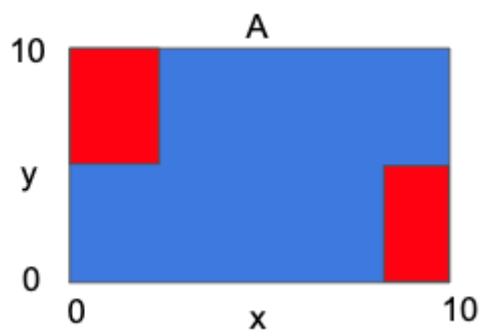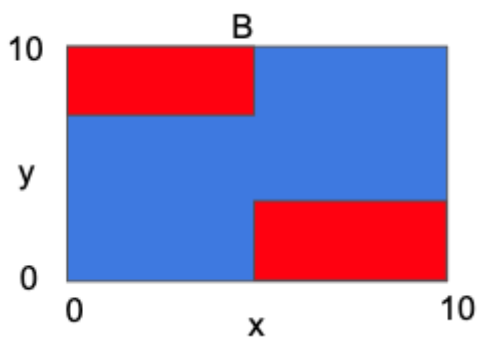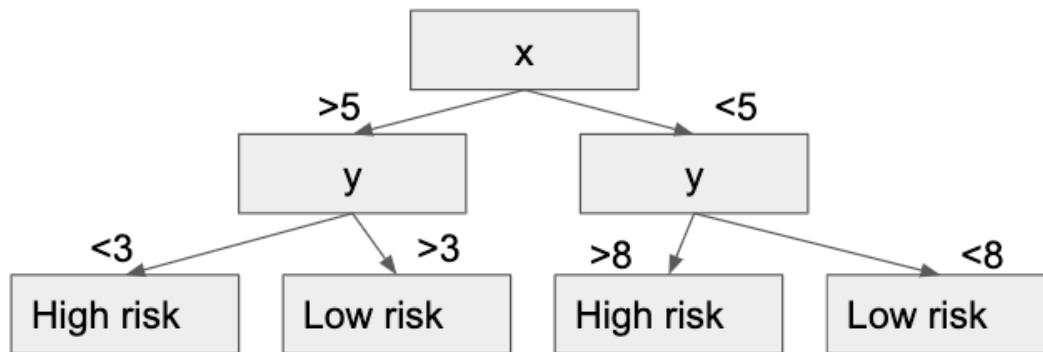# Week 2 Quiz

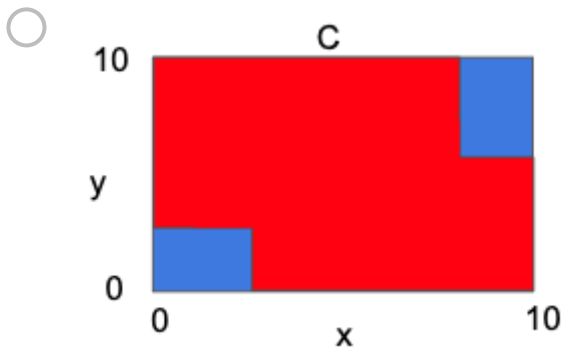**TOTAL POINTS 10**

---

1.  Which decision boundary corresponds to the following decision tree?

    In the options, red indicates high risk, blue indicates low risk.
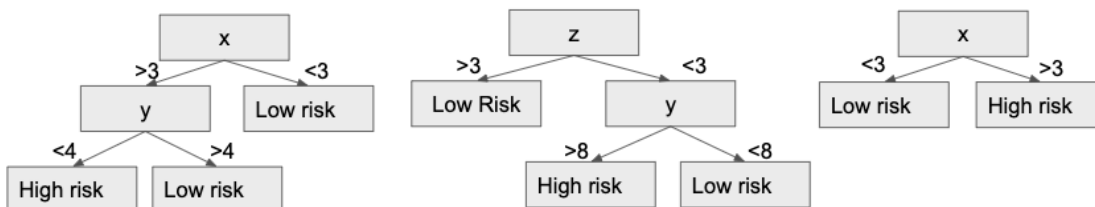


○ ●



○

2.  True or False: A tree of depth 1 is more expressive than a classical linear model.          1 point

    ○ True

    ⦿ False

3.  One way to aggregate predictions from multiple trees is by a majority vote. Using this          1 point
    aggregation rule, select the prediction of the following trees on the data point (x=4, y=7,
    z=2):



    ○ High risk

    ⦿ Low risk

4.  You've fit a random forest of 10 trees with max depth 20. Your training ROC is 0.99 and          1 point
    test ROC is 0.54. Which of the following is NOT a reasonable thing to try?

    ○ Increasing the number of trees

    ○ Trying a linear model

    ⦿ Increasing the max depth

○ Decreasing max depth

5. True or False: When your data is missing at random, then whether or not you are missing a covariate is completely independent of your outcome.    [ 1 point ]

HINT: the data is "missing at random" and not "missing completely at random".

◉ False

○ True

6. When is complete case analysis least likely to bias your model?    [ 1 point ]

○ Data is missing at random

○ Data is missing not at random

◉ Data is missing completely at random

○ Dropping data will bias your model in all of the above cases

7. You have created a model using mean imputation. At test time, you should fill in missing values with:    [ 1 point ]

○ 0.0

○ Mean from the test data

○ None of the above

◉ Mean from the train data

8. Let's say blood pressure (BP) measurements are more likely to be missing among young people, who generally have lower blood pressure. You use mean imputation to train your model. Which option correctly characterizes the mean BP (after imputation) in your training dataset?    [ 1 point ]

◉ It is higher than the true mean

○ It is lower than the true mean

○ It is the same as the true mean

○ None of the above

9. You have trained the following tree and want to make a prediction on someone, but all you know is they are 40 years old, and do not have their blood pressure.

1 point

From you dataset, you learn the linear relationship between blood pressure and age is:

$$BP = 1.7 \times Age + 60.$$

Using regression imputation, and the decision tree shown here, what is your prediction for this person's risk of heart attack?
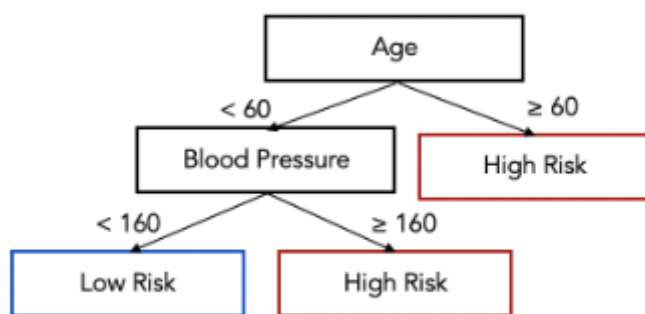


● Low Risk

○ High Risk

10. Assume you have missing data on one of your features, and are considering two options:

1 point

Option 1: Drop the feature that has missing values and fit a linear regression on the remaining features.

Option 2: Use imputation on the feature that has missing values, and fit a linear regression on all features.

True or False: "Both options have the same performance".

● False

○ True

☐ I, **Mirko Jerber Rodriguez Mallma**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code