1. Which of the following is an example of clustering? 1 point

- (●) Separate the data into distinct groups by similarity
- ( ) Creating a new representation of the data with fewer features
- ( ) Compress elongated clouds of data into more spherical representations
- ( ) Accumulate data into groups based on labels

2. Which of the following are advantages to using decision trees over other models? (Select all that apply) 1 point

- [x] Decision trees often require less preprocessing of data
- [ ] Decision trees are naturally resistant to overfitting
- [ ] Decision trees are highly efficient on high-dimensional data
- [x] Decision trees are easy to interpret and visualize
- [ ] Decision trees can learn complex statistical models using a variety of kernel functions

3. What is the main reason that each tree of a random forest only looks at a random subset of the features when building each node? 1 point

- ( ) To reduce the computational complexity associated with training each of the trees needed for the random forest.
- ( ) To increase interpretability of the model
- ( ) To learn which features are not strong predictors
- (●) To improve generalization by reducing correlation among the trees and making the model more robust to bias.

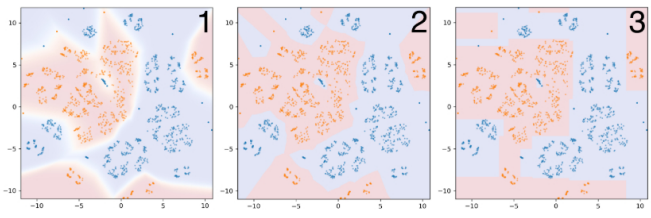
4. For which of the following supervised machine learning methods is it usually important to use some form of feature normalization/scaling? (Select all that apply) 1 point

- [x] Neural Networks
- [ ] Regularized logistic regression
- [ ] Decision Trees
- [x] Support Vector Machines
- [x] K-Nearest Neighbors (KNN)
- [ ] Naive Bayes

5. Select which of the following statements are true. 1 point

- [ ] For having an audience easily interpret the most important features in a fitted classification model, a **support vector machine** would be a better choice than a **decision tree**.
- [ ] For a fitted model that doesn't take up a lot of memory, **KNN** would be a better choice than **logistic regression.**
- [x] For predicting income over time from future sales of a new product, **linear regression** would be a better choice than a **k-nearest neighbors regressor**.
- [x] For a model that won't overfit a training set, **Naive Bayes** would be a better choice than a **decision tree**.

6. Match each of the prediction probabilities decision boundaries visualized below with the model that created them. 1 point

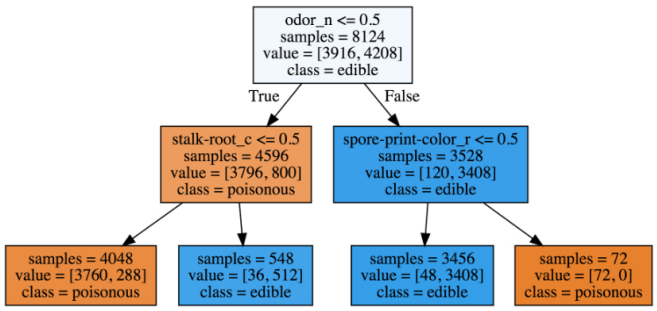


- ( )
  1. Neural Network
  2. Decision Tree
  3. KNN (k=1)
- ( )
  1. KNN (k=1)
  2. Decision Tree
  3. Neural Network
- ( )
  1. KNN (k=1)
  2. Neural Network

3. Decision Tree

- (●) 1. Neural Network
- ( ) 2. KNN (k=1)
- ( ) 3. Decision Tree

7. A decision tree of depth 2 is visualized below. Using the `value` attribute of each leaf, find the accuracy score for the tree of depth 2 and the accuracy score for a tree of depth 1. 1 point



What is the improvement in accuracy between the model of depth 1 and the model of depth 2? (i.e. accuracy2 - accuracy1)

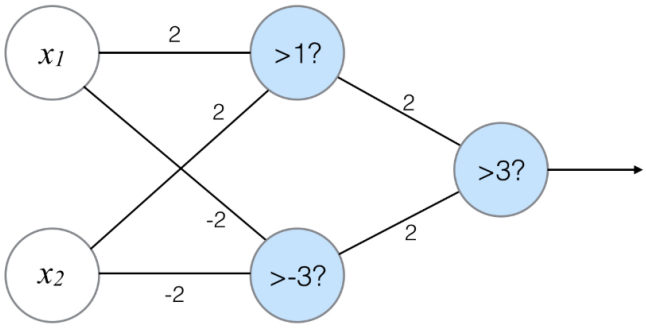
006745

8. Which of the following might be good ways to help prevent a data leakage situation? 1 point

- [ ] Perform a feature importance analysis on a fitted model
- [ ] Ensure that data is preprocessed outside of any cross validation folds.
- [x] Sanity check the model with an unseen validation set
- [x] If time is a factor, remove any data related to the event of interest that doesn't take place prior to the event.
- [x] Remove variables that a model in production wouldn't have access to

9. Given the neural network below, find the correct outputs for the given values of x1 and x2. 1 point

The neurons that are shaded have an activation threshold, e.g. the neuron with *>1?* will be activated and output 1 if the input is greater than 1 and will output 0 otherwise.



( )

x1	x2	output
0	0	0
0	1	0
1	0	0
1	1	1

( )

x1	x2	output
0	0	0
0	1	1
1	0	1
1	1	1

x1	x2	output
0	0	0
0	1	1
1	0	1
1	1	0

x1	x2	output
0	0	1
0	1	0
1	0	0
1	1	1

10. Which of the following are true statements about gradient boosted decision trees? (Select all that apply.) 1 point

- [x] Training gradient boosted decision trees usually requires significant computation and careful parameter tuning.
- [x] Like decision trees, gradient boosted decision trees easily handle a mixture of feature types.
- [ ] Like decision trees, gradient boosted decision tree models are easy to interpret.
- [x] Typically the number of weak estimators (n_estimators) parameter is adjusted first to best exploit computational resources, followed by other key parameters such as the boosting learning rate (learning_rate).
- [x] Gradient boosted decision trees have often achieved among the best 'off the shelf' results on many prediction problems with structured data.