Lab 3

Bayesian Learning and Boosting

DD2421

Yuqi Zheng
yuqizh@kth.se

Jingxuan Mao
jmao@kth.se

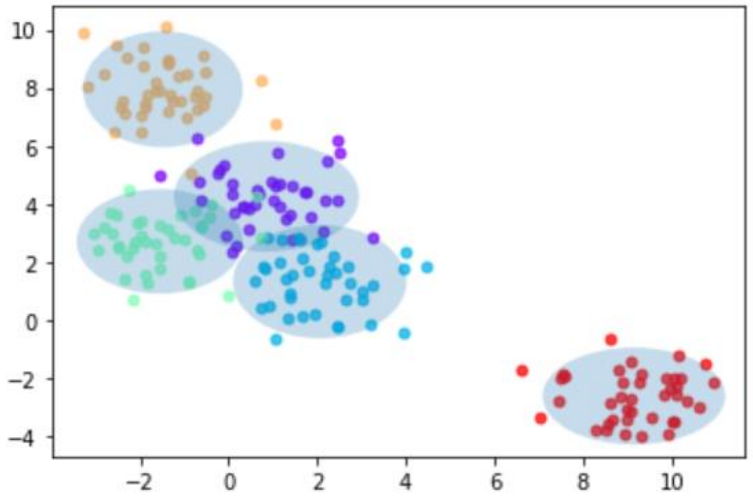
1 Bayesian Learning

Assignment 1

$$\boldsymbol{\mu}_k = \frac{\sum_{\{i|c_i=k\}} \mathbf{x}_i}{N_k}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{\{i|c_i=k\}} (\mathbf{x}_i - \boldsymbol{\mu}_k)^{\mathrm{T}} (\mathbf{x}_i - \boldsymbol{\mu}_k).$$

$$\boldsymbol{\Sigma}_k(m,m) = \frac{1}{N_k} \sum_{\{i|c_i=k\}} (\mathbf{x}_i(m) - \boldsymbol{\mu}_k(m))^2, \text{ and } \boldsymbol{\Sigma}_k(m,n) = 0 \text{ for } m \neq n.$$




```
Trial: 0 Accuracy 84.8
Trial: 10 Accuracy 90.5
Trial: 20 Accuracy 92.4
Trial: 30 Accuracy 90.5
Trial: 40 Accuracy 86.7
Trial: 50 Accuracy 85.7
Trial: 60 Accuracy 84.8
Trial: 70 Accuracy 92.4
Trial: 80 Accuracy 88.6
Trial: 90 Accuracy 88.6
Final mean classification accuracy  88.9 with standard
deviation 2.67
```

1 Bayesian Learning

Assignment 3

Accuracy of Iris Dataset:

```
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
```

Accuracy of Vowels Dataset:

```
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
```

1 Bayesian Learning

Assignment 3-1

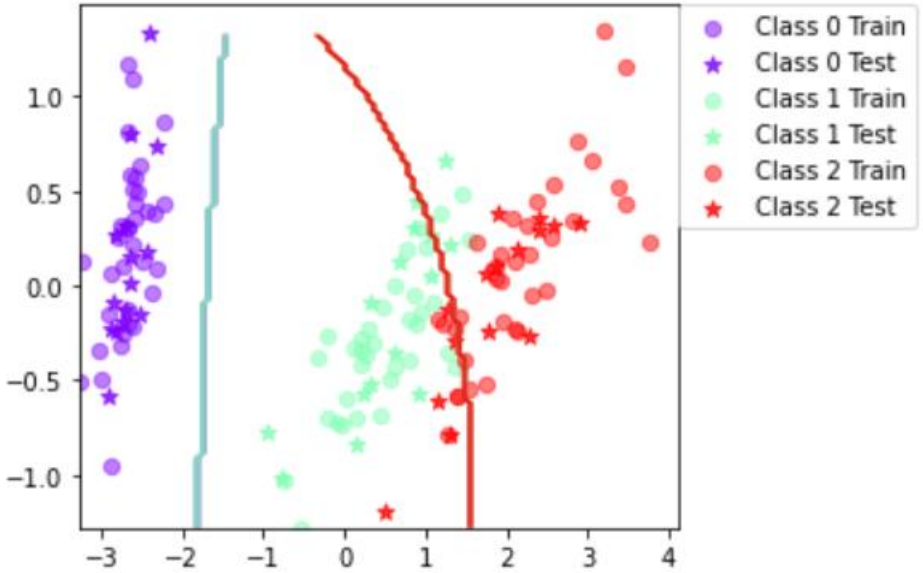
From the result, we can see that the accuracy of Iris dataset is much higher than that of the vowel dataset.

