Statistical Learning
HomeWork 1

.1 Prediction Problem

Given

$K = 96$

$n = 20$

(1) So a High-Dimensional data Set-up... It is obvious we can't use OLS, We have to Include some penalty...

(2) We also checked Multi-collinearity and turns out it's not a genuine issue (very low correlation) ...

So, the two most obvious choices are

.1 Ridge

.2 Lasso

furthermore, we Know we can use all 96 variables...

So, we used both Ridge and lasso, on average Ridge was scoring better (using cross validations)

finally, Even though Ridge was performing slightly better than lasso, we decided to predict using lasso as it was a random choice .. (very low error difference)...

The final equation then can be rewritten as

$$RSS + \lambda \sum_{J=1}^{P} |\beta_J| \quad \text{——— (A)}$$

Residual sum of square

The code can explain further ...

<u>Problem:</u> $$\hat{\beta}(\delta, M) = \underset{\beta \in R^n}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} -\ln p_\beta (y_i \mid \tilde{x}_i^{(m)})$$

$$= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} (y_i - x_i^T \beta)^2$$

1) Algorithm: ~~for each~~ ~~δ~~

$\delta \rightarrow$ A vector of probabilities

$M \rightarrow$ A vector of Integers

delta_run (δ, M, features, response):

for each δ:

for each M:

Create Bernoulli_Random_Vector

Create New features (add Noise)

Perform OLS

Calculate error

Save error

→ Finally Select δ, M. for minimum error.

2) We applied <u>Ridge Regression</u> & <u>Lasso Regression</u> to the Supernova Dataset & finally compared errors among all three models.

(3) Supernova Dataset was split into Train-1, Train2, Train 3, Train 4, and Test-1, Test-2, Test-3, Test-4 and Dropout technique was performed on all these splitted Datasets along with Ridge & Lasso.

4)

	Average Error	Minimum Error	Maximum Error
Dropout Technique:	167	0.1	5
Ridge Regression:	2	3.6	4
Lasso Regression:	4	1.2	4

Average Error: Average Mean square Error over all Test Datasets.

Minimum Error: Minimum Mean square Error among all Test Datasets

Maximum Error: Maximum mean square Error among all Test Datasets.

Remarks: Generating The Noisy Model:

Bernoulli Random Vector is created.

$$\xi_c \sim \text{Bernoulli}(1, \delta, m) \qquad \{ m \text{ is no of rows of our dataset} \}$$

$$\xi_c = \frac{\text{Bernoulli}(1, \delta, m)}{(1-\delta)}$$

So ξ_c is a vector of $\left(0s, \frac{1}{(1-\delta)}\right)$ values.

For every column in the original dataset, we multiply the

- ξ_c vector elementwise with every column.

PART-II

SIMULATION STUDY:

1) a) Generating Feature set

We generate 100 features from Uniform distribution with observations between 1 & 100.

$X_i \sim \text{Unif}(1, 100, 100)$

b) Generating Response

We generate Y_i vector from a normal distribution having mean as the overall mean of X_i's scaled by a factor of 100, and variance as 10.

$$Y_i \sim N(\text{mean}(X_i) * 100,\ 10, 75)$$

2) After Generating the feature set & Y_i, we repeat the same procedure as done in part-I on this dataset. Dropout techniques, Ridge & Lasso were used to fit the model.

3) The Generated Dataset was split into 5 samples, Train-1, Train-2, Train-3, Train-4, Train 5 & Test-1, Test-2, Test-3, Test-4, Test-5 and the average, maximum & minimum errors were computed over all these Datasets.

	Average Error	Minimum Error	Max Error
Dropout techniques:	34.1	2.2e-26	100
Ridge Regression:	154	56	178
LASSO Regression:	111	100	214