CSCI 467 Problem Set 4
Teresa Liu
March 13, 2020

True/False Questions

1. For linear regression, when including interactions in our model, we can omit the main effects if the p-values associated with those effects are not statistically significant
    a. ***False****: according to the Hierarchical Principle, you must include the main effect associated with interactions even if p-values are not significant*

2. KNN Regression is an example of a parametric model.
    a. ***False****: KNN Regression is a non-parametric model.*

3. Discriminant Analysis is preferred for data sets that are well separated, since the parameter estimates for logistic regression are unstable.
    a. ***True****: logistic regression does not perform well with well separated data sets.*

4. Naïve Bayes is robust to isolated noise samples, irrelevant attributes, and redundant attributes.
    a. ***False****: Naïve Bayes is not robust to redundant attributes since this model assumes that features are independent in each class, and redundancy violates this assumption.*

5. Leave One Out Cross Validation (LOOCV) produces less bias and doesn't overestimate errors. However, this method is very computationally intensive, which is why K-Fold Cross Validation is more popular.
    a. ***True****: Since LOOCV uses every data point individually as validation sets, it produces the lowest bias of the resampling methods, but this is too computationally difficult to perform, so K-Fold provides a compromise between bias and computational challenge.*

6. All algorithms perform equally when averaged over all problems
    a. ***True****: This is the fundamental No Free Lunch Theorem*

7. KNN Regression is known as a Lazy Learning algorithm as opposed to an Eager Learning algorithm.
    a. ***True****: KNN is known as lazy learning, since generalization beyond the training data is delayed until a query is made to the system, as opposed to in Eager Learning, where the system tries to generalize the training data before receiving queries.*

8. Dummy encoding and one-hot encoding are a commonly used method for converting a categorical input variable into continuous variables. With k-groups for a predictor, dummy encoding requires k coded variables and one-hot encoding requires k-1 coded variables.

a. ***False****: The second statement should be flipped. With k-groups, dummy encoding requires k-1 coded variables and one-hot encoding requires k coded variables.*

Conceptual/Fundamental Midterm Exam Questions

1. Which of the following examples are considered unsupervised learning?
    a. Taking a dataset of people's movie-watching history and transforming that data into a recommendations feature based on overlapping movie histories.
    b. Anomaly detection when researching a medical data set for cancer cells, but we don't know what the anomaly looks like.
    c. Using stock market data from the past 20 years to predict how the market will perform in the next 5 years.
    d. Taking a large set of images of people and grouping faces together so that users can search photos by people's faces.
    e. Training a factory robot to pick up and drop off boxes based on a set of rewards.
    f. ***Answer****: A B D*

2. The following are indications that you should use discriminant analysis over logistic regression for a dataset.
    a. When classes are well-separated.
    b. If n is small and the distribution of the predictors X is approximately normal in each of the classes.
    c. When we have more than 2 response classes and we wish to categorize data.
    d. When you have many outliers in your data.
    e. ***Answer****: all of the above*

    Observe the following multiple linear regression model

    $$\hat{y} = \hat{\beta_0} + \hat{\beta_1} x_1 + \hat{\beta_2} x_2 + \hat{\beta_3} x_3$$

    - $\hat{y}$: the predicted number of sales in 1000's of dollars
    - x_1: the number of ads put up on billboards
    - x_2: the number of TV commercials that run on a channel
    - x_3: a dummy variable with the ad containing a special deal
    - the value of coefficients: $\hat{\beta_0} = 1.1$, $\hat{\beta_1} = 5.2$, $\hat{\beta_2} = 1.3$, $\hat{\beta_3} = -0.2$

3. Which of the following statements are correct interpretations of $\hat{\beta_1}$?
    a. If the value of the number of ads put up on billboards increases by one unit, then the number of sales is predicted to increase by $5,200 assuming that all other predictors remain fixed.
    b. If the value of the number of ads put up on billboards decreases by 5.2, then the number of sales is predicted to decrease by $1,000 assuming that all other predictors remain fixed.
    c. If the value of the number of ads put up on billboards increases by one unit, then the number of sales is predicted to increase by $5.20.

d. If the value of the number of ads put up on billboards decreases by one unit, then the number of sales is predicted to decrease by $5,200 assuming that all other predictors remain fixed.

e. ***Answer****: A D*

4. Which of the following are valid methods of selecting important variables in linear regression?
    a. Sideways Selection, in which you start with a model with p/2 variables, and add variable or remove a variable when the p-value associated with that variable is outside of an accepted range, refit the model, and continue until a stopping rule is reached.
    b. Forward Selection, in which you begin with a null model, fit p simple linear regressions and add to the model the variable that results in the lowest RSS, and then repeat until a stopping rule is reached.
    c. Backward Selection, in which you begin with all variables, remove the variable with the largest p-value, fit the new p-1 model, and continue until a stopping rule is reached.
    d. LOOCV, in which you leave a variable out of the regression model and repeat N times to fit the model until you figure out which variable caused the most error, and then remove that variable, and repeat until stopping rule is reached.
    e. ***Answer****: B C*