Feature independence assumption can be reasonable in the case of image classification since the image pixels can be considered as independent from each other. While speech is an inherently dynamic process whose features cannot be assumed as independent. Therefore, Bayesian assumption is not suitable for speech classification.

1 Bayesian Learning

Assignment 3-2

Split = 0.7



We can see that Class 0 and Class 1 can be separated accurately, while Class 1 and Class 2 are not clearly split by the boundary since they are overlapped and scattered. SVM with slack variables might have better performance in this case, where we can project the data to hyperplanes.

2 Boosting

Assignment 5

$$\boldsymbol{\mu}_k = \frac{\sum_{\{i|c_i=k\}} \omega_i \mathbf{x}_i}{\sum_{\{i|c_i=k\}} \omega_i} \tag{13}$$

$$\boldsymbol{\Sigma}_k(m,m) = \frac{1}{\sum_{\{i|c_i=k\}} \omega_i} \sum_{\{i|c_i=k\}} \omega_i (\mathbf{x}_i(m) - \boldsymbol{\mu}_k(m))^2, \text{ and } \boldsymbol{\Sigma}_k(m,n) = 0 \text{ for } m \neq n. \tag{14}$$

$$\omega_i^{t+1} = \frac{\omega_i^t}{Z^t} \times \begin{cases} e^{-\alpha^t} & \text{if } h^t(\mathbf{x}_i) = c_i \\ e^{\alpha^t} & \text{if } h^t(\mathbf{x}_i) \neq c_i \end{cases},$$

$$\alpha^t = \tfrac{1}{2}\left(\ln(1-\epsilon^t) - \ln(\epsilon^t)\right).$$

2 Boosting

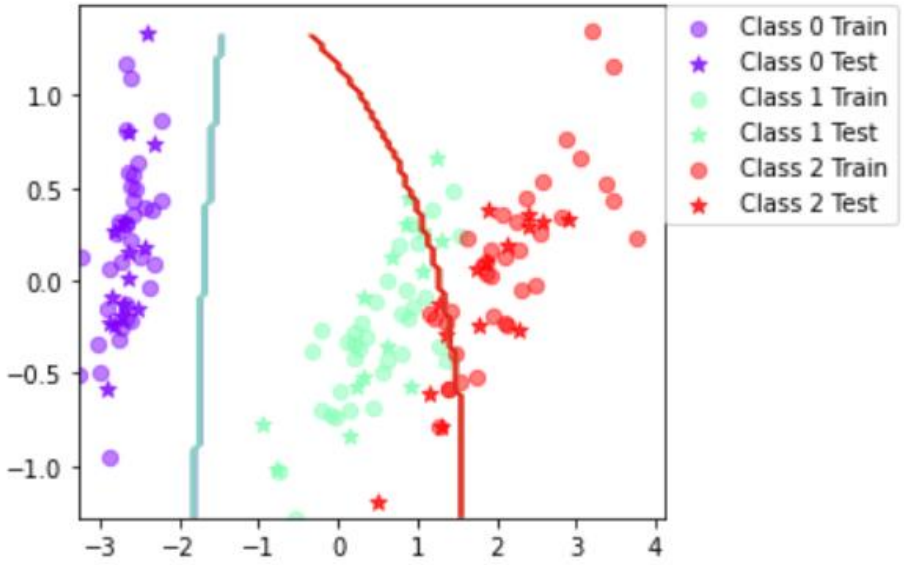
Original Vowels

```
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
```

