1  
point

1.

Which of the following are components in building a machine learning algorithm?



Statistical inference



Collecting data to answer the question.



Machine learning



Training and test sets



Artificial intelligence

1  
point

2.

Suppose we build a prediction algorithm on a data set and it is 100% accurate on that data set. Why might the algorithm not work well if we collect a new data set?



Our algorithm may be overfitting the training data, predicting both the signal and the noise.



We have too few predictors to get good out of sample accuracy.



We may be using bad variables that don't explain the outcome.v



We have used neural networks which has notoriously bad performance.

1  
point

3.

What are typical sizes for the training and test sets?



90% training set, 10% test set



50% in the training set, 50% in the testing set.



0% training set, 100% test set.



80% training set, 20% test set

1  
point

4.

What are some common error rates for predicting binary variables (i.e. variables with two possible values like yes/no, disease/normal, clicked/didn't click)? Check the correct answer(s).



Sensitivity



Median absolute deviation



Correlation



Root mean squared error



R^2

1  
point

5.

Suppose that we have created a machine learning algorithm that predicts whether a link will be clicked with 99% sensitivity and 99% specificity. The rate the link is clicked is 1/1000 of visits to a website. If we predict the link will be clicked on a specific visit, what is the probability it will actually be clicked?



0.009%



50%



89.9%



9%