10-601 Machine Learning: Assignment 9

- The assignment is due at 3:00pm (beginning of class) on **Monday, April 21, 2008**.

- The assignment is worth **50 points**.

- Write your name at the top right-hand corner of each page submitted.

- Each student must hand in a hard-copy writeup. See course page for collaboration policies.
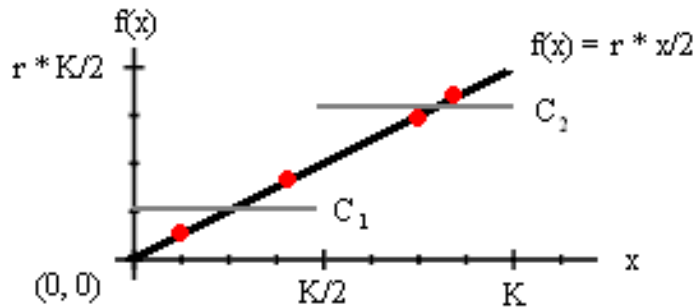
Q1: Bias versus variance [25 pts]

You are trying to estimate the function y = f(x). Suppose you sample m points, $x_i...x_m$, uniformly from the interval [0, K], and for each of those points, x_i, you are given its respective $y_i = f(x_i)$ from an oracle. You decide to use a piece-wise constant regression function, $\hat{f}(x)$, to estimate f(x). Your learning algorithm is simple: you partition your observations, $\langle X, Y \rangle$, into v equal intervals along the x-axis. Then, given a new observation, x^*, you see which of the v intervals x^* falls into. You then predict $y^* = \hat{f}(x^*)$ to be the mean of the y's for the x's you previously observed in that same interval. More precisely:

$$\hat{f}(x^*) = \begin{cases} c_1 = mean\,[Y|X \leq K/v] & \text{if } x^* \leq K/v \\ ... \\ c_j = mean\,[Y|(j-1)K/v < X \leq j*K/v] & \text{if } (j-1)K/v < x^* \leq j*K/v \\ ... \\ c_v = mean\,[Y|(v-1)K/v < X \leq K] & \text{if } (v-1)K/v < x^* \leq K \end{cases}$$

where $mean\,[Y|a < X \leq b]$ means 'take the mean of the y values for all the x's that fall between a and b.'

The picture below shows an example of such a task. The red dots represent your samples, the thick black line is f(x), while the two grey lines represent your piece-wise constant estimates for the two intervals. In the example shown, m = 4 and v = 2.
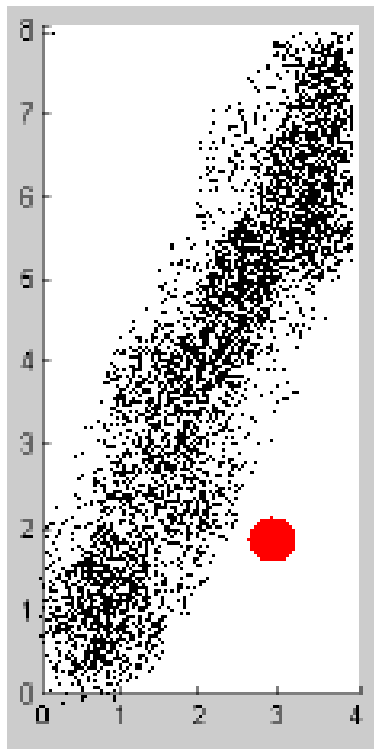

Given the example set up as above, please answer the following questions:

1. Suppose you only use a single piece to approximate f(x) (ie, v = 1). What is the bias of your estimate $\hat{f}(x)$? (answer in terms of X, K, and m)

2. Its variance?

3. What happens to the bias and variance if you increase v to 2? (Do they go up, down, neither, both?)

4. What if you double m (keeping v constant)?

HINT: You may find the equations for calculating the mean and variance of a sample useful.

Q2: Principal components analysis (PCA) [15 pts]

1. Draw, and clearly label, the first principal component of the small black points shown below (the points were drawn by hand and so may not be exactly correct, but trust the general idea: you should not need to use a computer).

2. Reconstruct the large red point at (x=3, y=2) both graphically, using a vector diagram, and algebraically, showing the form, along with the result, of the reconstruction:

   (a) using only the first principal component
   (b) using the first and second principal components

Q3: Neural networks [10 pts]

We follow-up on an idea presented in class. Slide 15 of the artificial neural networks lecture shows how an 8-bit pattern can be learned by a network with only 3 units in its single hidden layer. We assume here that the only inputs the network will be given are the 8 inputs shown in the slide.

1. Can you encode a similar 8-bit pattern using only a single hidden unit? If so, please give the weights for such a hidden unit. If not, explain why not.

2. Could the output layer decode from the hidden unit's representation to the 8-bit output pattern? Again, if so, please give the weights, and if not, explain why not.