15.095 Homework 2

Kim-Anh-Nhi Nguyen. MIT ID: 9137855521

2018-10-10

1 Algorithmic Framework for Linear Regression

Cf. notebook attached.

For dataset1:

- The variables with non-zero coefficient are: x[19], x[27] and x[56], i.e., x_4^2, $x_1 2^2$, and $log(x_{11})$
- The best hyperparameters are: $\Gamma = 0.001$ and $K = 3$
- Training R^2 value: 0.8432352378395275
- Validation R^2 value: 0.8127748811895958
- Testing R^2 value: 0.7697463987128148

For dataset2:

- The variables with non-zero coefficient are: x[16], x[18], x[19], x[22], x[24], x[27] and x[28], i.e., $\sqrt{x_2}$, $\sqrt{x_4}$, $\sqrt{x_5}$, $log(x_1)$, $log(x_3)$, $log(x_6)$, and $log(x_7)$
- The best hyperparameters are: $\Gamma = 0.01$ and $K = 7$.
- Training R^2 value: 0.9219065010999199
- Validation R^2 value: 0.8982248481520309
- Testing R^2 value: 0.9274666342602617

2 Convex Regression

Cf. notebook attached.

3 Primal and Dual Perspective on Sparse Regression

(a)

i.

To solve the problem the 'primal' way, we can simply minimize $f(\beta) = ||y - \beta||_2^2 + ||\beta||_2^2$ by calculating its gradient and find the values of β_i for which the gradient is null and such that the constraint $||\beta||_0 \leq k$ is satisfied.

$\nabla(f) = 0 \iff \forall i, \frac{\partial f}{\partial \beta_i} = 0$
$\frac{\partial f}{\partial \beta_i} = 0 \iff \frac{\partial}{\partial \beta_i}(\sum_{j=1}^{n}(y_j - \beta_j)^2 + (\beta_j)^2) = 0 \iff \frac{\partial}{\partial \beta_i}((y_i - \beta_i)^2 + (\beta_i)^2) = 0$
$\iff \frac{\partial}{\partial \beta_i}(y_i^2 - 2y_i\beta_i + \beta_i^2 + \beta_i^2) = 0 \iff -2y_i + 4\beta_i = 0 \iff \beta_i = \frac{y_i}{2}$

As we also have the constraint $||\beta||_0 \leq k$, we cannot set all the components β_i of β to $\frac{y_i}{2}$.
In order to minimize $||y - \beta||_2^2$, the optimal β satisfies the following:

Let I_k be the set : $I_k = \{i \;/\; |y_i| \; is\; among\; the\; largest\; k\; values\; of\; |y_i| \; for\; 1 \leq i \leq n\}$

So,

$$\beta_i^* = \begin{cases} \frac{y_i}{2} & if\; i \in I_k \\ 0 & otherwise \end{cases}$$

The optimal value of the cost function is then:
$f(\beta^*) = \sum_{i=1}^{n}(y_i^2 - 2y_i\beta_i^* + 2(\beta_i^*)^2) = \sum_{i\in I_k}(y_i^2 - 2y_i\frac{y_i}{2}) + 2(\frac{y_i}{2})^2 + \sum_{j\notin I_k} y_j^2 = \sum_{i\in I_k}(y_i^2 - \frac{y_i^2}{2}) + \sum_{j\notin I_k} y_j^2$
$f(\beta^*) = \sum_{i\in I_k}(\frac{y_i^2}{2}) + \sum_{j\notin I_k} y_j^2$

ii.

