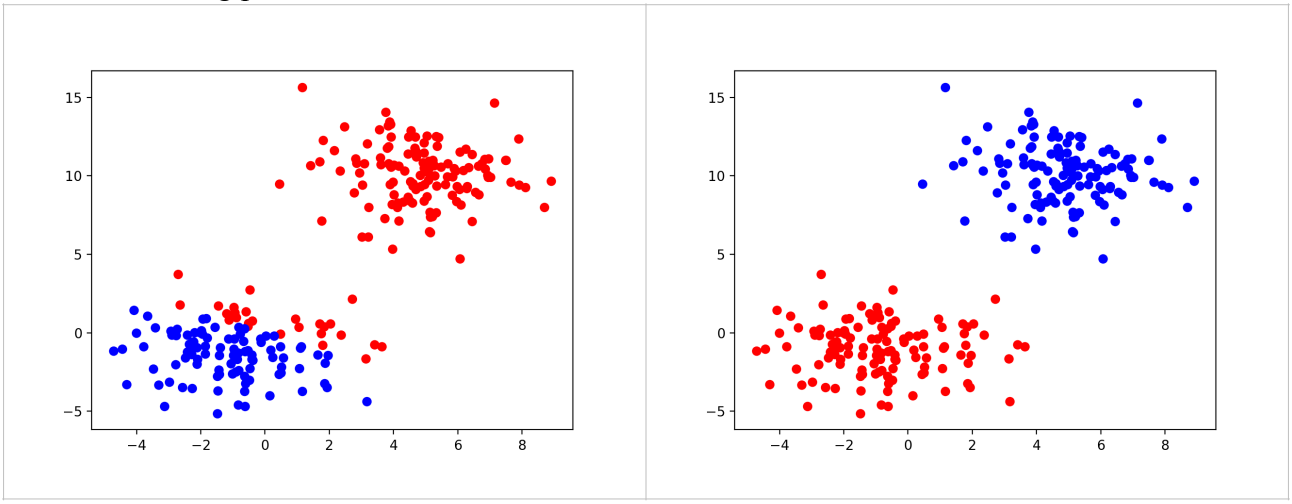
Assignment2

ZHANG Xinyu
1155091989

1.

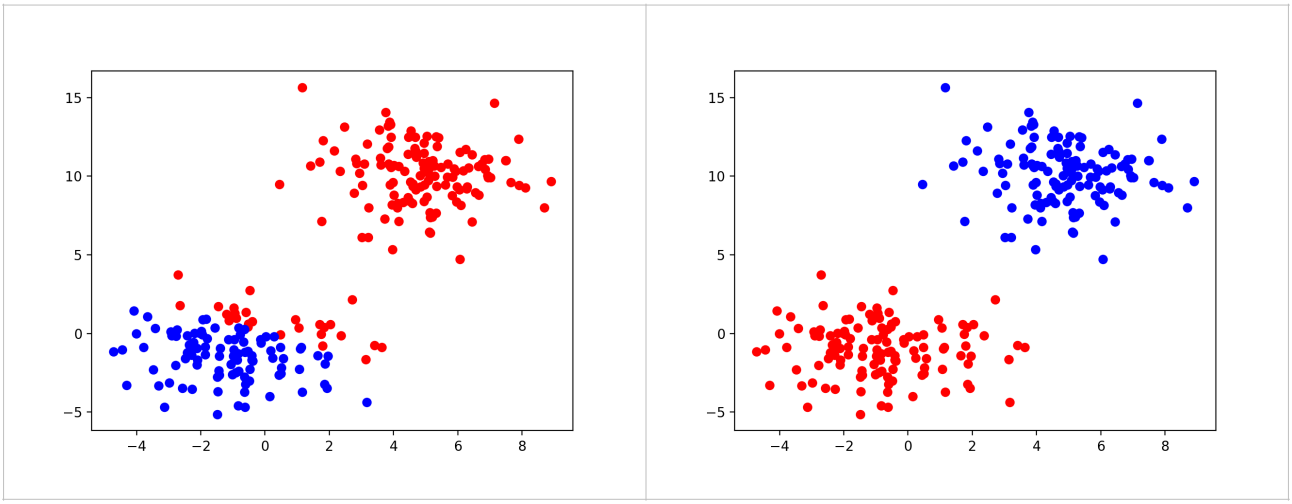
a. train()

