Solution to Quiz 1

Question 1

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

A: Asking the right question.

Question 2

Q: 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?

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

Question 3

Q: What are typical sizes for the training and test sets?

A: 80% training set, 20% test set. (or 50% in the training set, 50% in the testing set?)

Question 4

Q: 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).

A: Sensitivity.

Question 5

Q: 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?

A: The situation can be describes as:

Therefore, $\Pr(TP|+) = \frac{TP}{TP+FP} = \frac{99}{99+999} = 9.016\%$.