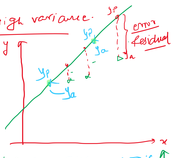
overfitting → Bias & variance

↳ Low Bias & High variance



$MSE = \frac{1}{n}\sum (y_a - y_p)^2$

$= 0$

Cost fun ⇒ C.F = 0

$MSE = \frac{1}{n}\sum (\quad)$ = ↑

Training Accuracy = 100% ↑

Re

overfitting

Test Acc. = 60% ↓

Regularization :-

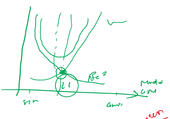
1) Lasso Regression (L1)

2) Ridge Reg. (L2)

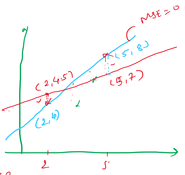
Underfitting → High Bias & Low vari

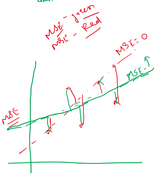
overfitting → Low Bias & High vari

Best Model → Low Bias & Low vari



Bias vari - tradeoff

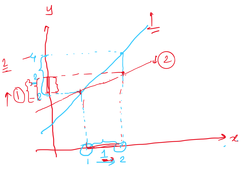




slope : $\frac{y_2 - y_1}{x_2 - x_1} = \frac{8-4}{5-2}$

$m = \frac{4}{3} = 1.33$

$m = \frac{7-4.5}{5-2} = \frac{2.5}{3} = 0.83$



1) Ridge (L2) :-

C.f : MSE = $\frac{1}{n}\sum (y_a - y_p)^2$ → λ (Slope)^2

↳ Reg. Parameter

= 0 + 1 × (1.3)^2

= 1.69

2) c.f : $\frac{1}{n}\sum (y_a - y_p)^2 + \lambda (\text{slope})^2$

= (small values) + 1 × (0.83)^2

= 0.5 + 0.69

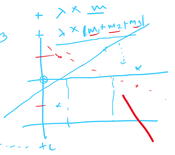
C.F = 1.19

Lasso Reg → $\frac{1}{n}\sum (y_a - y_p)^2 + \lambda |\text{slope}|$

↓

feature selection

m_1, m_2, m_3

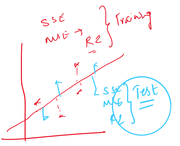


m

$y = mx + c$

$y = m_1x_1 + m_2x_2 + \dots + c$

$m \simeq 0$ → Remove

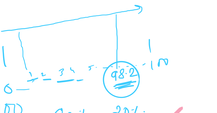


Multiple features → 100 → Lasso

Less → $m \simeq 0$

80

Less feat → Ridge Reg



Cross validation

CV = 5

k = 5