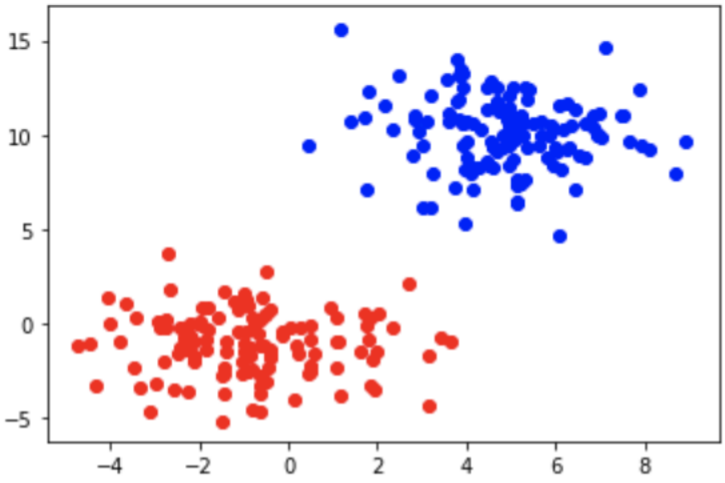
CSCI3320 Programming Assignment 2 Chung Tsz Ting / 1155110208

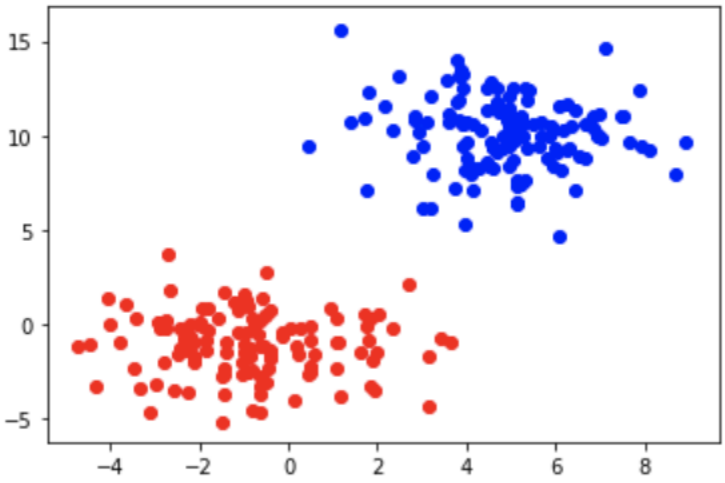
1)

Test with train()

```
w = train(X_train, y_train)
X_test_prediction = predict(X_test, w)
plot_prediction(X_test, X_test_prediction)
plot_prediction(X_test, y_test)

wrong = np.count_nonzero(y_test - X_test_prediction)
print ('Number of wrong predictions is: ' + str(wrong))
```



