Statistical learning

Report 4

Maciej Szczutko

2024-06-05

Elastic-net regression

The elastic net is another extension proposed by Hui Zou and Trevor Hastie in 2005 in "Regularization and variable selection via the elastic net". The main motivation for this method was handle the data with highly correlated data. The LOSS functon is defined as

$$\hat{\beta}_{\text{en}} = \text{argmin}_b \frac{1}{2}\|Y - Xb\|_2^2 + \lambda\left(\frac{1}{2}(1-\alpha)\|b\|_2^2 + \alpha\sum_{i=1}^{p}|b_i|\right)$$

to do: add caluclation.

$$\frac{\text{sgn}\left(\hat{\beta}_i^{\text{OLS}}\right)}{1+\lambda(1-\alpha)}\left(\left|\hat{\beta}_i^{\text{OLS}}\right| - \lambda\alpha\right)^+$$

What would be the value of the elastic net estimator with $\lambda = 1$ and $\alpha = 0.5$ if $\hat{\beta}_{\text{OLS}} = 3$?

According to above formula we got $\hat{B}_{en} = \frac{1}{1+1(1-0.5)} max(3-0.5, 0) = \frac{2}{3}\cdot\frac{5}{2} = \frac{5}{3}$.

Why do the LASSO, SLOPE, and elastic net perform variable selection, while ridge regression does not?

Because the LASSO, SLOPE and elastic net include L1 penalty when Ridge regression has additional penalty in terms of L2 norm. Ridge shrink some less important coefficient to 0 but not set them to 0. The other one do. This is because of the obtained estimator which perform some threshold selection.

Knockoff

The Knockoff method is designed to handle case when some variables is highly correlated with other within data and estimate coefficient might be large. Then Lasso or Slope might not detect is not important and shouldn't be use to predict response. In nutshell idea behind knockoff is as follow. We create copy of all features. Let say we have design matrix X and copy will be $\tilde{X}$. We want to preserve covariance between features but the covariance between exact copy of given feature should be as low as possible. Formally $\forall i \neq j \quad \text{cov}\left(X^i, \widetilde{X}^{(j)}\right) = \text{cov}\left(X^i, X^{(j)}\right)$.

Then we crate matrix $D = (X, \widetilde{X})_{nx2p}$. We using D as design matrix in LASSO model.

Then we need to deduce which feature are important on predicting response. We measure variable importance computing:

$$Z_j = \left|\hat{\beta}_j(\lambda)\right|, \quad \tilde{Z}_j = \left|\hat{\beta}_{j+p}(\lambda)\right|, \quad j = 1, \ldots, p$$

Note: Method description from Candes.

Then we compute

$$W_j = h\left(Z_j, \tilde{Z}_j\right) = -h\left(\tilde{Z}_j, Z_j\right), \quad j = 1, \dots, p$$

where the h mus be anti-symmetric. h is called symmetrized knockoff statistics.

Finally, the knockoff procedure selects predictors with large and positive values of W_j, according to the adaptive threshold defined as

$$T = \min\left\{t : \frac{1 + \#\{j : W_j \le -t\}}{\#\{j : W_j > t\}} \le \alpha\right\},$$

where α is the (desired) target FDR level.

For further work we use implementation *knockoff* library developed by Stanford scientist.

Simulation part

In this simulation we generate response vector from model

$$Y = X\beta + \epsilon$$

where $\epsilon \sim 2\mathcal{N}(0, I), \beta_i = 10$ for $i \in \{1, \dots, k\}, \beta_i = 0$ for $i \in \{k+1, \dots, 450\}$, and $k \in \{5, 20, 50\}$. The design matrix $X_{500x450}$ is orthogonal.

```
## NULL

## NULL
```

Table 1: Estimated FDRs and powers based on 100 replication.

	Lasso		Knockoff with Lasso		Knockoff with Ridge	
	FDR	Power	FDR.1	Power.1	FDR.2	Power.2
5	0.74	1	0.08	1	0.09	1
20	0.74	1	0.08	1	0.09	1
50	0.74	1	0.08	1	0.09	1

Table 2: Estimated means based on 100 replication.

	MSE			E(MEAN)		
	MSE_OLS	MSE_ridge	MSE_lasso	MSE_mean_OLS	MSE_mean_ridge	MSE_mean_lasso
5	829	381	19	1803	1918	354
20	829	381	19	1803	1918	354
50	829	381	19	1803	1918	354