Name: *Student ID:*

1/25	
2/25	
3/25	
4/25	
T/100	

STA 414/2104 — First Test — 2006-02-28

No books, notes, or calculators are allowed.
The four questions are worth equal amounts.

Question 1: Consider a classification problem in which there is one real-valued input, X, and a binary (0/1) response variable, Y. There are ten training cases, for which the values of X and Y are as follows (for your convenience, the training cases have been ordered by the value of X):

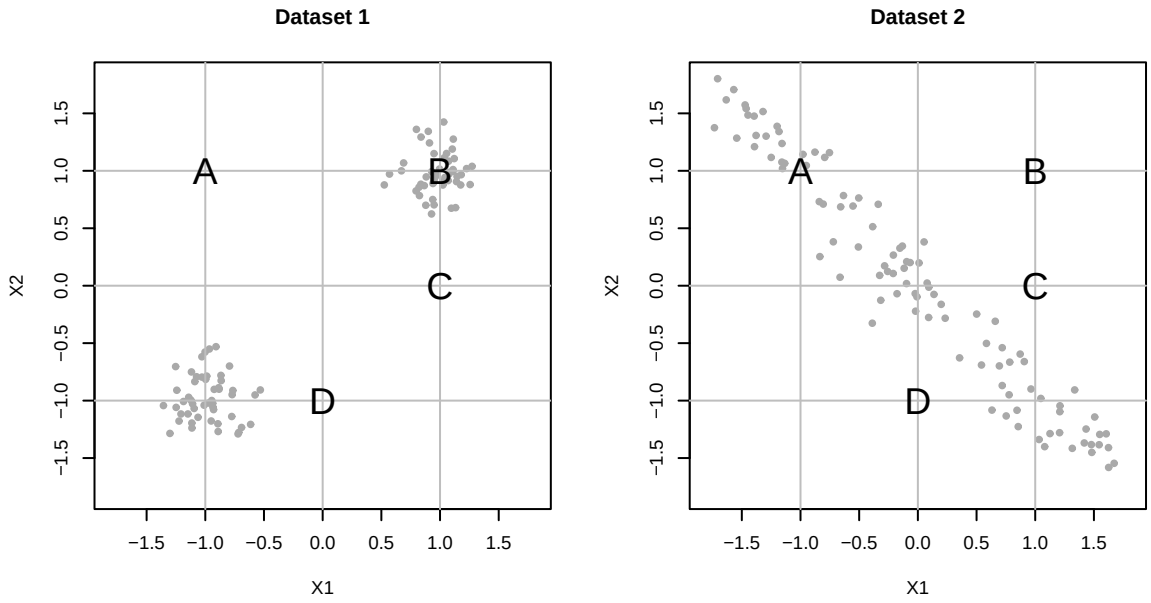
X	0.2	0.7	1.1	1.2	1.4	1.9	2.2	2.9	3.1	3.8
Y	1	1	0	1	1	0	0	0	1	0

a) Estimate the error rate of the one-nearest-neighbor (1-NN) classifier for this problem using leave-one-out cross validation. (Ie, cross validation in which each training case is predicted using all the other training cases.) Also, show for each training case whether it results in an error when doing cross validation.

b) Estimate the error rate of a three-nearest-neighbor (3-NN) classifier for this problem using leave-one-out cross validation, and show for each training case whether it results in an error when doing cross validation.

c) On the basis of the results in parts (a) and (b) above, if you had to use either 1-NN or 3-NN to make predictions for a test set, which would be the better choice?

Question 2: Below are scatterplots of the input variables, X_1 and X_2, in two datasets with 100 cases. For both datasets, the variables have been standardized to have mean zero and standard deviation one. (Note that the response variable is not shown here.)



Rather than use these two inputs for a classification or regression method, we would like to reduce them to only one variable by projecting the points onto the first principle component direction. If $x = (x_1, x_2)$ is a vector representing the inputs for some case, and v is a vector of length one pointing in the direction of the first principle component, the projection on the first principle component will be $x^T v$. Note that $-v$ would also be a valid indicator of the principle component direction; you may use either, but you must be consistent. Draw the vectors you use on the plots above.

The questions below ask you to find (approximately) the projections on the first principle component of the four points marked A, B, C, and D. You should determine the principle component directions by eye, not by trying to work numerically through some formula. (Note that you should figure out the the answers for dataset 1 and for dataset 2 separately — these datasets have nothing to do with each other.)

