Homework Week 6

Part 1 -Classification Trees

You will need the following image for a decision trees
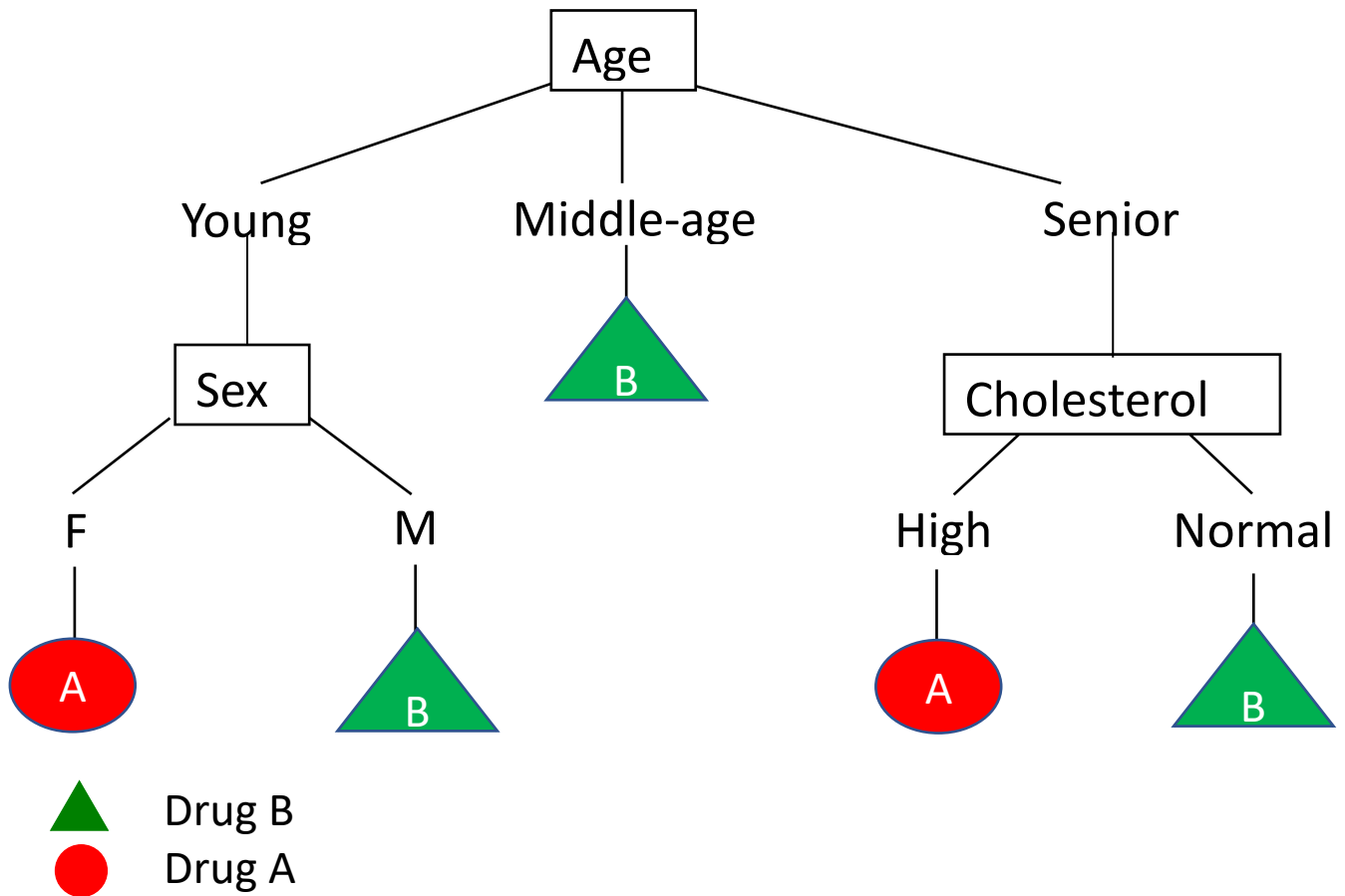

Image form Machine Learning with Python

1) What is the depth of this Decision Trees?
   A. 1
   B. 2
   C. 3
   D. 4

2) How many leaf nodes are in the following Decision Tree?

A. 1
B. 2
C. 3
D. 4

3) What is the class of the following feature $x = [\,young, male\,]$
    A. Drug-A
    B. Drug-B

4) You can use categorical features in a Decision Tree
    A. True
    B. False
5) You can use categorical features in a Decision Tree without sklearn without pre-processing
    A. True
    B. False

Part 2 : Bagging

6) Bagging can improve the generalization error of

    A. only trees
    B. any classifier
    C. classifiers that work well on the training data but not on the test data
    D. None above

7) You use Bagging for a classification problem, your training error is 0% your test error is 50% what should you do
    A. Increase the depth of the tree
    B. Increase the number of predictors
    C. decrease the number of predictors

8) When you are performing Bagging if you add more predictors it will lead to overfitting
    A. Always
    B. Rarely -if the dataset has lots of noise
    C. Never

9) Consider the following each column is a different tree constructed from a Random Forest algorithm

Tree 1	Tree 2	Tree 3	Tree 4	Tree 4
1	3	1	1	2

    A. 1
    B. 2
    C. 3

10) The main difference between Bagging with Trees and Random Forest is

    A. Only Bagging randomly samples data samples to build trees
    B. Only Random Forest randomly samples data samples to build trees
    C. Only Bagging randomly samples data features to build trees
    D. Only Random Forest randomly samples data features to build trees

11) What is the main reason we randomly construct trees?
    A. Prevent underfitting
    B. By making each tree a little different they are less likely all to make the same mistake
    C. If you don't have enough samples
    D. To preprocess data

12) Random Forest and Bagging are the best performing and most popular ensemble methods
    A. True, although there are other ensemble methods Random Forest and Bagging usually perform best
    B. False, Gradient Boosting is another popular ensemble method that performs comparable or even better, but it's more difficult to train due to overfitting

Solutions

    1. B
    2. D
    3. B
    4. A
    5. B
    6. C
    7. B

8. B
9. A
10. D
11. B
12. B