$: Var(f(x)) = E[(f(x) - E[f(x)])]$

Regularization-2

- Bias Variance Decomposition –

$$\text{Loss} = \text{bias} + \text{variance} + \text{irreducible error}$$
$$= \underbrace{\text{bias}^2 + \text{var.}}_{\text{reducible}} + \underbrace{\text{var}(\varepsilon)}_{\text{irreducible}}$$

→

Cgpa	iq	lpa	Prediz
—	—	9	9.1
—	—	8	7.9
—	—	7.9	7.5
—	—	7.6	8

we apply lr and get more results

$$Y = \beta_0 + \beta_1 \, \text{Cgpa} + \beta_2 \, \text{iq}$$

There are some errors in our predictⁿ. it is reducible & irreducible both.

Lets assume for irreducible error
mean = 0
$var = \sigma^2$

- Derivation

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n} = E[(y-\hat{y})^2]$$

$y = f(x) + \varepsilon = \theta + \varepsilon$

$\hat{y} = f'(x) = \hat{\theta}'$

$\theta = f(x)$
↳ constant

$\Rightarrow E[(\theta + \varepsilon - \hat{\theta})^2]$

$= E[(\theta - \hat{\theta})^2 + \varepsilon^2 + 2\varepsilon(\theta - \hat{\theta})]$

$= E[(\theta - \hat{\theta})^2] + E[\varepsilon^2] + E[2\varepsilon(\theta - \hat{\theta})]$

$+ 2\underbrace{E[\varepsilon]}_{0}E[\theta - \hat{\theta}]$

$$MSE = E[(\theta - \hat{\theta})^2] + E[\varepsilon^2]$$

$\Rightarrow var(\varepsilon) = \sigma^2 = E[(\varepsilon - \underbrace{E[\varepsilon]}_{0})^2]$

$\sigma^2 = E[\varepsilon^2]$

$$MSE = E[(\theta - \hat{\theta})^2] + var(\varepsilon)$$

$\Rightarrow E[(\theta - \hat{\theta})^2] = E[(\underbrace{\theta - E(\hat{\theta})}_{a} + \underbrace{E(\hat{\theta}) - \hat{\theta}}_{b})^2]$

$= E[(\theta - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \hat{\theta})^2 + 2(\theta - E(\hat{\theta}))(E[\hat{\theta}] - \hat{\theta})$

$E[(\theta - E(\hat{\theta}))^2] + E[(E(\hat{\theta}) - \hat{\theta})^2] + 2E[\theta - E(\hat{\theta})]$

$E[\cancel{E(\hat{\theta}) - \hat{\theta}}]$

$\boxed{E[E[\cancel{const}]] = E[\cancel{constant}]}$

$$+2(\theta - E[\hat{\theta}])\left\{E[\underbrace{E[\hat{\theta}]}_{\text{const}}] - E[\hat{\theta}]\right\}$$

$$E[\hat{\theta}] - E[\hat{\theta}] = 0$$

So

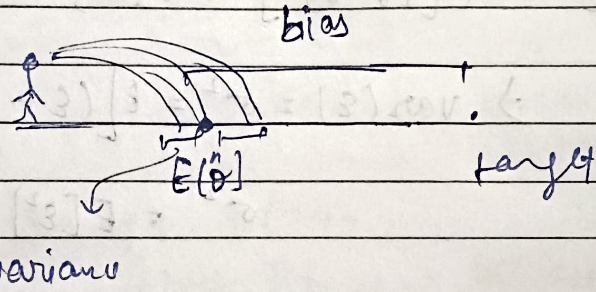
$$E[(\theta - E[\hat{\theta}])^2] + E[\underbrace{(E[\hat{\theta}] - \hat{\theta})^2}_{var}]$$

$$(\theta - E[\hat{\theta}])^2$$

$$(bias)^2$$

$$mse = (bias)^2 + var + var(\varepsilon)$$

- Sir has given a great example of golf



- Regularization reduces variance means it reduces overfitting

- we first get loss and then we decompose it into bias, var.

$$loss = \underbrace{(bias)^2}_{\downarrow} + \underbrace{var}_{\downarrow}$$

we can use complex model to reduce it

we use techniques like for → Ridge, LASSO, Elastic

Page No.

Date

- When to use Regularization -
  - → To prevent overfitting
  - → When we have high dimensionality data
  - → Reduce multicollinearity
  - → feature selection