✅ **Congratulations! You passed!**
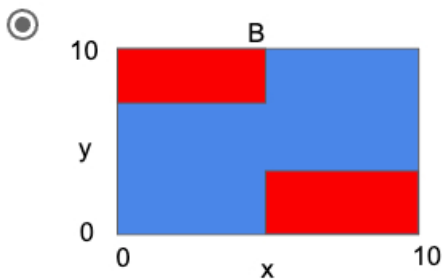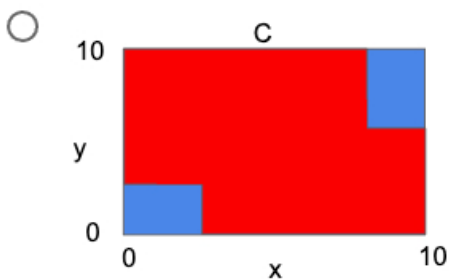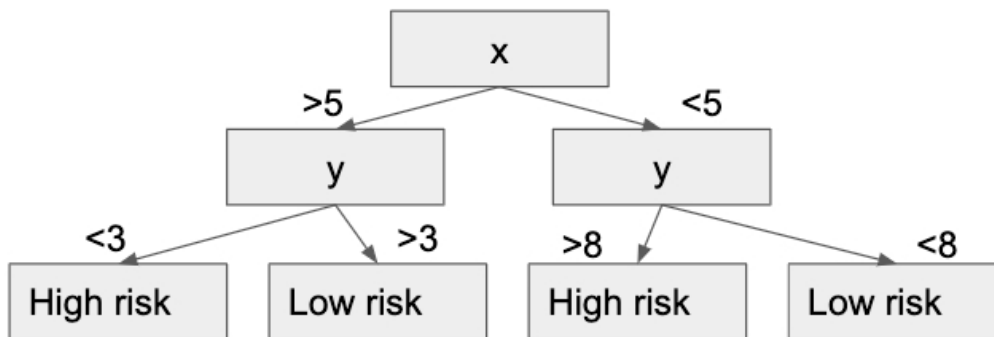
Grade received 100%   **To pass** 80% or higher

Go to next item

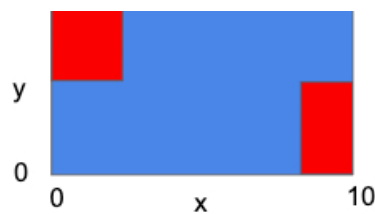**1.** Which decision boundary corresponds to the following decision tree?

1 / 1 point

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

```
                    x
           >5 /          \ <5
             y            y
      <3 /      \ >3   >8 /    \ <8
  High risk  Low risk  High risk  Low risk
```

○

C

◉

B

○

A

y
0

0   x   10

---

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

1 / 1 point

◉ False

○ True

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

1 / 1 point
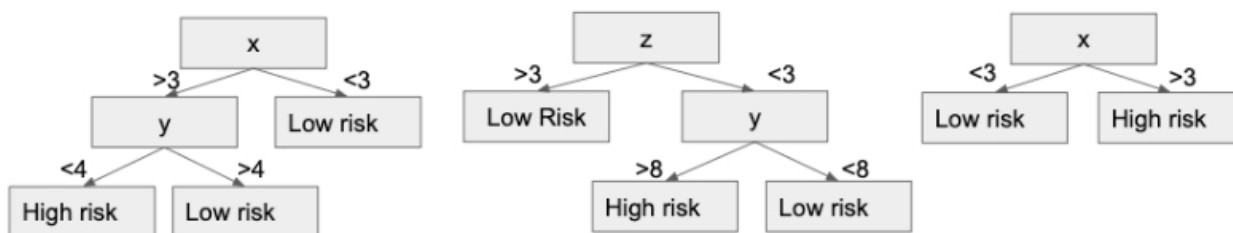
○ High risk

◉ Low risk

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

1 / 1 point

○ Trying a linear model

◉ Increasing the max depth

○ Increasing the number of trees

○ Decreasing max depth

> ✓ **Correct**
> The model is overfitting, since it's performing poorly on the test set compared to the training set.
>
> Increasing the depth INCREASES the variance of the trees, which you would do if the model was UNDER-fitting. Since the model is already OVER-fitting, increasing the model's variance would make it under-fit even more.
>
> Therefore, increasing max depth is NOT a reasonable thing to try in this case.

5. When is complete case analysis least likely to bias your model?

1 / 1 point

○ Data is missing at random

○ Data is missing not at random

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

◉ Data is missing completely at random

> ✓ **Correct**
> If data is missing completely at random, then there is no difference in the distribution of your data with and without the rows with missing data, so as long as you have enough complete data, dropping those rows should not bias your model.

6. You have created a model using mean imputation. At test time, you should fill in missing values with:

1 / 1 point

○ Mean from the test data

○ None of the above

○ 0.0

◉ Mean from the train data

> ✓ **Correct**
> Explanation: When testing your model, you should use the mean from your training data. Otherwise, your test statistics will not be reflective of your training procedure and how they will perform in the real world, since in the real world you will not have access to the entire collection of test examples beforehand.

7. 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 / 1 point

○ None of the above

○ It is lower than the true mean

◉ It is higher than the true mean

○ It is the same as the true mean

8. 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 / 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?
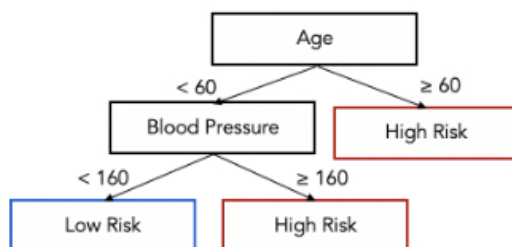
◉ Low Risk

○ High Risk

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

   1 / 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".

○ True

◉ False