Project 10 Structured sparsity in genetics

Team K - Minghang Li

May 05, 2023

Problem 28: Uniqueness of predictions from the lasso

Proof. To prove that $\mathbf{X}\hat{\beta}_1 = \mathbf{X}\hat{\beta}_2$, it's equivalent to prove $\hat{\beta}_1 = \hat{\beta}_2$. We'll prove by contradiction.

Let the solution set be denoted as S, we have $\hat{\beta}_1, \hat{\beta}_2 \in S$.

Since S is convex, we also have $\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2 \in S$.

So that we have:

$$\frac{1}{2}\left\|\mathbf{y} - \mathbf{X}\left(\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right)\right\| + \lambda\left\|\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right\| = c^* \tag{1}$$

Define the loss function as $f : S \mapsto \|\mathbf{y} - \mathbf{X}\beta\|_2^2$. We know that the lineare loss function is stirctly convex, so that we have:

$$f\left(\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right) < \alpha f(\hat{\beta}_1) + (1-\alpha)f(\hat{\beta}_2) \tag{2}$$

Similarly, since l_1-nrom is convex, we have:

$$\left\|\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right\|_1 \leq \alpha\|\hat{\beta}_1\| + (1-\alpha)\|\hat{\beta}_2\| \tag{3}$$

Combining Equation 2 and Equation 3, we can start the following derivation:

$$\begin{aligned}
&\frac{1}{2}f\left(\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right) + \lambda\left\|\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right\| \\
&< \alpha\left(f(\hat{\beta}_1) + \lambda\|\hat{\beta}_1\|\right) + (1-\alpha)\left(f(\hat{\beta}_2) + \lambda\|\hat{\beta}_2\|)\right) \\
&= \alpha c^* + (1-\alpha)c^* \\
&= c^*
\end{aligned}$$

Just a re-write:

$$\frac{1}{2}f\left(\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right) + \lambda\left\|\alpha\hat{\beta}_1 + (1-\alpha)\hat{\beta}_2\right\| < c^* \tag{4}$$

Equation 4 contradicts with Equation 1, which suggests that our assumption $\hat{\beta}_1 \neq \hat{\beta}_2$ is wrong.

□

Problem 29: Ridge regression solution

Given $n \times p$ matrix $\mathbf{X}$, response n-vector $\mathbf{y}$, and the parameter p-vector β, consider augmenting $\mathbf{X}$ with rows corresponding to $\sqrt{\lambda}$ times a $p \times p$ identity matrix I:

$$\mathbf{X}_* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda}\mathbf{I} \end{pmatrix}$$

and $\mathbf{y}$ is similarly augmented with p zeros at its end into $\mathbf{y}_*$.

Now, the least square objective function on the modified dataset turns out to be:

$$\begin{aligned} &(\mathbf{y}_* - \mathbf{X}_*\beta)^T(\mathbf{y}_* - \mathbf{X}_*\beta) \\ &= \mathbf{y}_*^T\mathbf{y}_* - 2\mathbf{y}_*^T\mathbf{X}_*\beta + \beta^T\mathbf{X}_*^T\mathbf{X}_*\beta \end{aligned} \tag{5}$$

The solution to the linear regression on the modified dataset hence follows:

$$\hat{\beta} = (\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T\mathbf{y} \tag{6}$$

Since we are just padding $\mathbf{y}$ with zeros, the augmented part of $\mathbf{X}_*$ will not have any effect when it is multiplied by $\mathbf{y}_*^T$:

$$\mathbf{y}_*^T\mathbf{X}_* = \mathbf{y}^T\mathbf{X} \tag{7}$$

And finally with augmented dataset $\mathbf{X}_*$ we have:

$$\begin{aligned} &\mathbf{X}_*^T\mathbf{X}_* \\ &= \begin{pmatrix} \mathbf{X}^T & \sqrt{\lambda}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda}\mathbf{I} \end{pmatrix} \\ &= \mathbf{X}^T\mathbf{X} + \lambda\mathbf{I} \end{aligned} \tag{8}$$

Plug Equation 7 and Equation 8 back into Equation 6, we delightfully sees that performing linear regression on the modified dataset gives:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \tag{9}$$

which is just the ridge regression solution.

Problem 30: Variable selection under various norms

TODO