CSP 571 Data Preparation and Analysis

Quiz - 1

Question 1

Prediction of a categorical or qualitative output for a supervised learning problem is also referred to as:

- ○ None of the Above
- ◉ Classification
- ○ Regression
- ○ Latent Variable Model

Question 2

Distance-based non-parametric methods for regression, such as kNN, suffer from the curse of dimensionality as distances to local points increase as the number of feature dimensions:

- ○ Decrease
- ○ Go to 0
- ○ None of the Above
- ◉ Increase

Question 3

An example of a non-parametric method/model for estimating $\hat{f}$ would be:

- ○ Simple Linear Regression
- ◉ k-Nearest-Neighbors
- ○ Polynomial Regression
- ○ None of the Above

Question 4

In order to obtain a Least Squares estimate of a linear model, the RSS quantity is:

- ○ Maximized
- ○ Set to 0
- ◉ Minimized
- ○ Set to ∞

Question 5

When computing the t-statistic for an estimated coefficient in a linear regression model, we are checking the number of standard deviations the specific β_i is away from:

- ○ 1
- ○ ∞
- ○ -1
- ◉ 0

Question 6

A model whose estimated parameters change significantly when a single training observation is changed is said to have high:

- ○ Linearity
- ◉ Variance
- ○ MSE
- ○ Bias

Question 7

In multiple linear regression, the F-statistic will be close to this value when there is no relation between the response and any of the predictors:

- ○ None of the Above
- ◉ 1
- ○ π
- ○ ∞

Question 8

Including the square of a predictor within a linear model for polynomial regression still results in a:

- ◉ Linear Model
- ○ Non-Parametric Model
- ○ None of the Above
- ○ Both of the Above

Question 9

10

A small p-value for the association between a predictor and response within a linear regression model allows us to reject the following Hypothesis regarding the possible relationship between predictor and response:

- ○ Alternative
- ○ None of the Above
- ◉ Null
- ○ Supervised

Question 10

When including a qualitative predictor within a regression model, if the predictor can be represented as a factor with two levels, then the following type of variable can be included within a linear model:

- ◉ Dummy Variable
- ○ None of the Above
- ○ Random Variable
- ○ Exogenous Variable

Question 11

Any structure or pattern in the following plot of regression results may indicate a problem with linear model results:

- ○ Time Series Plot
- ○ Feature Plot
- ◉ Residual Plot
- ○ None of the Above

Question 12

When two or more predictors are closely related to each other, with high covariance/correlation, it is referred to as:

- ○ None of the Above
- ○ Outliers
- ◉ Collinearity
- ○ Leverage

Question 13

When diagnosing a regression model, the ratio of variance of the response explained by the model and the total variance of the response is calculated to be:

- ○ RSS
- ○ TSS
- ○ RSE
- ◉ R^2

Question 14

The accuracy of $\widehat{Y}$ as a prediction for Y consists of Reducible Error and Irreducible Error - the latter Irreducible Error stems from the variance of which term that Y is a function of?

- ○ β
- ◉ ε
- ○ $f(X)$
- ○ $\sigma(X)$

Question 15

As a model's flexibility increases (degrees of freedom for smoothing splines/curves), the Training MSE will :

- ◉ Decrease
- ○ None of the Above
- ○ Both of the Above
- ○ Increase

Question 16

Within a model, attempting to determine the exact form of $\widehat{f}$ instead of treating it as a black box is know as:

- ○ Prediction
- ◉ Inference
- ○ Expected Value
- ○ Clustering

Question 17

10 points

The β_0 and β_1 coefficient estimates from a given sample in simple linear regression are neither systematically over-estimated or under-estimated with respect to the population parameters by the least squares estimate, and are thus:

- ○ None of the Above
- ◉ Unbiased
- ○ Equal to the population coefficients
- ○ Biased

Question 18

For a given test observation x_o, along with the error term ε we can decompose the expected Test MSE into the following components:

- None of the Above
- $\left[Bias(\widehat{f(x_0)})\right]^2$
- $Var(\widehat{f(x_0)})$
- Both of the Above