Jillian Egland

DS 303 Exam 2

Question 1: Concept Review

1. Yes, I agree with her approach for estimating test MSE. She is using her ridge regression to predict new values for the test set using the best lambda found and calculating test MSE for it.
2. Multicollinearity
   1. Least Squares Linear Regression
      1. For least squares, if there is multicollinearity, you can still use the least squares for prediction, but you cannot do any inference on your model. The multicollinearity will result in a large standard error which reduces the power of the hypothesis test.
      2. \*\* cannot take the inverse
   2. Ridge Regression
      1. Works well in the presence of multicollinearity. When there is multicollinearity, there is large variance in the coefficient estimates. Ridge Regression adds bias to the model, which will reduce the overall expected test MSE.
      2. \*\* can handle perfect correlation
   3. LDA
      1. Multicollinearity makes the regression coefficients and significance variable and untrustworthy, but the predictions are still accurate
      2. Causes problems: Collinearity makes the k class matrix close to the singular matrix, so it makes it difficult or inaccurate to calculate the inverse matrix
      3. \*\* can’t take inverse
   4. QDA
      1. Causes problems for QDA. We have to take the inverse for QDA which is problematic with multicollinearity. We can still take the inverse if it is not full rank, but the classifier gets really large. If the data is perfectly correlated, we cannot take the inverse.
3. A regularized model can achieve better test error but still be wrong. Since the beta regression estimates are overestimated and underestimated, they have high variance. Regularized regression adds bias into the model, which is why the betas are biased. Adding bias offsets the variance in the bias-variance tradeoff, which lowers the overall test MSE of the model. While the model is systematically wrong, it has lower variance which can balance and make a better test MSE.
4. P(Y=1|X) = P(X|Y=K)\*P(Y=K)/P(X)
   1. P(Y=K) = sum of y’s
   2. P(X) = πk
   3. P(X|Y=K) = 1/(σk\*sqrt(2π))\*exp(-1/(2(σk)2)\*(x-μk)2
   4. A test observation will be assigned to Y=1 if the above formula is greater than .5