Boosting Vowels

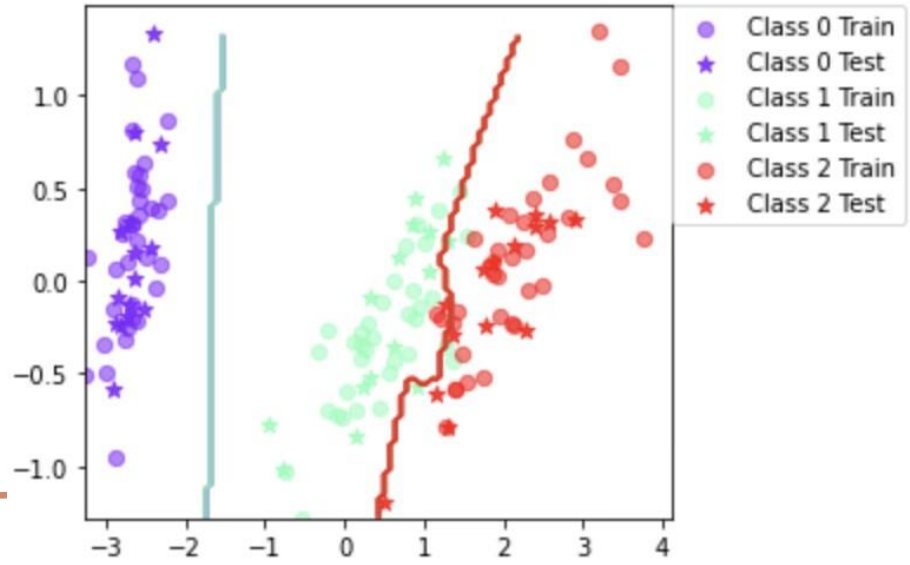
```
Trial: 0 Accuracy 76.6
Trial: 10 Accuracy 86.4
Trial: 20 Accuracy 83.1
Trial: 30 Accuracy 80.5
Trial: 40 Accuracy 72.7
Trial: 50 Accuracy 76
Trial: 60 Accuracy 81.8
Trial: 70 Accuracy 82.5
Trial: 80 Accuracy 79.9
Trial: 90 Accuracy 83.1
Final mean classification accuracy  80.2 with standard deviation 3.52
```

2 Boosting



Original Iris

```
Trial: 0 Accuracy 84.8
Trial: 10 Accuracy 90.5
Trial: 20 Accuracy 92.4
Trial: 30 Accuracy 90.5
Trial: 40 Accuracy 86.7
Trial: 50 Accuracy 85.7
Trial: 60 Accuracy 84.8
Trial: 70 Accuracy 92.4
Trial: 80 Accuracy 88.6
Trial: 90 Accuracy 88.6
Final mean classification accuracy  88.9 with standard deviation 2.67
```



Boosting 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
```

2 Boosting

Assignment 5

There are significant improvements in the classification accuracy of both datasets. This is because we increase the weights of the misclassified data points in each iteration, forcing the classifier to focus on them. Also, the classifier with higher error rate will be given lower Alpha values, which means they are less trusted in the voting.

We can use boosting to make up for not using a more advanced model in the basic classifier.

2 Boosting

Assignment 6

Original Vowels

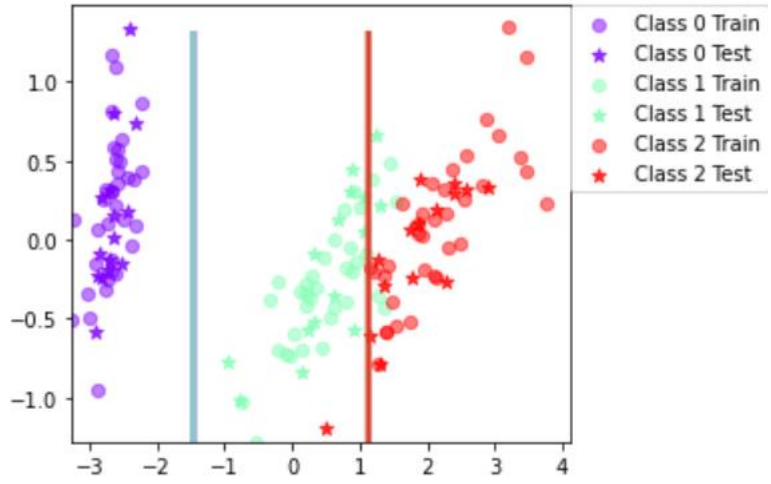
```
Trial: 0 Accuracy 63.6
Trial: 10 Accuracy 68.8
Trial: 20 Accuracy 63.6
Trial: 30 Accuracy 66.9
Trial: 40 Accuracy 59.7
Trial: 50 Accuracy 63
Trial: 60 Accuracy 59.7
Trial: 70 Accuracy 68.8
Trial: 80 Accuracy 59.7
Trial: 90 Accuracy 68.2
Final mean classification accuracy  64.1 with standard deviation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

