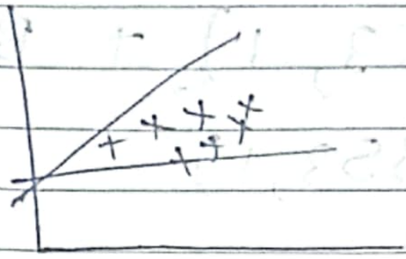
Ridge Regression

$y = mx + b$

$m \downarrow$ slope, $b \downarrow$ intercept

* When slope (m) is high overfitting occurs



We apply Bias Variance Trade off

we have to reduce the value of m to generalize the model.
If we reduce m that means Bias will ↑ and variance ↓

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

Panelty

m (Slope)

$\frac{\partial L}{\partial b}$ $\qquad$ $\frac{\partial L}{\partial m}$

$\boxed{b = \bar{Y} - m\bar{x}}$

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

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

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

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

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

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

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

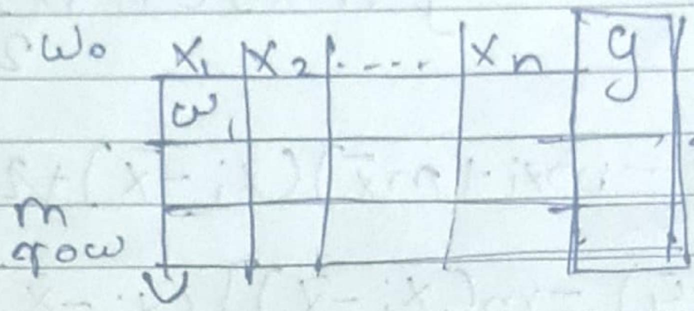
$$= \lambda m + m\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$$

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

Ridge Regression

Hyperparameters

Ridge Regression on n dimensional Data



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

in matrix form

$$(Xw - y)^T (Xw - y)$$

$$\begin{bmatrix} 1 & x_{11} & x_{12} \cdots & x_{1n} \\ 1 & \vdots & \vdots & x_{1n} \\ 1 & \vdots & \vdots & \vdots \\ 1 & x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

Now.

$$L = (Xw - y)^T (Xw - T)$$

This is for normal Linear Regression

$$L = (Xw - Y)^T (Xw - Y) + \lambda \|w\|^2$$

$$\lambda(w_0^2 + w_1^2 \cdots w_n^2$$

So, Ridge regression

$$\boxed{L=(XW-Y)^T(XW-Y)+\lambda W^TW}$$

$$L=W^TX^TXW-2W^TX^TY+Y^TY+\lambda W^TW$$

$$\frac{\partial L}{\partial W}=\cancel{2}X^TXW-\cancel{2}X^TY+0+\cancel{2}\lambda W=0$$

$$X^TXW+\lambda W=X^TY$$

$$(X^TX+\lambda I)W=X^TY$$

$$\boxed{W=(X^TX+\lambda I)^{-1}X^TY}$$

for ~~Rigde~~ Ridge Regrex.