Final Exam

**Part 1: Concept Review**

1. True. When we perform multiple tests, we are more likely to encounter a Type 1 error (false positive) by chance.
2. False. In bagging, the construction of each tree is independent of the other trees in the ensemble. Bagging creates multiple subsets of the dataset by randomly sampling with replacement and then constructing separate decision trees for each subset. The trees are grown independently of each other.
3. False. Boosting creates an ensemble of trees where each tree contributes to the final prediction.
4. True. Overfitting happens when a model is too complex and fits the training data too closely, including the noise, which makes it sensitive to small changes in the training data, leading to poor generalization to new data.
5. False. K-means is a greedy algorithm and it’s possible for the objective function to increase at some steps during its iterations. This can happen temporarily when data points are being reassigned in the assignment step, but the overall objective of K-means is to minimize the objective function over subsequent iterations.
6. False. When SVD is performed on a matrix with dimensions n\*p where n < p, the number of non-zero singular values is at most n, not p.
7. i) Naïve Bayes. Not significantly affected by the curse of dimensionality. Naïve Bayes assumes that there is independence between features given the class, helping mitigate the effects of having large number of predictors.

ii) Logistic Regression. Can be affected by the curse of dimensionality. As the number of predictors increases, especially if there are irrelevant or high multicollinearity among predictors, it may lead to unstable estimates and break down.

iii) QDA. Is affected by the curse of dimensionality. Estimating the covariance matrix accurately can be more challenging, especially when the number of predictors is close to or exceeds the number of observations.

iv) Hierarchical clustering. Is affected by the curse of dimensionality. As the number of predictors increases, the idea of distance becomes less meaningful, and clusters may become less distinct.

v) Lasso. Not significantly affected by the curse of dimensionality. Lasso is designed to address this by encouraging sparsity by shrinking some coefficients to exactly zero, effectively selecting a subset or relevant predictors.

1. i) After changing the units by multiplying each predictor by c, the relationship between the coefficients can be expressed as αhatj = Bhatj / c for j = 1,2,…p. The coefficient for each transformed predictor Zj in the second model αhatj is equal to the original coefficient Bhatj­ divided by the constant c by which the units were changed. The intercept term remains the same so αhat0 = Bhat0

ii) While the underlying patterns in the data are the same after the linear transformation, the specific clustering assignments obtained by K-means clustering may be different because of the scaling effect. K-means relies on distances between data points. So, when you change the units of the predictors by multiplying with a constant, we are stretching or compressing the space along these dimensions.

**Part 2: Coding**

1. My optimal m was 3

Here are the CV errors for each m:

M=2, CV error = 0.33

M=3, CV error = 0.32

M=4, CV error = 0.32

M=5, CV error = 0.32

1. Using m=3, misclassification rate on the test set = 0.34
2. Using m=3, OOB misclassification rate on the full data set = 0.3051948
3. Phat(Y=1|X) = 0.754
4. Standard error = 0.4530577