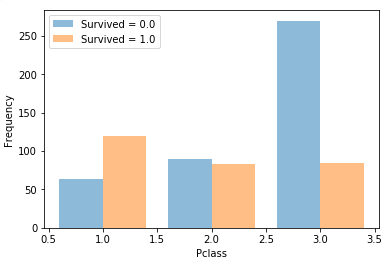
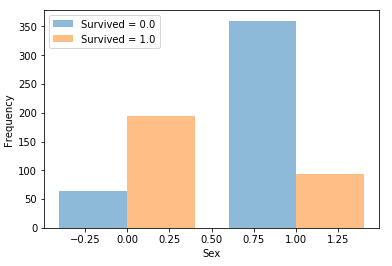
(a)



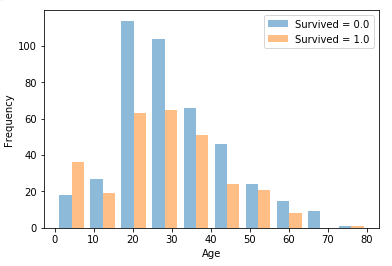
Pclass:

The first class has the highest survival rate. Then the second class. And the third class has the lowest survival rate.



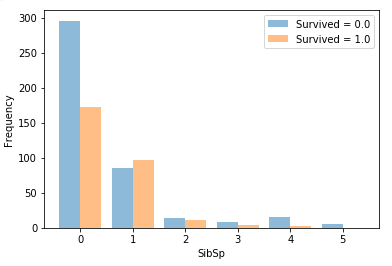
Sex:

Females’ survival rate is higher than males.



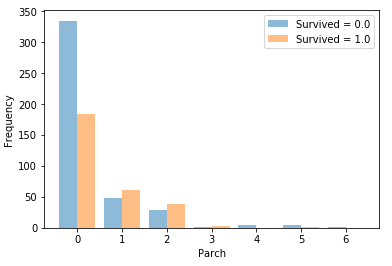
Age:

People whose age is around 20-50 has lowest survival rate. People under 10 are more likely to survive.



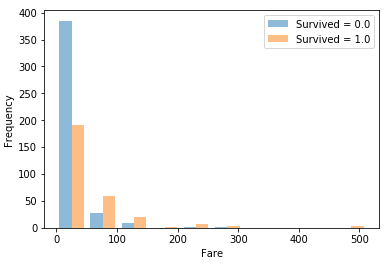
SibSip:

People who traveled with at least 1 sibling or spouse have higher survival rate than those who traveled alone.



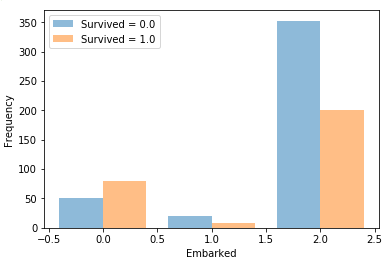
Parch:

People who traveled with at least one parent or child are more likely to survive than those who traveled alone.



Fare:

People who paid more for fare have a higher survival rate.



Embarked:

People embarking from Cherbourg have higher survival rate than embarking from other places.

(c)

The training error of this Decision Tree Classifier is 0.014

(d)

training error (n\_neighbors=3): 0.167

training error (n\_neighbors=5): 0.201

training error (n\_neighbors=7): 0.240

(e)

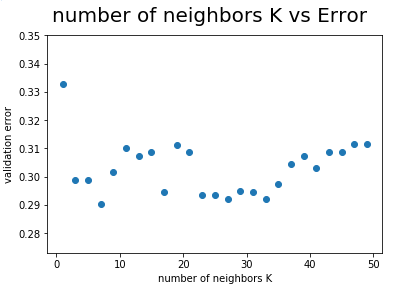
Training error and testing error for MajorityVoteClassifier are: 0.404, 0.407

Training error and testing error for RandomClassifier is: 0.489, 0.487

Training error and testing error for DecisionTreeClassifier is: 0.012, 0.241

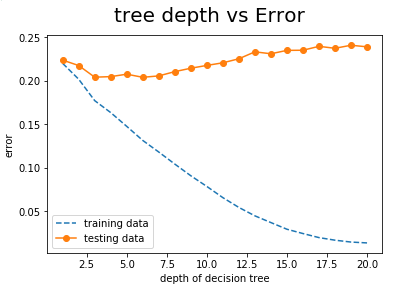
Training error and testing error for KNeighborsClassifier is: 0.212, 0.315

(f)



At first when k is bigger, validation error decreases. But later on we can observe overfitting, that is, when k is still getting bigger, validation error will increase. The best k is 7. The validation error for K=7 is 0.290

(g)



The best depth limit is 6. When the tree gets deeper, training error always decreases, but testing error will first decrease and then increase due to overfitting.

(h)

