CS 5565,465R,ECE5590CI, ECE 490CI HW1(Introduction) 130 pts.

Name _____

1. (6 points) You are given the three vectors below,

$$\mathbf{a} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 5 \\ -4 \end{bmatrix} \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ -2 \\ 1 \end{bmatrix} \mathbf{c} = \begin{bmatrix} 0 \\ -3 \\ 5 \\ -1 \\ 2 \end{bmatrix}$$

   (a) (1 point) What is $\mathbf{a} + \mathbf{b}$

   (b) (1 point) What is $\mathbf{b} + \mathbf{c}$

   (c) (1 point) What is $3\mathbf{a} + 4\mathbf{b} - 2\mathbf{c}$

   (d) (1 point) Show the dot product of $\mathbf{a}$ and $\mathbf{b}$

   (e) (1 point) Show the dot product of $\mathbf{a}$ and $\mathbf{c}$

   (f) (1 point) Show the dot product of $\mathbf{b}$ and $\mathbf{c}$

2. (12 points) You are given the three vectors below,

$$\mathbf{a} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \mathbf{c} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

   (a) (2 points) Find the angle between $\mathbf{a}$ and $\mathbf{b}$

   (b) (2 points) Find the angle between $\mathbf{a}$ and $\mathbf{c}$

   (c) (2 points) Find the angle between $\mathbf{b}$ and $\mathbf{c}$

   (d) (2 points) Find the plane that goes through point [1,1,1] that is perpendicular to $\mathbf{a}$

   (e) (2 points) Find the plane that goes through point [1,1,1] that is perpendicular to $\mathbf{b}$

   (f) (2 points) Find the plane that goes through point [1,1,1] that is perpendicular to $\mathbf{c}$

3. (12 points) You are given the two matrices below,

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} \mathbf{B} = \begin{bmatrix} 1 & 3 & 0 \\ 0 & -2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

   (a) (4 points) For each matrix find the Characteristic Polynomial.

   (b) (4 points) For each matrix find the Eigenvalues.

   (c) (4 points) For each matrix find the Eigenvectors.

4. (15 points) Explain whether each scenario is a classification or regression problem, and indicate whether we are most interested in inference or prediction. Finally, provide n and p.

(a) (5 points) We collect a set of data on 10,000 home sales in a particular city. For each house sold, we record the price, number of floors (ranch, 1 1/2, reverse 1 1/2, 2 story), age, total number of rooms, number of bedrooms, number of bathrooms, basement or no basement, lot size, and school district. We are interested in understanding which factors affect the house sales price.

(b) (5 points) We are interested in predicting the % change in the USD/Euro exchange rate in relation to the weekly changes in the world stock markets. Hence we collect weekly data for all of 2016. For each week we record the % change in the USD/Euro, the % change in the US market, the % change in the British market, and the % change in the German market.

(c) (5 points) We are considering an analysis on group 100 employees in a petrochemical plant to determine whether they have either a high or low risk of being diagnosed with a certain disease. We collect data on 30 different chemical compounds that they may be exposed to and record the amount of time they are exposed to each chemical each day. For each employee, we want to determine if they have high risk or a low risk of getting the disease.

5. (10 points) Consider the bias-variance decomposition.

   (a) (5 points) Provide a sketch of typical (squared) bias, variance, training error, test error, and Bayes (or irreducible) error curves, on a single plot, as we go from less flexible statistical learning methods towards more flexible approaches. The x-axis should represent the amount of flexibility in the method, and the y-axis should represent the values for each curve. There should be five curves. Make sure to label each one.

   (b) (5 points) explain why each of the five curves has the shape displayed in part (a).

6. (20 points) For each of parts (a) through (d), indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.

   (a) (5 points)
   The sample size n is extremely large, and the number of predictors p is small.

   (b) (5 points)
   The number of predictors p is extremely large, and the number of observations n is small.

   (c) (5 points)
   The sample size n is extremely large, and the number of predictors p is small.

   (d) (5 points)
   The relationship between the predictors and response is highly non-linear.

   (e) (5 points)
   The variance of the error terms, i.e. $\sigma^2 = \text{Var}(\epsilon)$, is extremely high.

7. (15 points) Suppose you are given the following data.

X	Y
1.00	0.8
2.00	1.4
3.00	3.8
4.00	3.7
5.00	6.2
6.00	7.0
7.00	9.2
8.00	9.3
9.00	10.1
10.00	12.2

(a) (5 points) Use linear regression to find the coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$

(b) (5 points) What is the predicted value of Y_{15}?

(c) (5 points) What is the predicted value of Y_{18}?

You may use the expression in the book or the following:

$$\begin{aligned} \beta_1 &= \frac{n\sum_{t=1}^{n} X_t Y_t - \left(\sum_{t=1}^{n} X_t\right)\left(\sum_{t=1}^{n} Y_t\right)}{n\sum_{t=1}^{n} X_t^2 - \left(\sum_{t=1}^{n} X_t\right)^2} \\ \beta_0 &= \bar{Y} - \beta_1 \bar{X}, \end{aligned}$$

where $\bar{Y} = \left(\sum_{t=1}^{n} Y_t\right)/n$ and $\bar{X} = \left(\sum_{t=1}^{n} X_t\right)/n$

8. (10 points) In your own words, explain the advantages and disadvantages of a very flexible (versus a less flexible) approach for regression or classification. Under what circumstances might a more flexible approach be preferred to a less flexible approach? When might a less flexible approach be preferred?

9. (10 points) Describe the differences between a parametric and a non-parametric statistical learning approach. What are the advantages of a parametric approach to regression or classification (as opposed to a nonparametric approach)? What are its disadvantages?

10. (20 points) The table below provides a training data set containing ten observations, three predictors, and one qualitative response variable.

Obs.	X_1	X_2	X_3	Y
1	0	3	0	Red
2	2	0	0	Red
3	3	1	1	Red
4	0	1	2	Green
5	-1	0	1	Green
6	1	1	1	Red
7	0	1	2	Green
8	-2	0	2	Green
9	1	2	2	Red
10	1	2	1	Red

Suppose we wish to use this data set to make a prediction for Y when $X_1 = X_2 = X_3 = 0$ using K-nearest neighbors.

(a) (4 points) Compute the Euclidean distance between each observation and the test point, $X_1 = X_2 = X_3 = 0$.

(b) (4 points) What is our prediction with $K = 1$? Why?

(c) (4 points) What is our prediction with $K = 3$? Why?

(d) (4 points) Answer the previous two questions using Manhattan distance as the metric (sum of distances in each orthogonal direction)? How does the use of a different measure affect the outcome?

(e) (4 points) If the Bayes decision boundary in this problem is highly nonlinear, then would we expect the *best* value for K to be large or small? Why?