Week 1 Quiz

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

* **Asking the right question.**
* Statistical inference
* Machine learning
* Training and test sets
* Artificial intelligence

1. 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?

* We have used neural networks which has notoriously bad performance.
* We may be using bad variables that don’t explain the outcome.v
* **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.

1. What are typical sizes for the training and test sets?

* 20% training set, 80% test set.
* 90% training set, 10% test set
* 100% training set, 0% test set.
* **60% in the training set, 40% in the testing set.**

1. 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).

* Median absolute deviation
* Root mean squared error
* Correlation
* **Accuracy**
* R^2

1. 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?

P(Click|Visits) = (P(Visits|Click) \* P(Click)) / (P(Visits|Click) \* P(Click) + P(Visits|NoClicks) \* P(NoVisits))

(0.99 \* (1/1000)) / (0.99 \* (1/1000) + (1-0.99) \* (1-1/1000))

## [1] 0.09016393

* 50%
* 99.9%
* 99%
* **9%**