

Homework 7

Make a copy of this document or answer in a separate document.

1. (1 point) *What is the difference between Lasso and Ridge Regression? When would you use one over the other? How would you decide?*

The difference between Ridge Regression and Lasso is the penalty term. Ridge regression is the sum of squares of coefficients while lasso is the sum of absolute values of coefficients. Ridge regression shrinks coefficients toward zero but rarely makes them exactly zero. Lasso regression can shrink some coefficients to exactly zero which performs feature selection. You use ridge regression when there are many correlated predictors and you want to keep all features in the model. You use lasso regression when you suspect only a subset of features are important and want a more interpretable model. I would decide depending on how many features I want and how interpretable I want the model to be.

2. (2 points) *Build a model using the pima indians dataset to predict BMI using ordinary least squares. Evaluate your model using a test set which is made of 30% of the original data. Report your test MSE.*

Test set MSE: 47.5293001851732

3. (1 point) *Evaluate your model using k -fold cross validation ($k = 5$) and LOOCV, report both cross validated MSEs.*

5-Fold CV MSE: 46.5946
LOOCV MSE: 46.8664

4. (1 point) *What is PCA and what are principal components?*

PCA/Principal Component Analysis is a dimensionality reduction technique used to simplify a dataset while preserving as much as the original variability as possible. It transforms the original correlated features into a new set of uncorrelated variables called principal components. Principal components are linear combinations of the original features. Each principal component captures a portion of the variance in the data.

5. (1 point) *ISLP 6.6 problem 2.*

- a) iii. less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance
- b) iii. less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance
- c) ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias

6. (1 point) Build model with Ridge using RidgeCV to predict BMI. Report your optimal lambda (alpha, tuning parameter) value and your best MSE.

```
Optimal lambda (alpha): 46.4159
Best Ridge MSE: 45.4824
```

7. (1 point) Build a model with Lasso to predict BMI. Report your optimal lambda (alpha, tuning parameter) value and your coefficients. Which are eliminated by your model?

```
Optimal lambda (alpha): 0.0433
Lasso Coefficients:
Pregnancies      : -0.1655
Glucose          : 0.5967
BloodPressure    : 1.5292
SkinThickness    : 2.5500
Insulin          : -0.1313
DiabetesPedigreeFunction: 0.1688
Age              : -0.2490
Outcome          : 1.7782

Eliminated features: []
```

Many problems should be solved using Python. The following notebook has been initialized with the packages and data that you need : [link here](#)

To turn in your assignment please take **SCREENSHOTS of your results and insert them into (this or another) document for submission. You will then **CONVERT THAT DOC TO A .PDF** before submitting that. Please embed a hyperlink* to your google notebook **INSIDE** .pdf that you submit.**

* : by “embed the link” I mean that I do NOT want you to copy/paste the disgusting hyperlink. Instead type a word, highlight the word and right click “insert link.” Make sure it works! Then save as .pdf and the link should preserve!

[Tran_HW7](#)