Assignment #1 - chapter 2

Question 1

a.Using Caret tuning function, optimal alpha is 0.97 and optimal lambda is 1. MSE is 975.9 and MAE is 19.64668 R Code

R output Results

```
##                         Model   MSE       MAE Alpha Lambda
## 1 Optimal Elastic Net Caret 975.9 19.64668  0.97      1
```

b.Using own 10-FOld CV function, optimal alpha is 0.05 and lambda is 4. Due to inefficient code and cpu constraints lambda sequence was used between 1 and 4. Best MSE is 984.5199 and best MAE is 19.89811. As caret is well-tested and has optimal code base, it outperformed my own function but end result of own cv function is not way too off from caret optimal estimates.

R code output

```
##                            Model      MSE      MAE Alpha Lambda
## 2 Optimal Elastic Net No Caret 984.5199 19.89811  0.05      4
```

c.Both caret and own optimal elastic net regression models performed worse than models on slide 83 except simple OLS regression, which suggests that there are highly correlated variables in the model which leads to multicollinearity and therefore affect coefficients or the model complexity of the elastic model is too high which leads to overfitting. Multicollinearity problem is dealt well by the Ridge regression model, while Lasso regression performs feature selection which leads to the end model to become less complex and therefore prevent overfitting. In this terms, elastic net as a compromise between Ridge and Lasso seems not a good fit for ames dataset.

Model	MAE	MSE
Basic Lasso	18.14	898.66
Ridge regression	**17.64**	**871.50**
OLS regression	19.64	1278.40
Lasso + OLS	18.15	908.20
Relaxed Lasso	18.30	905.13
Group Lasso	19.51	1264.62
Exponential group Lasso	20.63	1006.84

Question 2

After fitting various regression models on music dataset, own optimal elastic net cross validation model performed the a little better in terms of MSE (265.5898), while forward selection model performed a little better than others in terms of MAE (12.55716). It should be noted that all models performed similarly in terms of error and there is no obvious best model which outperforms others by a huge margin, however overall, elastic net models performed a little better than lasso, ridge and OLS.

R code end output with results

```
##                             Model      MSE      MAE
## 1                             OLS 280.7687 12.79714
## 2                        backward 273.4687 12.61967
## 3                         forward 267.6096 12.55716
## 4                        stepwise 270.2106 12.61318
## 5                Ridge Regression 270.3563 12.89190
## 6                Lasso Regression 266.1379 12.73492
## 7           Elastic Net Alpha=0.2 265.6397 12.74908
## 8           Elastic Net Alpha=0.5 265.8992 12.73696
## 9           Elastic Net Alpha=0.8 266.0715 12.73534
## 10    Optimal Elastic Net Caret 265.7537 12.73732
## 11 Optimal Elastic Net No Caret 265.5898 12.75316
## 12               Relaxed Lasso 267.5887 12.74739
```

Question 3: Multiple Choice Questions

1.Which of the following statements is true? (Only one statement is true)

A.Lasso regression L1-norm penalty's purposes are feature selection, fight with overfitting and smoothing.

Answer: False. Primary purpose of L1 penalty is to fight with overfitting. In Lasso, as the penalty increases, more features are driven to zero which is used for feature selection for the best model. However, L1 loss function does not have continuous derivatives and therefore smoothing is not one of the primary goals of Lasso regression.

B. As Ridge regression regularization parameter increases, regression coefficients also increase.

Answer: False. While regularization parameter increases, coefficients are reduced to near zero, however they are not discarded like in Lasso regression and therefore Ridge regression is not used for feature selection but to deal with multicollinearity.

C. The Lasso Regression regularization parameter increases the sparsity of the resulting solutions.

Answer: True. Lasso regression regularization parameter is L1-norm penalty which means as the penalty increases, more parameters will be driven to zero, therefore more sparse resulting solutions.

D. Multicollinearity affects the prediction capability of the model.

Answer: False. Multicollinearity affects the coefficient estimates, standard errors and p-values estimates of predictors and intercept which makes it much harder to understand the effect of the predictor on the target variable but it does not affect the prediction capability or goodness of fit.

2.What can be concluded about a model based on a tuning lambda parameter of Ridge Regression?

A. In case of very large lambda; bias is low, variance is high.

Answer: False. Very large lambda means the low complexity of the model which by bias-variance tradeoff signifies a model with high bias and low variance.

B. In case of very small lambda; bias is low, variance is high.

Answer: True. Very small lambda means the high comlexity of the model which by bias-variance tradeoff signifies a model with high variance and small bias.

C. When lambda is 0, model is uninterpretable.

Answer: False. When lambda is 0, Ridge Regression will act as an OLS regression and therefore all the predictors can be interpreted.

D. Lambda value affects the coefficients of predictors which can be used as a guideline for feature selection.

Answer: False. It's true that large lambda will shrink coefficient parameters towards zero, however it will not discard less important predictors by setting them to zero. Therefore, Ridge regression is not suitable for feature selection, while Lasso regression is used for that.