hw2test

Subhankar Ghosh

February 23, 2018

(a)

(1) (a) $$f(\beta, \beta_0) = \frac{1}{2n} \| y - \beta_0 - X\beta \|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

This can be written as

$$f(\beta, \beta_0) = \frac{1}{2n} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ji}\beta_j \Big)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The best β_0 that would optimize $f(\beta, \beta_0)$ the most can be found by partially differentiating w.r.t. β_0 and setting it to 0.

$$\frac{\partial f(\beta, \beta_0)}{\partial \beta_0} = \frac{1}{2n} \sum_{i=1}^{n} 2\,(y_i - \beta_0 - X\beta)(-1) = 0.$$

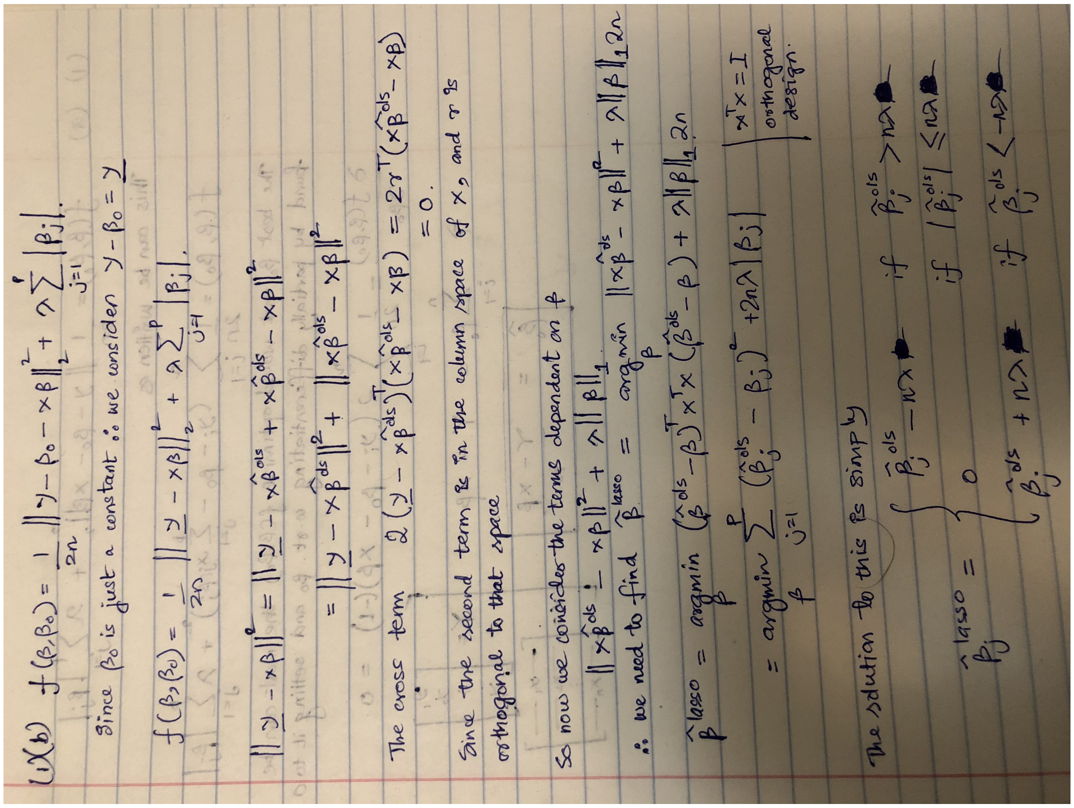
$$\Rightarrow \sum_{i=1}^{n} \beta_0 = Y - X\beta$$

$$\boxed{\hat{\beta}_0 = \frac{Y - X\beta}{n}}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} - x_1 - \\ - x_2 - \\ \vdots \\ - x_n - \end{bmatrix}$$

The one-variable optimization problem of Beta_0 based on this objective function.

(b)



(1)(b) $f(\beta,\beta_0) = \frac{1}{2n} \| y - \beta_0 - X\beta \|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|$.

Since β_0 is just a constant $\therefore$ we consider $y - \beta_0 = \underline{y}$

$$f(\beta,\beta_0) = \frac{1}{2n} \| \underline{y} - X\beta \|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|.$$

$$\| \underline{y} - X\beta \|^2 = \| \underline{y} - X\hat{\beta}^{ols} + X\hat{\beta}^{ols} - X\beta \|^2$$

$$= \| \underline{y} - X\hat{\beta}^{ols} \|^2 + \| X\hat{\beta}^{ols} - X\beta \|^2$$

The cross term $2(\underline{y} - X\hat{\beta}^{ols})^T (X\hat{\beta}^{ols} - X\beta) = 2r^T(X\hat{\beta}^{ols} - X\beta)$
$= 0.$

Since the second term is in the column space of X, and r is orthogonal to that space

So now we consider the terms dependent on β

$$\| X\hat{\beta}^{ols} - X\beta \|^2 + \lambda \| \beta \|_1$$

$\therefore$ we need to find $\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \| X\hat{\beta}^{ols} - X\beta \|^2 + \lambda \|\beta\|_1 2n$

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} (\hat{\beta}^{ols} - \beta)^T X^T X (\hat{\beta}^{ols} - \beta) + \lambda \|\beta\|_1 2n$$

$$= \underset{\beta}{\operatorname{argmin}} \sum_{j=1}^{p} (\hat{\beta}_j^{ols} - \beta_j)^2 + 2n\lambda |\beta_j|$$

$X^T X = I$ orthogonal design.

The solution to this is simply

$$\hat{\beta}_j^{lasso} = \begin{cases} \hat{\beta}_j^{ols} - n\lambda & \text{if } \hat{\beta}_j^{ols} > n\lambda \\ 0 & \text{if } |\hat{\beta}_j^{ols}| \leq n\lambda \\ \hat{\beta}_j^{ols} + n\lambda & \text{if } \hat{\beta}_j^{ols} < -n\lambda \end{cases}$$

The one-variable optimization problem of Beta_j based on this objective function.