4.1: Penalized Regression

Dr. Bean - Stat 5100

1 Why Penalized Regression?

What are some undesirable consequences of having estimates of β_k's with inflated variance?

Why is it critical that we standardize our variables prior to using any of the penalized regression techniques?

Which of the following is NOT a good scenario to used penalized regression techniques? Why?

1. **Facebook is trying to create a model to predict the likelihood of a user responding positively to a certain type of ad.**
2. **The Huntsman Cancer institute is trying to determine which active genes in a person's DNA increase the likelihood of Pancreatic cancer.**
3. **The USU Agriculture Experiment Station is trying to determine if a change in the composition of feed significantly influences the milk output of dairy cows.**

Which method does NOT get estimated coefficients exactly equal to zero as the penalty parameter increases? Why?

- Ridge Regression
- LASSO
- Elastic Net

Given the following output, determine the value of the intercept for the following ridge regression model.

Obs	_TYPE_	_RMSE_	aluminate	trisilicate	ferrite	disilicate
1	PARMS	2.44601	1.55110	0.51017	0.10191	-0.14406
2	SEB	2.44601	0.74477	0.72379	0.75471	0.70905
3	RIDGEVIF	.	3.16388	5.67511	3.12746	5.94881
4	RIDGE	2.46291	1.31521	0.30612	-0.12902	-0.34294
5	RIDGESEB	2.46291	0.21499	0.10885	0.19630	0.10360

The MEANS Procedure

Variable	Mean
calories	95.4230769
aluminate	7.4615385
trisilicate	48.1538462
ferrite	11.7692308
disilicate	30.0000000