CSPB 4622 - Truong - Machine Learning

Started on	Monday, 16 September 2024, 2:18 PM
State	Finished
Completed on	Monday, 16 September 2024, 2:30 PM
Time taken	12 mins 8 secs
Marks	11.00/11.00
Grade	**10.00** out of 10.00 (**100**%)

Question **1**

Correct

Mark 1.00 out of 1.00

In which of the following categories do you think the decision tree approach lies in?

- ⦿ a. It is a high variance method ✔
- ○ b. It is a high bias method
- ○ c. It is a low variance method

Your answer is correct.

Question **2**

Correct

Mark 1.00 out of 1.00

How do you decide which feature to pick while splitting the decision tree? Select all that apply:

☑

d. Compute information gain for each feature and split on the one with largest information gain ✔

☑

b. Select that feature to split on which decreases impurity the most ✔

☐

c. Select that feature to split on which increases impurity the most

☐

a. Compute information gain for each feature and split on the one with smallest information gain

Question **3**

Correct

Mark 1.00 out of 1.00

If you split on feature x_k at internal node J of the decision tree, you will never need to split on x_k again in any of J's descendants. Select one.

○ a. True

◉ b. False   ✔

> Your answer is correct.

Question **4**

Correct

Mark 1.00 out of 1.00

Consider you have a decision tree model, and it overfits. What can you do to reduce the overfitting? (Select all that apply)

☑

a Reduce the max_depth (maximum depth of the tree)

✔

☐

b Reduce the min_sample_leaf (minimum number of samples in a leaf node)

☑

d Increase the ccp_alpha (post pruning tuning parameter)

✔

☑

c Reduce the max_feature (maximum number of features to consider in a split)

✔

☑

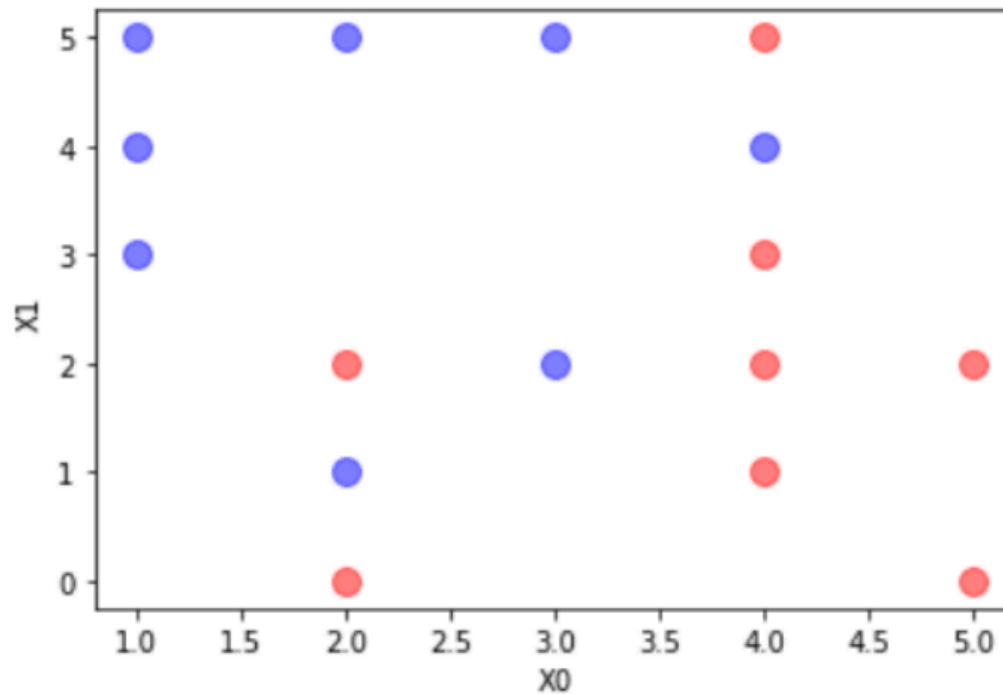e Increase min_impurity_decrease (minimum threshold of impurity decrease to split)

✔

☐

f Reduce the number of trees

Question **5**

Correct

Mark 1.00 out of 1.00

Consider the data X with the two features X0 and X1 and binary class Y (red or blue) as shown in the following graph.

You created a decision tree classifier model. What is the models first split criteria?

○ a. X0\<=1.5

○ b. X0\<=2.5

◉ c. X0\<=3.5   ✔

○ d. X0\<=4.5

> Your answer is correct. We want to obtain nodes as pure as possible for a Decision Tree classifier. We want the initial split to lead to the most considerable reduction in entropy possible. For the data in this question, that occurs when X0\<=3.5.

Question **6**

Correct

Mark 1.00 out of 1.00

What is the Gini of the box that satisfies the split condition above? (True for the condition) Please round up at the 4th digit below decimal.

Answer: 0.5000   ✔

Question **7**

Correct

Mark 1.00 out of 1.00

Entropy is a measure of impurity in a set of examples. When do you think entropy has the largest value?