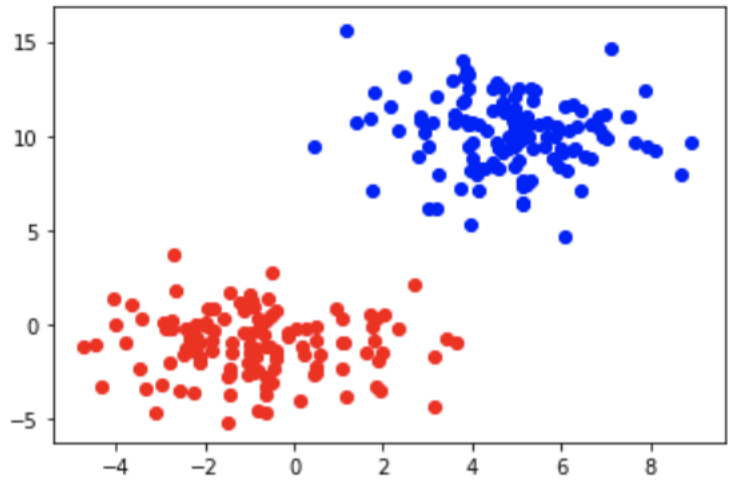


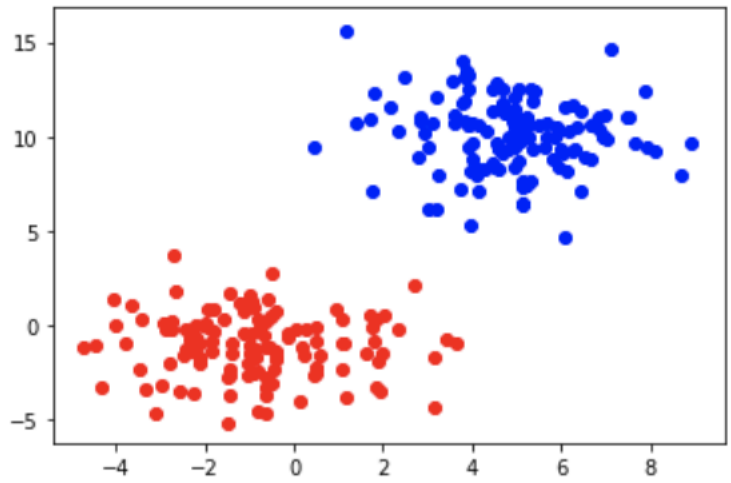

```
Number of wrong predictions is: 0
```

Test with train_matrix()

```
w = train_matrix(X_train, y_train)
X_test_prediction = predict(X_test, w)
plot_prediction(X_test, X_test_prediction)
plot_prediction(X_test, y_test)

wrong = np.count_nonzero(y_test - X_test_prediction)
print ('Number of wrong predictions is: ' + str(wrong))
```

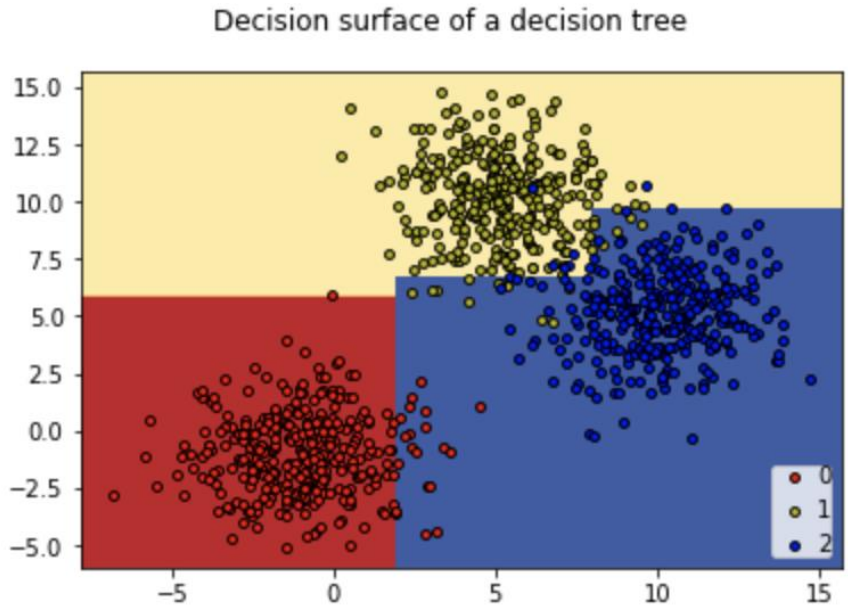





```
Number of wrong predictions is: 0
```

