# Quiz # 4. AMS 580

**Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dear all, you have the rest of the lecture time to do this quiz. It is open book however you must do the work entirely on your own. Please turn on your video and mute your audio. For those of you who are not in the classroom, please connect to zoom with both your computer and your cellphone. Please turn in your quiz by 9:50am. Please submit your answer in two documents: 1. The Rmd file, and 2. The output file (word or pdf) – to the SBU Brightspace (BS). If you have difficulty submitting to BS, you may then email the solutions (with the subject of “Quiz 3, AMS 580”, and in one email with the two attachments), to your TA Ian at:** [**weihao.wang@stonybrook.edu**](mailto:weihao.wang@stonybrook.edu)

# Please refer to the following website website for steps for Penalized Regressions in R:

# <http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/153-penalized-regression-essentials-ridge-lasso-elastic-net/>

#### Penalized Regression with the Red Wine Data

The dataset ‘wine.csv’ is related to red variants of the Portuguese "Vinho Verde" wine [Cortez et al., 2009, <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>]. Our goal is to establish a regression model predicting the output variable wine ‘quality’ using the input variables – and in particular, using various penalized regression techniques to perform variable selection.

**Attribute Information:**  
Input variables (based on physicochemical tests):   
1 – fixed\_acidity   
2 – volatile\_acidity   
3 – citric\_acid   
4 – residual\_sugar   
5 - chlorides   
6 – free\_sulfur\_dioxide   
7 – total\_sulfur\_dioxide   
8 - density   
9 - pH   
10 - sulphates   
11 - alcohol   
Output variable (based on sensory data):   
12 - quality (score between 0 and 10)

1. Please use the random seed 123 to divide the data into 75% training and 25% testing.

1. Please first find the best Ridge Regression model using the training data. Please (a) find the best **λ** valuethrough cross-validation and display this value; (b) display the coefficients of the fitted model; and (c) make prediction on the testing data, and report the RMSE and the Coefficient of Determination .
2. Please first find the best LASSO model using the training data. Please (a) find the best **λ** valuethrough cross-validation and display this value; (b) display the coefficients of the fitted model; and (c) make prediction on the testing data, and report the RMSE and the Coefficient of Determination .
3. Please first find the best Elastic Net model using the training data. Please (a) find the best **tuning parameter** valuesthrough cross-validation and display these values; (b) display the coefficients of the fitted model; and (c) make prediction on the testing data, and report the RMSE and the Coefficient of Determination .



<https://en.wikipedia.org/wiki/Vinho_Verde>