16/3/24

Loss Function or cost function

↓ D.P. one

↓ entire data

Gradient Techniques

① $\boxed{\text{mean square Error}}$

$$\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$$

⇒ also called variance.

Error = $(y-\hat{y})$

square Error = $(y-\hat{y})^2$

mean SQ = $\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$

note :- most of the time we will use this method bcoz after squaring we can easily find out the error.

Knowing our destination is completing half the journey.

② Mean Absolute Error → MAD

$$\frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}|$$

③ Root Mean Square Error

$$\frac{1}{n}\sum_{i=1}^{n}\sqrt{(y-\hat{y})^2} \Rightarrow \sqrt{var} = \text{std dev}$$

optimization / convergence algo

m – slope
c – intercept

gradient decent

$$\boxed{m_{new} = m_{old} - \eta\frac{\partial loss}{\partial m}} \rightarrow \text{derivative}$$

η → learning rate

$$\boxed{c_{new} = c_{old} - \eta\frac{\partial loss}{\partial c}}$$

Use case

let $x = 1, 2, 3$

$y = mx$

let $m = 1$, $m = 0.5$, $m = 0$

$x_1 = 1$
$\hat{y} = 1\times1 = 1$
$x_2 = 2$
$\hat{y} = 1\times2 = 2$
$x_3 = 3$
$\hat{y} = 1\times3 = 3$

$x_1 = 1$
$\hat{y} = 0.5\times1 = 0.5$
$x_2 = 2$
$\hat{y} = 0.5\times2 = 1$
$x_3 = 3$
$\hat{y} = 0.5\times3 = 1.5$

$x_1 = 0$
$\hat{y} = 0$
$x_2 = 0$
$\hat{y} = 0$
$x_3 = 0$
$\hat{y} = 0$

Cost fun $= \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2$ here $n = 3$ (x, 1, 2, 3)

for $m = 1$

$= \frac{1}{3} [(1-1)^2 + (2-2)^2 + (3-3)^2]$

$= 0$

for $m = 0.5$

$= \frac{1}{2} [(1-0.5)^2 + (2-1)^2 + (3-1.5)^2]$

$= 1.16$

for $m = 0$

$= \frac{1}{3} [(1-0)^2 + (2-0)^2 + (3-0)^2]$

$= \frac{1}{3} [1+4+9]$

$= 4.6$

So we will select $m=1$ & c value $=0$.
So it is our optimized parameter

→ it gives us no loss. so the m & c value we will select.

Ridge Regression (L2) & Lasso Regression (L1 reg)

Cost fun $= \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2 + \beta (m)^2$ (Ridge)

$= \frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2 + \cancel{\beta c}\ \beta |m|$ (Lasso)

β = hyperparameter