Select one:

○ a. 75% examples are positive while 25% negative

○ b. All examples are in the same class

◉ c. Samples are equally balanced i.e equal number of examples of each class ✔

> Your answer is correct.
>
> When samples are equally balanced, entropy = 1(largest).

Question **8**

Correct

Mark 1.00 out of 1.00

When we talk about splitting strategy, what approach does the decision tree algorithm take for splitting, since we can't check all possible labelings?

Select one:

○ a. Using optimal substructures like in dynamic programming approach

◉ b. Greedy approach ✔

○ c. None of the above

○ d. Brute-force approach

> Your answer is correct.
>
> Decision Tree always takes a Greedy approach for splitting.

Question **9**

Correct

Mark 1.00 out of 1.00

N data samples are split into a train and a test data sets with a ratio of 60:40. After fitting with a k=1 KNN, the average prediction accuracy for the whole data set was 80%. What is the prediction accuracy of the test data set?
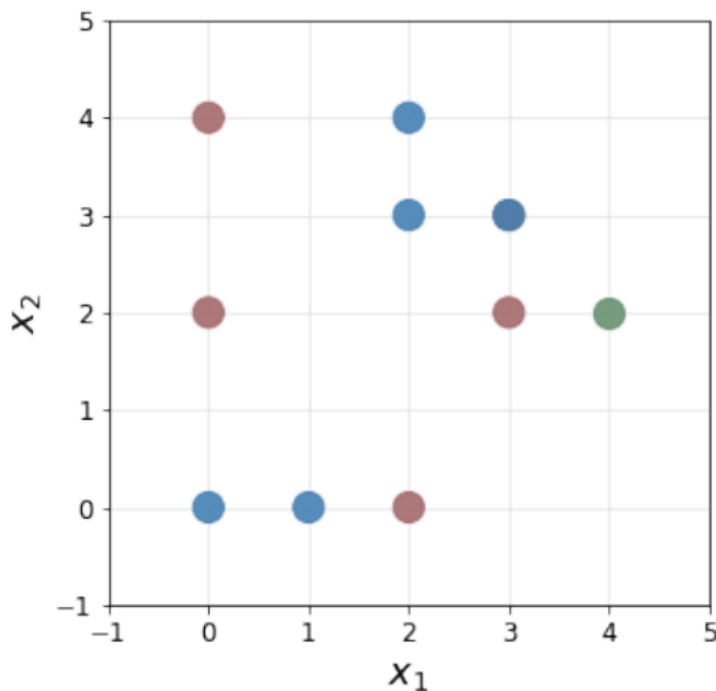
Answer: 0.5000   ✔

Question **10**

Correct

Mark 1.00 out of 1.00

Consider the training set shown above, with points labeled as red and blue. Suppose we want to predict the label of the green query point, $\mathbf{x}$. Use the standard (i.e. unweighted) formulation of KNN to answer the following questions. If you need to break a tie, decrease k by 1 until you break the tie. With $K = 2$, predict the label of $\mathbf{x}$
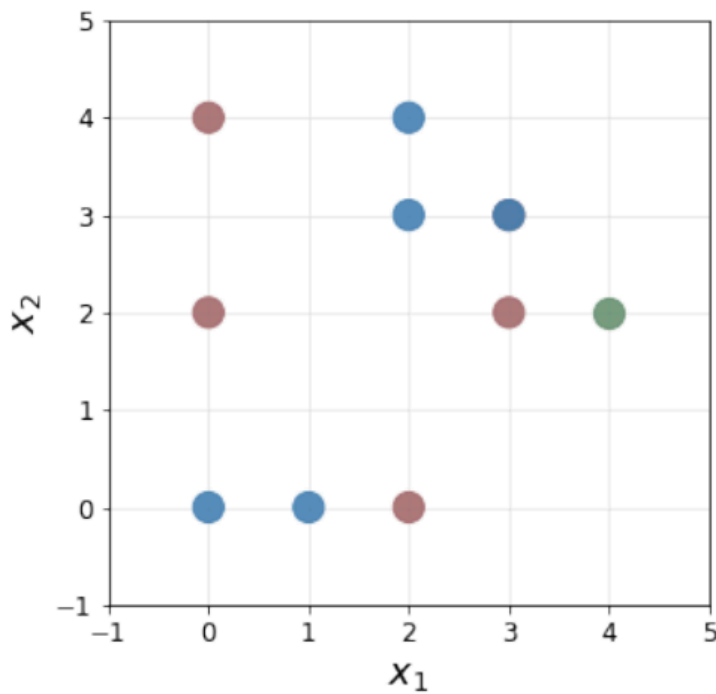
○ a. Blue

⊙ b. Red   ✔

Your answer is correct.

Question **11**

Correct

Mark 1.00 out of 1.00

Consider the training set shown above, with points labeled as red and blue. Suppose we want to predict the label of the green query point, $\mathbf{x}$. Use the standard (i.e. unweighted) formulation of KNN to answer the following questions. If you need to break a tie, decrease k by 1 until you break the tie. With $K = 3$, predict the label of $\mathbf{x}$

- ⦿ a. Blue   ✔
- ◯ b. Red

Your answer is correct.