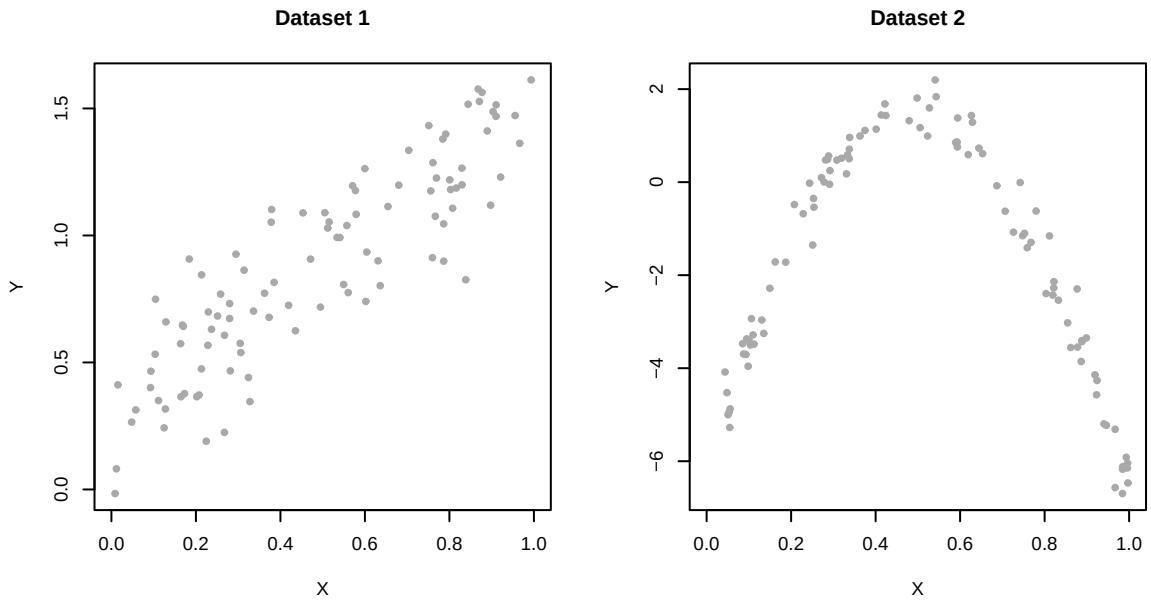
a) For **dataset 1**, and for each point marked in the plot, circle the number below that is approximately the projection of that point on the first principle component direction:

Point A:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0
Point B:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0
Point C:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0
Point D:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0

b) For **dataset 2**, and for each point marked in the plot, circle the number below that is approximately the projection of that point on the first principle component direction:

Point A:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0
Point B:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0
Point C:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0
Point D:	-2.0	-1.4	-1.2	-1.0	-0.7	-0.3	0.0	0.3	0.7	1.0	1.2	1.4	2.0

Question 3: Here are scatterplots of two datasets. For each dataset, there is only one input variable, X, which plotted on the horizontal axis, and a real-valued response variable, Y, which is plotted on the vertical axis.



The questions below ask, for each dataset, about the bias and variance of each of three regression methods — least squares linear regression (LR), one-nearest neighbor (1-NN), and twenty-nearest-neighbor (20-NN). In answering these questions, you should assume that the input variable for a test point in which we are interested is chosen uniformly from $(0, 1)$.

For **dataset 1**, circle the phrase that best describes the **bias** of each method:

LR:	Close to zero	Small	Moderate	Large
1-NN:	Close to zero	Small	Moderate	Large
20-NN:	Close to zero	Small	Moderate	Large

For **dataset 1**, circle the phrase that best describes the **variance** of each method:

LR:	Close to zero	Small	Moderate	Large
1-NN:	Close to zero	Small	Moderate	Large
20-NN:	Close to zero	Small	Moderate	Large

For **dataset 2**, circle the phrase that best describes the **bias** of each method:

LR:	Close to zero	Small	Moderate	Large
1-NN:	Close to zero	Small	Moderate	Large
20-NN:	Close to zero	Small	Moderate	Large

For **dataset 2**, circle the phrase that best describes the **variance** of each method:

LR:	Close to zero	Small	Moderate	Large
1-NN:	Close to zero	Small	Moderate	Large
20-NN:	Close to zero	Small	Moderate	Large

Question 4: Consider a regression problem with three inputs, X_1, X_2, and X_3, and a real-valued response variable, Y. We know the values of these variables in ten training cases, as follows:

X_1	X_2	X_3	Y
2.2	1.9	0.0	2.1
0.0	0.0	1.1	5.4
1.3	2.2	0.0	1.9
3.5	0.7	0.0	3.0
0.0	0.0	1.1	3.4
8.0	2.8	0.0	1.8
7.1	3.3	0.0	3.3
2.2	1.9	0.0	2.2
3.4	4.0	0.0	2.1
4.4	5.2	0.0	2.7

a) Suppose we use a linear regression model without an intercept, of the form

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

If we estimate β_1, β_2, and β_3 by least squares, what will be the estimate for β_3? (Note: Elaborate calculations are not necessary.)

b) If we use the least squares estimates for the model without intercept in part (a), what will be the predicted value of Y for a test case in which $X_1 = 0$, $X_2 = 0$, and $X_3 = 1.1$?

c) If we use the least squares estimates for the model without intercept in part (a), what will be the predicted value of Y for a test case in which $X_1 = 0$, $X_2 = 0$, and $X_3 = 2$?

d) Suppose we use a linear regression model with an intercept, of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

If we estimate β_0, β_1, β_2, and β_3 by least squares, will the estimate for β_3 be the same as for the model without intercept (as in part (a))?

d) If we use the least squares estimates for the model with intercept in part (d), will the predicted value of Y for a test case in which $X_1 = 0$, $X_2 = 0$, and $X_3 = 1.1$ be the same as for the model without an intercept (as in part (b))?