Sparse Linear Regression

2019 年 9 月 16 日

Motivation: To improve OLSE

1. **prediction**

$$\text{MSE} = \text{bias}^2 + \text{variance}.$$

   trade bise for variance. "Shrinkage toward zero".

2. **interpretation** sparsity. 需要较少的变量

3. **stability** not sensitive to small perturbations of data.

1 Best xx selection

$$minimize RSS\left(\beta\right) s.t2 \|\beta\|_0 \leq k.$$

Note that $\|\beta\|_0 = \#\{\beta \neq 0\}$.

2 StepWise Selection

3 Shrinkage Methods

3.1 Ridge Regression

$$\begin{aligned}\hat{\beta} &= argmin\left(RSS\left(\beta\right) + \lambda\|\beta\|_2^2\right) \\ &= argmin(\|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2).\end{aligned}$$

Note that this is a convex optimization problem, thus can be addressed as follows:

$$\text{minimize } RSS\left(\beta\right) \text{ s.t } \|\beta\|_2^2 \leq \lambda$$

And this problem as an explict solution:

$$\hat{\beta}^{ridge} = \left(X^T X + \lambda I\right) Xy.$$

SVD:

$$X = UDV^T.$$

where X is $n \times n$ and U is n x n, and V is $p \times p$, and D is $n \times p$. And $rank\left(X\right) = r$

For OLS, we have,

$$X\hat{\beta}^{ols} = X\left(X^T X\right)^{-1} X^T y = uu^T y.$$

For Ridge Regression, we have

$$\begin{aligned}
X\hat{\beta}^{ridge}\left(\lambda\right) &= X\left(X^T X + \lambda I\right)^{-1} X^T y \\
&= UDV^T\left(VD^T U^T UDV^T + \lambda I\right)^{-1} VDU^T y \\
&= UDV^T\left(VD^T DV^T + \lambda I\right)^{-1} VDU^T y \\
&= UDV^T V(D^T D + \lambda)^{-1} V^T VDU^T y \\
&= UD(D^T D + \lambda)^{-1} DU^T y \\
&= \sum_{j=1}^{r} \frac{d_j^2}{d_j^2 + \lambda} u_j u_j^T y.
\end{aligned}$$

3.2 Ridge Regression from Baysesian Horizon

y $\sim N(x\beta, \sigma 2I)\beta_j \sim N(0, r^2)$