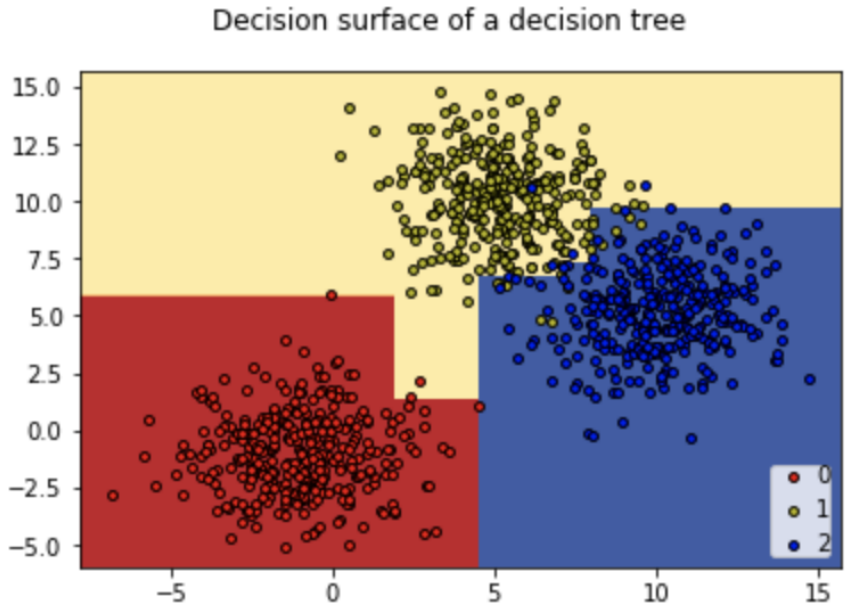
2)

max_dept set to 3:



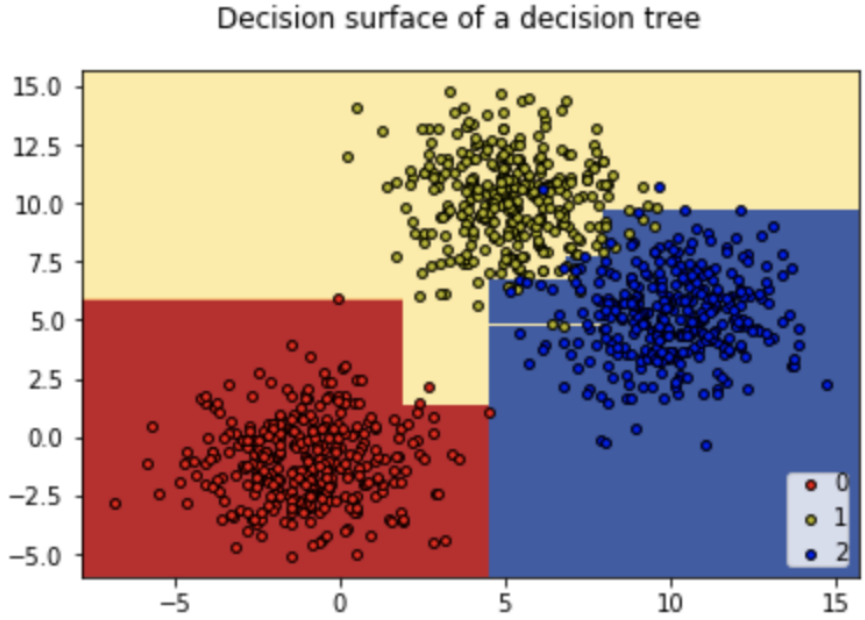

```
Number of wrong predictions is: 15
```

max_dept set to 5:




```
Number of wrong predictions is: 8
```

max_dept set to 7:




```
Number of wrong predictions is: 8
```

Explanation:
The number of wrong predictions reduces when the value of max_dept increases. This is because of the reduction in underfitting and reduce the bias while increase the variance for classification. Therefore, the number from max_dept equal to 3 increasing to 5 can reduce the number of wrong predictions (decrease from 15 to 8). However, when the max_dept allowed is further increased, the model start to overfit and we can see from the graph there is a yellow gap inside the blue region but that region which is initialized with center (5,10) should be classified as blue and the yellow instances are outliners. Thus, the starting of overfitting stop the accuracy from increasing and therefore there is no further improvement on the number of wrong predictions (i.e. still equal to 8).