INFO 7390 – ADS
Practice Exam One

Student Name: _____
Professor: Nik Bear Brown

Rules:
1. NO COMPUTER, NO PHONE, NO DISCUSSION or SHARING.
2. Ask if you don't understand a question.
3. You may use one 8½"×11" sheets of notes (you may use both sides, written or printed as small as you like).
4. Time allowed. Until end of class.
5. Bring pen/pencil. The exam will be written on paper.

Q1 (5 Points)

What is a low bias model? What are its advantages/disadvantages? How do we counter act its disadvantages?

Q2 (5 Points) Assume regression is being used to predict whether a student will get drunk or not. The dependent variable is **drunk,** which indicates drunk or not. Assume the only independent variable is **beers,** which indicates the number of beers consumed. The stats for the fit are shown in the table below.

drunk	Coef.	Std. Err.
beers	0.7	0.035
intercept	-2.5	0.1

A. Write an equation that describes the model.
B. Is the coefficient *beers* significant? How does one interpret the meaning of its value?
C. Is the coefficient *intercept* significant? How does one interpret the meaning of its value?
D. What is the probability of getting drunk after 2 beers?
E. What is the likelihood of getting drunk after no beers?

Q3 (5 Points) Write pseudocode for the bootstrapping algorithm.

Q4 (5 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?

Q5 (5 Points) What is the difference between Ridge and Lasso regression? Why use Ridge or Lasso regression? Why would I use one or the other? Are there hyperparameters in use Ridge or Lasso regression, and if so, how does one determine their value?

Q6 (5 Points) How does one adjust the support in a Support Vector Machine? How does one adjust the bias in a Support Vector Machine other than changing the kernel? Are their hyperparameters that adjust the support and bias? If so, how does one determine their values?

Q7 (5 Points) Write the Ridge regression equation. Are there coefficients not effected by the tuning parameter?

Q8 (5 Points) What is the equation for multiple logistic regression with two independent variables?

Q9 (5 Points) Describe an interaction variable in regression. How does one create them? How does one know whether their effect is significant?

Q10 (5 Points) Create an algorithm for aggregation of base models in bagging that uses another machine learning model rather than numerical aggregation.

Q11 (5 Points) Write pseudocode for the K-Means Clustering algorithm.

Q12 (5 Points) Assume one wants to use the three-class categorical variable high/medium/low as an independent variable in linear regression. How can we encode it?

Write an equation that describes the model. Assume we are fitting the intercept and one other continuous independent variable, and our dependent variable is called 'y'.

Q13 (5 Points) What is a standardized regression coefficient? How do we calculate them?

Q14 (5 Points) Describe k-fold cross validation in pseudocode.

Q15 (5 Points) Describe the difference between bagging, boosting and stacking? Which are ensemble methods?

Q16 (5 Points) Is k-means a parametric or a non-parametric statistical learning approach. Why or why not?

Q17 (5 Points) Write the equation for multiple linear regression with three continuous independent variables and one categorical independent variable with two classes. Are there any assumptions of the error model?

Q18 (5 Points) What is the difference between a t-test and Z-test? How are z-scores calculated?

Q20 (5 Points) You are asked to build a classification model about meteorites impact with Earth (important project for human civilization). After preliminary analysis, you get 97% accuracy. Is this a good result? Why or why not?