1 Week 1

Question 1

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

i Artificial intelligence

ii Data mining

iii Training and test sets

iv Creating features.

Question 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?

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

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

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

iv We may be using a bad algorithm that doesn't predict well on this kind of data.

Question 3

What are typical sizes for the training and test sets?

i 90% training set, 10% test set

ii 60% in the training set, 40% in the testing set.

iii 20% test set, 80% training set.

iv 0% training set, 100% test set.

Question 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)?

i R^2

ii Median absolute deviation

iii Sensitivity

iv P-values

Question 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?

i 89.9%

ii 0.009%

iii 50%

iv 9%