**Department of Econometrics and Statistics**

**Pakistan Institute of Development Economics, Islamabad**

**Topics in Advanced Econometrics Zahid Asghar**

Note: Attempt all the questions. Question No.1 has 15 points.

Question #1

1. While doing a homework assignment, you fit a Linear Model to your data set. You are thinking about changing the Linear Model to a Quadratic one. Which of the following is most likely true:
2. Using the Quadratic Model will decrease your Irreducible Error.
3. Using the Quadratic Model will decrease the Bias of your model.
4. Using the Quadratic Model will decrease the Variance of your model
5. Using the Quadratic Model will decrease your Reducible Error
6. A hypercube with side length 1 in d dimensions is defined to be the set of points (x1, x2, ..., xd) such that0≤*xj*≤1 for all j = 1, 2, ..., d. Define the boundary region of the hypercube to be the set of all points such that there exists a j for which 0≤*xj*≤.05 or .95≤*xj*≤1 (i.e., it is the set of all points that have at least one dimension in the most extreme 10% of possible values). What proportion of the volume of a hypercube of dimension 50 is in the boundary region? Please give your answer as a value between 0 and 1 with 3 significant digits. If you think the answer is 50.52%, you should say 0.505: (2 points)
7. True or False: A fitted model with more predictors will necessarily have a lower Training Set Error than a model with fewer predictors.
8. Why is linear regression important to understand? Select all that apply:
9. Top of Form

a) The linear model is often true

1. Linear regression is very extensible and can be used to capture nonlinear effects
2. Simple methods can outperform more complex ones if the data are noisy
3. Understanding simpler methods sheds light on more complex ones
4. In the expression Sales ≈ f(TV, Radio, Newspaper), "Sales" is the:
5. Response
6. Training Data
7. Independent Variable Feature
8. Which of the following are true statements? Select all that apply:
9. Top of Form
   1. A 95% confidence interval is a random interval that contains the true parameter 95% of the time
   2. If I perform a linear regression and get confidence interval from 0.4 to 0.5, then there is a 95% probability that the true parameter is between 0.4 and 0.5.
   3. The true parameter (unknown to me) is 0.5. If I sample data and construct a 95% confidence interval, the interval will contain 0.5 95% of the time.
10. Which of the following indicates a fairly strong relationship between X and Y?
11. R2=0.9
12. The p-value for the null hypothesis β1=0 is 0.0001
13. The t-statistic for the null hypothesis β1=0 is 30
14. What is the difference between lm(y ~ x\*z) and lm(y ~ I(x\*z)), when x and z are both numeric variables?
15. The first one includes an interaction term between x and z, whereas the second uses the product of x and z as a predictor in the model.
16. The second one includes an interaction term between x and z, whereas the first uses the product of x and z as a predictor in the model.
17. The first includes only an interaction term for x and z, while the second includes both interaction effects and main effects.
18. The second includes only an interaction term for x and z, while the first includes both interaction effects and main effects.

9. Suppose we collect data for a group of students in a statistics class with variables X1= hours studied, X2= undergrad GPA, and Y= receive an A. We fit a logistic regression and produce estimated coefficients β^o=−6, β^1=0.05,β^2=1.Estimate the probability that a student who studies for 40h and has an undergrad GPA of 3.5 gets an A in the class:

10. Which of the following is NOT a linear function in x:

1. f(x)=a+b2x
2. The discriminant function from LDA
3. δk(x)=xμkσ2−μ2k2σ2+log(πk)
4. logit(P(y=1|x)) where P(y=1|x) is as in logistic regression
5. P(y=1|x) from logistic regression

Q#2

1. This problem relates to the QDA model, in which the observations within each class are drawn from a normal distribution with a classspecific mean vector and a class specific covariance matrix. We consider the simple case where *p* = 1; i.e. there is only one feature.

Suppose that we have *K* classes, and that if an observation belongs to the *k*th class then *X* comes from a one-dimensional normal distribution, *X ∼ N*(*μk, σ*2*k*). Recall that the density function for the one-dimensional normal distribution is given in (4.11). Prove that in this case, the Bayes’ classifier is *not* linear. Argue that it is in fact quadratic.

1. Carefully explain the differences between the KNN classifier and KNN regression methods.

Q#3

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

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

(c) The relationship between the predictors and response is highly non-linear.

(d) The variance of the error terms, i.e. σ2 = Var(), is extremely high.

**Best of Luck**