Lasso Regression. (L1)

Only difference b/w Ridge and lasso is lasso consider $|w|$ mod of coefficient.

$$\therefore \text{MSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |w|$$

by increase alpha or λ coefficient can be zero.

Indirectly we can use lasso to feature selection for high dimentional data.

Key points.

1) Coefficient can be zero. This property can be used to remove unnecessary column (feature selection)

2) Coef having higher value decrease fastly and Intermediate value of λ give good feature selection.

4 Lasso shows sparsity, which means making coef of zero.

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |m|$$

$$b = \bar{y} - m\bar{x}$$

mathematical convenience

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |m|$$

$$= \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda |m|$$

$$\frac{\partial L}{\partial m} = 2 \sum (y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) + 2\lambda = 0$$

$$-2 \sum [(y_i - \bar{y}) - m(x_i - \bar{x})](x_i - \bar{x}) + 2\lambda = 0$$

$$-\sum [(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2] + \lambda = 0$$

$$-\sum (y_i - \bar{y})(x_i - \bar{x}) + m \sum (x_i - \bar{x})^2 + \lambda = 0$$

$$m \sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda$$

$$\boxed{m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda}{\sum (x_i - \bar{x})^2}}$$

LR

Loss

for $m > 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda}{\sum (x_i - \bar{x})^2}$$

for $m = 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

for $m < 0$

$$m = \frac{\sum (y_i - y)(x_i - \bar{x}) + \lambda}{\sum (x_i - \bar{x})^2}$$

Page No.
Date

Important question is that why lasso create sparsity and Ridge don't answer is in formula only

- Ridge : $m = \frac{\Sigma (y_i - \bar{y})(x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2 + \lambda}$.

as alpha is at denominator ∴ for m to the zero value in ~~numerat~~ numerator has to the zero. (in most of the cases numerator is positive)

- lasso.
alpha value is in numerator therefore by changing alpha value we can easily manipulate m. which create sparsity there are three cases where $m < 0$, $m = 0$, $m > 0$.