Congratulations! You passed!

TO PASS 70% or higher

Keep Learning

GRADE
100%

Phase 2. Project 2

LATEST SUBMISSION GRADE
100%

1. The D-DIMER values are highly concentrated <1k, but there are many samples that are several orders of magnitude apart from the rest of the samples. What is the most likely explanation for this? (Hint: look at the data samples, particularly the exam metadata.) **1 / 1 point**

- ○ The data was collected from two cohorts from two different time periods
- ○ There is a large disparity in D-DIMER lab values across patient age
- ◉ The data collected from one the clinics may use different units
- ○ There is a large disparity in D-DIMER lab values across patient gender

✓ **Correct**
Clinic B D-DIMER lab values are 3 orders of magnitude apart, implying a different recordkeeping protocol from the other clinics.

Phase 2. Project 2
Graded Quiz • 30 min

Due Mar 22, 2:59 PM CST

Check all that apply.

- ☑ A tree-based model, such as random forest, can be trained directly on the data with missing values

✓ **Correct**
Tree-based models are able to adapt to settings where there are missing values.

- ☑ A tree-based model, such as random forest, can be trained after the missing values are synthetically generated, using a process known as imputation

✓ **Correct**
Tree-based models are already able to adapt to settings where there are missing values, therefore imputation does not affect its feasibility.

- ☑ A logistic regression model can be trained after the missing values are synthetically generated, using a process known as imputation

Phase 2. Project 2
Graded Quiz • 30 min

Due Mar 22, 2:59 PM CST

Logistic regression models need to be able to see all features in a sample in order to make a prediction, and imputation fills in the missing values.

- ☐ A logistic regression model can be trained directly on the data with missing values

3. Which of the following is FALSE regarding logistic regression models? **1 / 1 point**

- ○ Logistic regression produces values between 0 and 1, regardless of the scale of the features
- ○ Logistic regression is commonly used for classification problems
- ○ Logistic regression uses the sigmoid activation function
- ◉ Logistic regression can take unstructured inputs, such as images or text

✓ **Correct**
Images (2-dimensional number grids) and text (sequences of 1-dimensional vectors) cannot be fed into logistic regression models, because logistic regression models can only receive a single 1-dimensional vector as input.

4. Which of the following is FALSE regarding random forest models? **1 / 1 point**

- ○ Random forest models are highly interpretable
- ○ Random forest models learn multiple decision trees that each learn on a subset of the available features
- ○ Random forest models are a type of decision tree algorithm
- ◉ Random forest models require feature normalization (i.e. scaling the features such that they are between 0 and 1) in order to work effectively

✓ **Correct**