In the 'dual' way, the dual reformulation that eliminates the variable β is :

$$\min\; c(s) = y'(I_n + \frac{1}{\lambda}\sum_j s_j K_j)^{-1} y$$
$$s.t.\; s \in S$$

Where $S := \{s \in \{0,1\}^p : 1's \leq k\}$
Here, $X = In$ and $\lambda = 1$, so we are left with:

$$c(s) = y'(I_n + \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_n \end{pmatrix})^{-1} y$$

$$c(s) = y' \begin{pmatrix} 1+s_1 & 0 & \dots & 0 \\ 0 & 1+s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1+s_n \end{pmatrix}^{-1} y$$

$$c(s) = y' \begin{pmatrix} \frac{1}{1+s_1} & 0 & \dots & 0 \\ 0 & \frac{1}{1+s_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{1+s_n} \end{pmatrix} y$$

$$c(s) = \sum_{i=1}^{n} \frac{y_i^2}{1+s_i}$$

The dual formulation is then :

$$\min \sum_{i=1}^{n} \frac{y_i^2}{1+s_i}$$

$$s.t.\, s \in S$$

The optimal solution for this dual formulation is clearly:

$$s_i^* = \begin{cases} 1 & if\ i \in I_k \\ 0 & otherwise \end{cases}$$

The optimal value of the cost function is then:

$$c(s^*) = \sum_{i \in I_k} (\frac{y_i^2}{2}) + \sum_{j \notin I_k} y_j^2$$

We can notice that this optimal cost is the same as obtained previously for the "primal" way.

(b)

Let's apply the cutting plane method for the dual formulation.

Step 1: We pick $s_1 = [0, 0, ..., 0] \in S$ and $C_1 = s_1$

Step 2: We calculate

$$z_1^* = \min_{s \in S} [\max_{\bar{s} \in C_1} (c(\bar{s}) + \sum_{i=1}^{n} \frac{\partial c(\bar{s})}{\partial s_i} (s_i - \bar{s}_i))]$$

For a given s_i, $\frac{\partial c(s)}{\partial s_i} = \frac{\partial}{\partial s_i} (\sum_{j=1}^n \frac{y_j^2}{1+s_j}) = \frac{\partial}{\partial s_i} (\frac{y_i^2}{1+s_i}) = y_i^2 \frac{\partial}{\partial s_i} (\frac{1}{1+s_i}) = y_i^2 \frac{\frac{1}{1+s_i}\Big|_{s_i=1} - \frac{1}{1+s_i}\Big|_{s_i=0}}{1-0} = y_i^2 (\frac{1}{2} - 1)$

So, for $1 \leq i \leq n$:, $\frac{\partial c(s)}{\partial s_i} = \frac{-y_i^2}{2}$

There only is one element in C_1, so :

$$\max_{\bar{s} \in C_1} (c(\bar{s}) + \sum_{i=1}^{n} \frac{\partial c(\bar{s})}{\partial s_i} (s_i - \bar{s}_i)) = c(s_1) + \sum_{i=1}^{n} \frac{\partial c(s_1)}{\partial s_i} (s_i - (s_1)_i) = \sum_{i=1}^{n} \frac{y_i^2}{1 + (s_1)_i} + \sum_{i=1}^{n} \frac{-y_i^2}{2} s_i = \sum_{i=1}^{n} y_i^2 - \frac{y_i^2}{2} s_i$$

$$= \sum_{i=1}^{n} y_i^2 (1 - \frac{s_i}{2})$$

To minimize this sum over S, we need to minimize each component $y_i^2 (1 - \frac{s_i}{2})$, so maximize s_i for the biggest values of y_i^2, i.e., $|y_i|$.

Therefore, the optimal value is :

$$(s_1^*)_i = \begin{cases} 1 & if\ i \in I_k \\ 0 & otherwise \end{cases}$$

$$z_1^* = \sum_{i=1}^{n} y_i^2 (1 - \frac{(s_1^*)_i}{2}) = \sum_{i \in I_k} (\frac{y_i^2}{2}) + \sum_{j \notin I_k} y_j^2$$

$$c(s_1^*) = \sum_{i \in I_k} (\frac{y_i^2}{2}) + \sum_{j \notin I_k} y_j^2$$

Step 3: Therefore, $z_1^* = c(s_1^*)$, so the algorithm stops here, after 1 iteration.

We would only add one cut. We can notice again that the final step that we got gives the same solutions s^* and the same cost $c(s^*)$ as previously.