Boosting Vowels

```
Trial: 0 Accuracy 87
Trial: 10 Accuracy 89.6
Trial: 20 Accuracy 86.4
Trial: 30 Accuracy 91.6
Trial: 40 Accuracy 80.5
Trial: 50 Accuracy 81.8
Trial: 60 Accuracy 89
Trial: 70 Accuracy 87.7
Trial: 80 Accuracy 84.4
Trial: 90 Accuracy 85.7
Final mean classification accuracy  86.5 with standard deviation 2.95
```

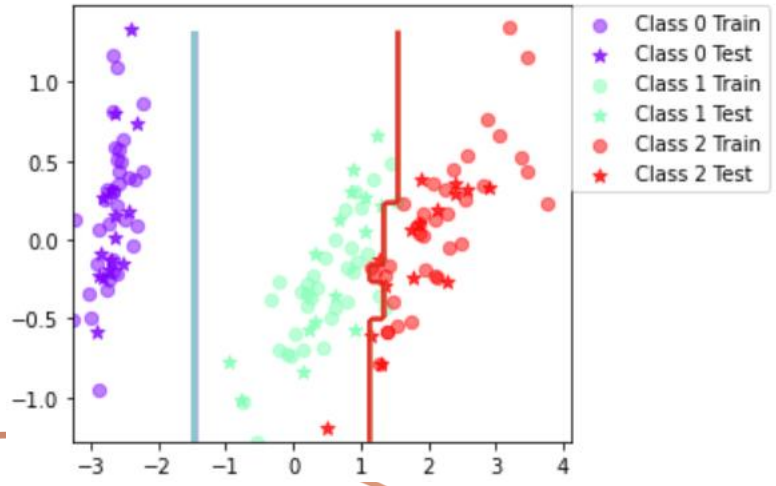
2 Boosting

Assignment 6




```
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
```

Original Iris




```
Trial: 0 Accuracy 95.6
Trial: 10 Accuracy 100
Trial: 20 Accuracy 95.6
Trial: 30 Accuracy 91.1
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
```

Boosting Iris

2 Boosting

Assignment 7

Outliers:

We should choose naive Bayesian classifier. The boosting classifiers will give too many weights to outliers. While the decision tree may lead to overfitting problems.

Irrelevant inputs:

We should choose decision tree classifier. Decision tree is good at dealing with independent data, using information gain to select the attribute in each level.

2 Boosting

Assignment 7

Predict power: Boosting version of the two classifiers

		mean	std
Iris	Bayesian	89	4.16
	Boosting Bayesian	94.7	2.82
	Decision Tree	92.4	3.71
	Boosting Decision Tree	94.6	3.67
Vowels	Bayesian	64.7	4.03
	Boosting Bayesian	80.2	3.52
	Decision Tree	64.1	4
	Boosting Decision Tree	86.5	2.95

2 Boosting

Assignment 7

Mixed types of data:

For continuous data, generative model, like Bayesian classifier performs better. While decision tree outperforms in the case of discrete data.

Scalability:

Decision tree classifier performs better when the dataset is large. We find Bayesian classifier take much longer time than the decision tree in the olivetti classification. And boosting will be even more time-consuming.

