Try again once you are ready
Grade received 75%
To pass 80% or higher
 Try again

1. Ensemble learning is most effective when the outputs are aggregated from models that follow the exact same methodology all using the same dataset.

    1 / 1 point

    ◯ True
    ⦿ False

    ✓ Correct
    Ensemble learning is most effective when the outputs are aggregated from models that follow different methodologies—for instance, a logistic regression, a Naive Bayes model, and a decision tree classifier. In this way, any errors will be uncorrelated.

2. What are some of the benefits of ensemble learning? Select all that apply.

    1 / 1 point

    ☑ The predictions have less bias than other standalone models.

    ✓ Correct
    Ensemble learning combines the results of many models to help make more reliable predictions. Also, these predictions have less bias and lower variance than other standalone models. In order to work, ensemble learning requires numerous base learners, all trained on a random subset of the training data.

    ☑ It combines the results of many models to help make more reliable predictions.

    ✓ Correct
    Ensemble learning combines the results of many models to help make more reliable predictions. Also, these predictions have less bias and lower variance than other standalone models. In order to work, ensemble learning requires numerous base learners, all trained on a random subset of the training data.

    ☐ It requires few base learners trained on the same dataset.

    ☑ The predictions have lower variance than other standalone models.

    ✓ Correct
    Ensemble learning combines the results of many models to help make more reliable predictions. Also, these predictions have less bias and lower variance than other standalone models. In order to work, ensemble learning requires numerous base learners, all trained on a random subset of the training data.

3. In a random forest, what type of data is used to train the ensemble of decision-tree base learners?

    1 / 1 point

    ⦿ Bootstrapped
    ◯ Duplicated

○ Sampled

○ Unstructured

⊘ Correct
In a random forest, bootstrapped data is used to train the ensemble of decision-tree base learners. Bootstrapping refers to sampling with replacement. So, a random forest model will grow each of its trees by taking a random subset of the available features in the training data, then splitting each node at the best feature available to that tree.

4. Fill in the blank: When using a decision tree model, a data professional can use _____ to control the threshold below which nodes become leaves.

0 / 1 point

◉ min_samples_leaf

○ max_depth

○ max_features

○ min_samples_split

⊗ Incorrect
When using a decision tree model, a data professional can use min_samples_split to control the threshold below which nodes become leaves.