Data Science Quiz Day 4

|  |  |  |  |
| --- | --- | --- | --- |
| Instructor: | Stuart Whipp, Ben Whalley | Name: |  |
|  |  | Date: |  |

Try your best to answer these questions based on topics discussed thus far. Expected to take 15-30 minutes with discussion.

1. What is the Gini index for?
   1. Deciding how decision trees split into different branches based on training data
   2. Selecting different types of neural networks
   3. Selecting number of epochs in neural networks

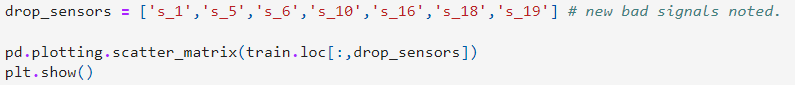
A picture containing diagram

Description automatically generated

1. Supposing that model\_gs is a Random Forest, what does the code above convey?
   1. A plot of the entire random forest
   2. A plot of our best model
   3. A plot of the first tree within a list of trees contained in the ‘forest’
2. What are two examples of ensembling trees?
   1. Bagging and Boosting
   2. Weights and Bias
   3. Convolutions and ReLU



1. What is the purpose of the above code?
   1. To make all columns of the data have a mean of 1 and standard deviation 0
   2. To make all columns of the data have a mean of 0 and standard deviation 1
   3. To make all columns of the data have a mean of 1 and standard deviation 1



1. In the above code sample, we are using ‘loc’ to index locations in our data where column names occur. If we wished to perform indexing by ‘integers’ rather than using string (text) information, what might we need to do?
   1. To investigate the .iloc method
   2. To investigate the .int\_loc method
   3. To make all columns have numeric column names

Text

Description automatically generated

1. What is the purpose of the above code?
   1. To perform GridSearch
   2. To perform gradient descent on a neural network
   3. To evolve a model’s settings or ‘hyper-parameters’

Text

Description automatically generated

Diagram

Description automatically generated with low confidence

Graphical user interface, application, email

Description automatically generated

1. The above diagram is drawn in network library and represents a ‘directed graph network’. Which node has the highest *degree centrality?* 
   1. 3
   2. 5
   3. 6
   4. 4
2. Which node has the highest *betweeness centrality?*
   1. 3
   2. 5
   3. 6
   4. 4
3. What do these two terms mean?
   1. The number of connections, and instances where paths must pass this node on the route to elsewhere
   2. How central a node appears visually on the diagram
4. Which node has the highest *betweeness centrality* out of node 6 and node 4*?*
   1. 4
   2. 6