2 Boosting

Voluntary Assignment

Naive Bayesian

```
Trial: 0 Accuracy 88.3
Trial: 10 Accuracy 90.8
Trial: 20 Accuracy 85
Trial: 30 Accuracy 89.2
Trial: 40 Accuracy 89.2
Trial: 50 Accuracy 84.2
Trial: 60 Accuracy 91.7
Trial: 70 Accuracy 82.5
Trial: 80 Accuracy 81.7
Trial: 90 Accuracy 86.7
Final mean classification accuracy
87.7 with standard deviation 3.03
```

Decision Tree

```
Trial: 0 Accuracy 65.8
Trial: 10 Accuracy 57.5
Trial: 20 Accuracy 49.2
Trial: 30 Accuracy 50
Trial: 40 Accuracy 53.3
Trial: 50 Accuracy 44.2
Trial: 60 Accuracy 49.2
Trial: 70 Accuracy 54.2
Trial: 80 Accuracy 50
Trial: 90 Accuracy 52.5
Final mean classification accuracy
48.4 with standard deviation 6.45
```

Boosting Bayesian

```
Trial: 0 Accuracy 88.3
Trial: 10 Accuracy 90.8
Trial: 20 Accuracy 85
Trial: 30 Accuracy 89.2
Trial: 40 Accuracy 89.2
Trial: 50 Accuracy 85
Trial: 60 Accuracy 91.7
Trial: 70 Accuracy 82.5
Trial: 80 Accuracy 80
Trial: 90 Accuracy 86.7
Final mean classification accuracy
87.5 with standard deviation 3.12
```

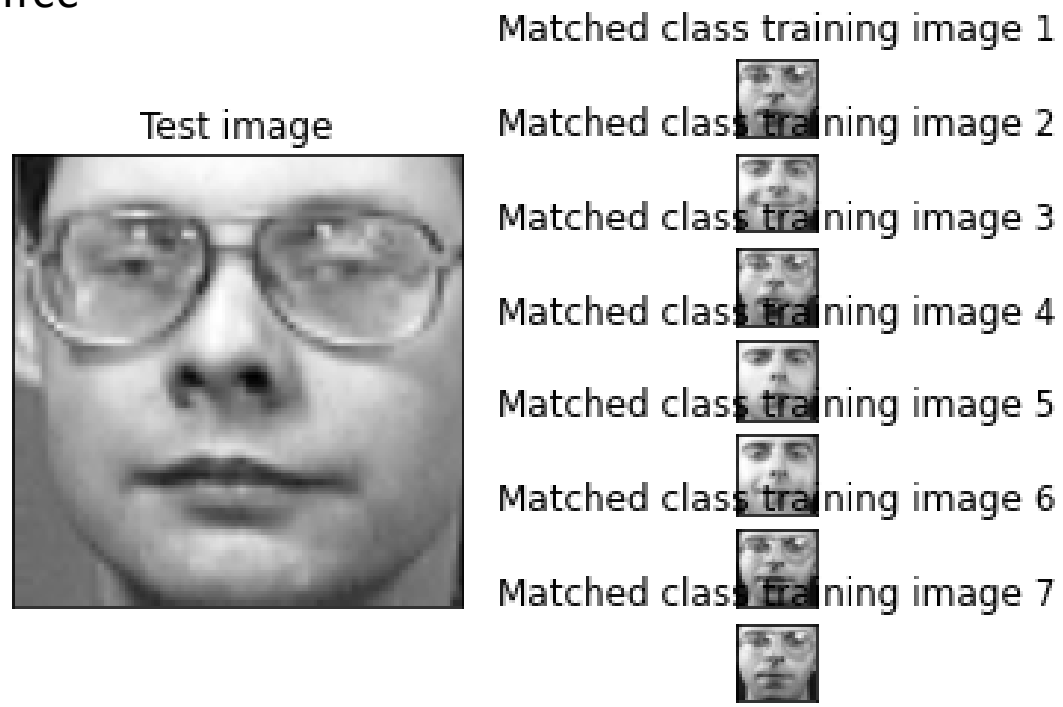
Boosting Decision Tree

```
Trial: 0 Accuracy 74.2
Trial: 10 Accuracy 71.7
Trial: 20 Accuracy 76.7
Trial: 30 Accuracy 73.3
Trial: 40 Accuracy 74.2
Trial: 50 Accuracy 66.7
Trial: 60 Accuracy 72.5
Trial: 70 Accuracy 63.3
Trial: 80 Accuracy 65.8
Trial: 90 Accuracy 70.8
Final mean classification accuracy
70.7 with standard deviation 6.73
```

2 Boosting

Voluntary Assignment

Decision Tree



Thanks for listening :)