ENGN 2520 / CSCI 1950-F Homework 1
Due Friday February 8 by 4pm

Problem 1

Suppose we have three boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. Consider the following process: a box is chosen at random with probabilities $p(r) = 0.1$, $p(b) = 0.3$, $p(g) = 0.6$, and a piece of fruit is selected uniformily at random from the box. What is the probability that this process will select an apple? If we selected an orange, what is the probability that it came from the green box?

Problem 2

Suppose we have two dice, one is fair and the other always comes up 6. I will toss a biased coin to decide which die to roll. The coin comes up heads with probability p. If the coin turns up heads I will toss the fair die and if the coin turns up tails I will toss the other die.

Let x be the value of the die roll obtained under this process.

(a) What is the expectation of x?

(b) What is the variance of x?

Problem 3

Show that if two random variables x and y are independent, then their covariance is zero.

Problem 4

In this problem you will implement the linear basis method for regression with least squares error. The data for this problem is available on the course website. You will learn functions using the training data and evaluate their generalization on the test data.

The picture below was generated by loading the training data into matlab and plotting the sample points using the command scatter as follows:

> load Xtrain;
> load Ytrain;
> scatter(Xtrain, Ytrain);
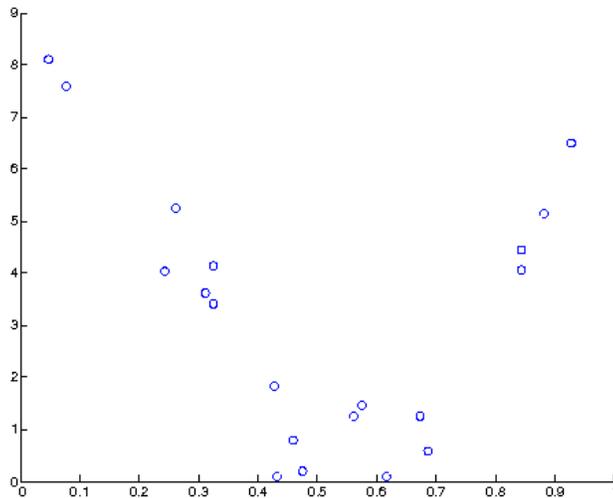
You will use a polynomial basis function:

$$\phi_1(x) = 1, \ \phi_2(x) = x, \ \phi_3(x) = x^2, \ \phi_4(x) = x^3, \ ...$$

(a) Report the training and test error (in terms of sums of squared differences) obtained using polynomials of degrees 1 through 10. Note that a polynomial of degree d is represented using $d + 1$ basis functions.

(b) Submit a plot showing the training data and the degree 3 polynomial you learned.

(c) Submit a plot showing the training data and the degree 10 polynomial you learned.

(d) Submit your Matlab source code along with your writeup and code.