset.[[1]](#footnote-2)  
    
  Creating a decision tree classifier was quicker than Bayes, this was very noticeable when boosting was utilised. But, the implementation of Bayes could be more optimised so its not an entirely fair comparison.

1. Comment from stackoverflow… might be false. [↑](#footnote-ref-2)