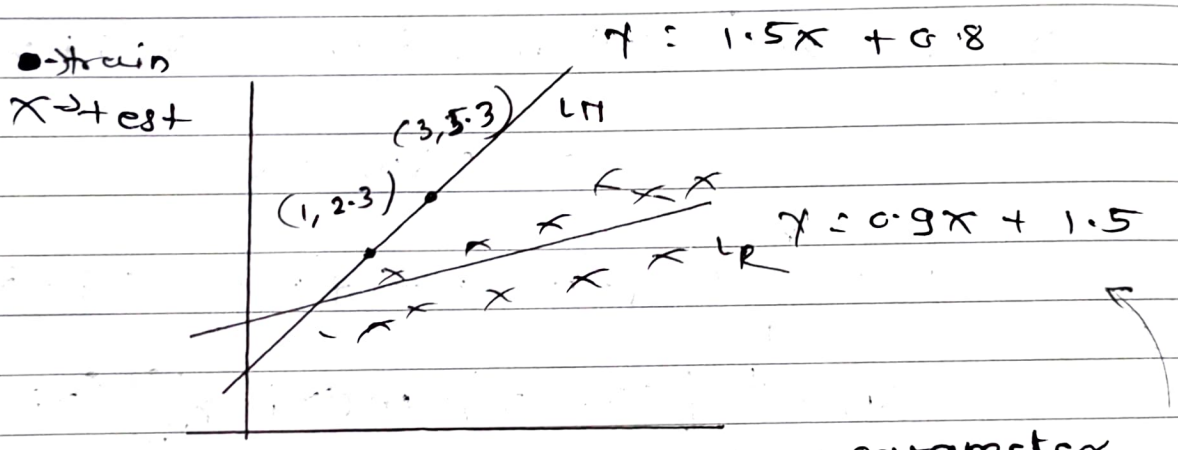
Regularization.

Page No.
Date

A Regularization is used to reduce overfitting which is caused because of using complex models

1) Ridge (L2)
2) LASSO (L1)
3) Elastic Net

for linear reg if value of m is more than it is overfitting. vice versa.



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

hyperparameter

Loss LH	Loss LR
$\lambda = 1$ $0 + (1.5)^2$ $= 2.25$	$\lambda = 1$ $(2.3 - 0.9 - 1.5)^2 +$ $(5.3 - 0.9(3) - 1.8) +$ $+ (0.9)^2$ $= 2.031$

As we take square of m^2
$\therefore$ L2

Derivation (Ridge Regression) [OLS]

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda m^2 \quad -(1)$$ (exactly similar Like linear Regression).

$$\frac{\partial L}{\partial b} = 0$$

$$\boxed{b = \bar{y} - m\bar{x}} \quad -(2)$$

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

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

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

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

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

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

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

As $\lambda \uparrow$ $m \downarrow$.

For n dimensions

$$L = (Xw - y)^T (Xw - y) + \lambda w^T w$$

$$L = [(Xw)^T - (y)^T][(Xw - y)] + \lambda w^T w$$

$$= (w^T X^T - y^T)(Xw - y) + (\lambda w^T w)$$

$$= w^T X^T Xw - \lambda X^T y - y^T Xw + y^T y + \lambda w^T w$$

$$= w^T X^T Xw - 2w^T X^T y + y^T y + \lambda w^T w$$

$$\frac{dL}{dw} = 2X^T Xw - 2X^T y + 0 + 2\lambda w = 0$$

$$X^T Xw + \lambda w = X^T y$$

$$(X^T X + \lambda I)w = X^T y$$

$$\boxed{w = (X^T X + \lambda I)^{-1} X^T y} \quad \text{- Ridge}$$

for linear reg

$$\boxed{w = (X^T X)^{-1} X^T y}$$