03-03-24

LASSO REGRESSION

* it is also called L1 Regularization. it helps to reduce overfitting.

$$L2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda ||W||^2 \rightarrow \text{Ridge Regression}$$

$$L1 = MSE + \underbrace{\lambda ||W||}_{\lambda\left[|W_1| + |W_2| + |W_3| + |W_4| + \dots |W_n|\right]} \rightarrow \text{Lasso Regression}$$

* L2 Norm is square and L1 Norm is mod
* in Lasso Regression if λ is very low then formula became the linear Regression. (overfitting) and if λ is very high then under fitting
  here $\lambda \geq 0$

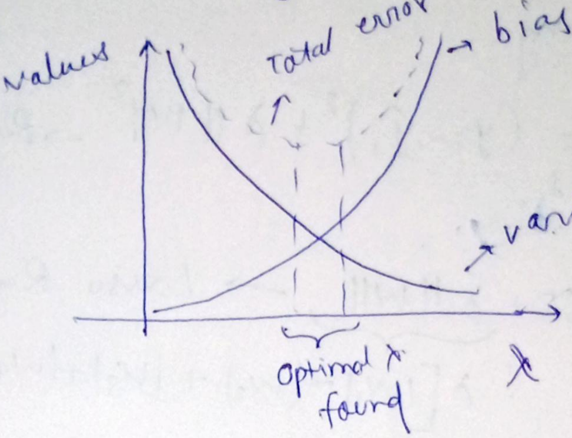
* here if alpha(λ) value is very high then coefficient became the zero. in Lasso Regression this situation not occur in Ridge Regression

* for example if we have a very high dimension data. in which overfitting scenario probability is very high (polynomial Regression and degree continuously increase) but if we apply Ridge Regression non- mana the coefficient is zero whereas in lasso if increase the value of λ (alpha) then which column is less important then there corresponding coefficient become zero(

X_1	X_2	X_3	X_4 ...	X_n	Y
⋮	⋮	⋮	⋮		

it is seems like work as a feature selection. hence at last the dimension of data is decrease

* for high dimension data prefer the Lasso regression rather than Ridge Regression.



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

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

now partial differentiation apply.

but here $\lambda|w|$ not differentiable hence we apply the cases.

(i) if $m > 0$ then $|m| = m$

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

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

rearrange

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

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

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

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

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

(ii) for $m=0$

$$m=\frac{\sum(y_i-\bar{y})(x_i-\bar{x})^{0}}{\sum(x_i-\bar{x})^2}$$ = Simple linear Regression

(iii) for $m<0$

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

Ques why sparsity occur in Lasso Regression.

from (i) if $m=\frac{YX-\lambda}{X^2}$ $\left\{\begin{array}{l} YX=100 \\ X^2=50\end{array}\right.$

$m=\frac{100-\lambda}{50}$ $\left\{\begin{array}{l|l|l} \lambda=0 & \lambda=10 & \lambda=100 \\ m=2 & m=9/5 & m=0\end{array}\right.$

$\longrightarrow$
m decrease

if $\lambda>100$
then we use formula from (iii)

$$m=\frac{YX+\lambda}{X^2}=\frac{100+\lambda}{50}=\frac{100+150}{50}=5$$

we want m is decrease but $\lambda>100$ the m is increase from 2 to 5 due to formula (iii)
hence here the worst situation is occur.
hence lasso stop at zero.