Ridge Regularization

(Ad)ding something extra in (MSE)L, so that overfitting tendency becomes low.

- Ridge (L_2)
- Lasso (L_1)
- Elastic Net.

Geometric intution

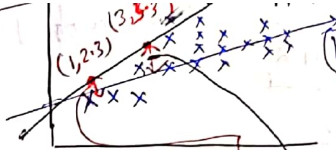
$y = mx + b$

- Extremly high ⇓ Overfitting ⇓ Soln ⇓ Reduce value of m to some extent ⇒ Regularization
- Extremly Low ⇓ underfitting ⇓
- high ⇓ y को निकालने में X का importance ज्यादा
- Low ⇓ y को निकालने में X का imp. कम

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

λ is hyperparameter $[0, \infty]$

this is regularization term.



L_R $y = 0.9x + 1.5$ → Assumed line After regularization.

(given) After training model

We will Calculate Loss ~~for~~ $= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda(m^2)$ for both L_R & L_N regression line.

Loss L_N (on train data) $\lambda = 1$

∴ No Loss because L_N passing through exactly through train data.

$= 0 + 1(m^2)$

$= 0 + 1 \times (1.5)^2$

$= 2.25$.

Loss LR (on train data After regularization) $\lambda = 1$

$(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \lambda(m^2)$

$= (2.3 - 0.9 \times 1 - 1.5)^2 + (5.3 - 0.9 \times 3 - 1.5)^2 + 1(1.5)^2$

$= 2.03$

$2.25 > 2.03$

Finally After regularization we decreased Loss.

It means even though L_N is passing exactly with train data but on adding $\boxed{\lambda(m^2)}$ its loss became high while L_R is having less Loss After adding $\boxed{\lambda(m^2)}$ eventhough L_R passing away from training data but variance is low.

Ridge regression Part 2

Mathematical formulation.

$y = mx + b$ → we regularized slope such a way that it ~~lowers heig~~ neither overfitt nor underfitt.

May be After regularization bias may ↑ but variance ↓.

$$L_R = \sum_{i=1}^{n} (y_i - \hat{y}_i) + \boxed{\lambda m^2}$$ → regularization term.

(L_R Loss of regg after regularization)

$$= \sum_{i=1}^{n} (y_i - mx_i - b)^2 - \lambda m^2$$

- $\frac{\partial L_R}{\partial b} = 0$ → it will be same as of simple linear reg.
- $\frac{\partial L}{\partial m} = 0$ → changes will be here since we have added λm^2 extra in eqn & derivati in term of m.

$$\frac{\partial L_R}{\partial b} = 0 \Rightarrow b = \underbrace{\bar{y} - m\bar{x}}$$

mean of y and x

$$L = \sum_{i=1}^{n} (y_i - mx_i - b)^2 + \lambda m^2$$

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

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

$\because$ to find value of $m \Rightarrow \frac{\partial L}{\partial m} = 0$

$$\therefore -2\sum_{i=1}^{n} (y_i - \bar{y} - mx_i + m\bar{x})(x_i - \bar{x}) + 2\lambda m = 0$$

$$\Rightarrow \lambda m - \sum_{i=1}^{n} [(y_i - \bar{y}) - m(x_i - \bar{x})](x_i - \bar{x}) = 0$$

$$\Rightarrow \lambda m - \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2 = 0$$

$$\Rightarrow \lambda m - \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) + m\sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$$

$$\Rightarrow \lambda m - \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})} + m\sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$$

Move to RHS:

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

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

$$\Rightarrow \boxed{m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x}) + \lambda}} \qquad \boxed{m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Equation of slope after regularization

this extra in denom penalizing

After normal regression