Number of wrong predictions is: 27

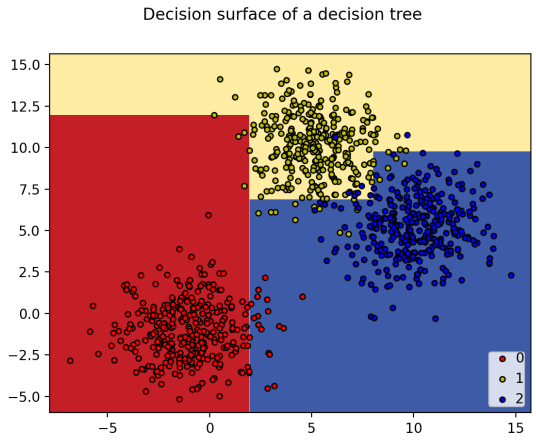
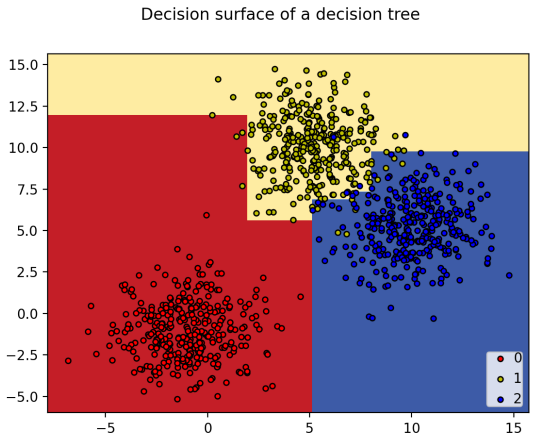
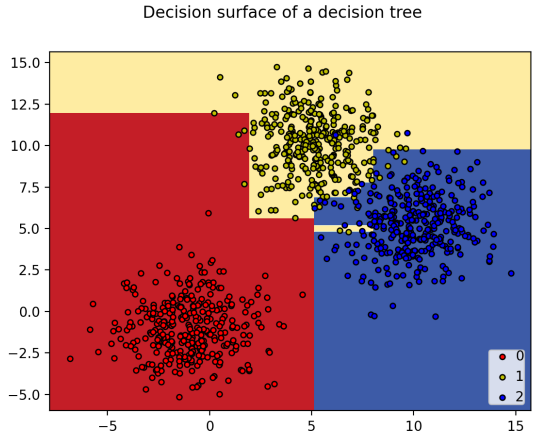


b. train_matrix()

Number of wrong predictions is: 27



2.

max_depth = 3	max_depth = 5
Number of wrong predictions is: 18	Number of wrong predictions is: 9
	
max_depth = 7	
Number of wrong predictions is: 10	
	

When the max_depth is 5, it gives the minimum number of wrong predictions. Given max_depth is 3, the error number is 18. And it drops to 9 when we add the max_depth up to 5. But if we set the max_depth higher, that is, to 7, the number of error growth to 10.

This may because when the decision tree that is too small (max_depth is 3) is susceptible to underfitting, that is, the tree's complexity is smaller than the underlying of data. Here, as we need to increase the complexity of decision tree, the number of wrong prediction decreases. But if the decision tree is too complex (max_depth is 7), the data is not enough to constrain it, which is overfitting. In such a case, we do not have a good generalization. In above examples, the case max_depth = 5 makes a good balance between the complexity of decision tree and generalization of model.