Revision

→ Assumption of LR,
→ MLR.
→ Multicollinearity → $x_1 \sim (x_2\ x_3\ x_4)$
↳ Variance Inflation factor (VIF)

$$VIF_i = \frac{1}{1 - R_i^2}$$

→ Need of Regularisation

Agenda

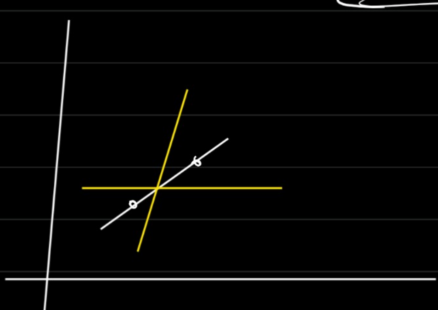
→ Practical implementation of Multicoll.
→ Lasso, Ridge, Elastic Net
→ Practical "
→ Polynomial Regression Implement

* Logistic Regression

* Regularisation

$$\boxed{CF = \frac{1}{n}\sum_{i=1}^{n}(y_{act} - mx_i - c)^2}$$

⇓
introduce some bias.



$$CF + \underbrace{\lambda\,(\text{slope})^2}_{\Uparrow}$$

bias or training error.

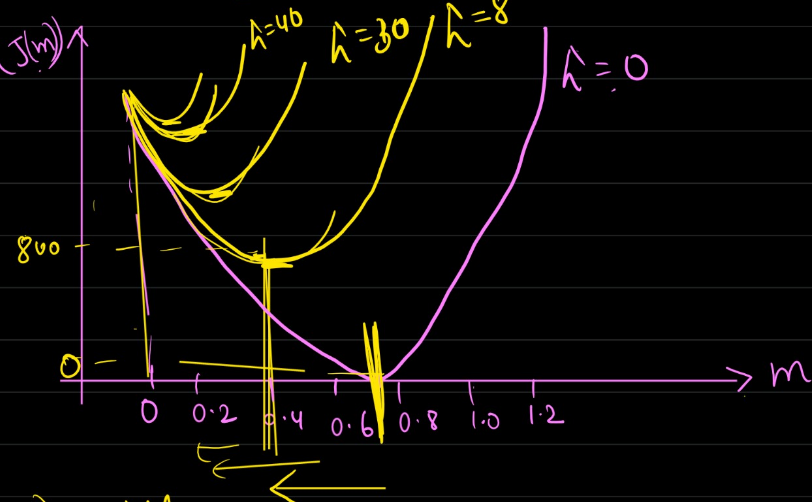
① ($L_2$ Regularisation, $L_2$ Norm) Ridge Regression

$$\boxed{\text{Ridge} = \frac{1}{n}\sum_{i=1}^{n}(y_{act} - mx_i - c)^2 + \lambda\, m^2}$$

$\lambda$ → 0 to $\infty$

relationship b/w $\hat{\lambda}$, costfn, m.

$\hat{\lambda}\uparrow$ Costfn $\uparrow$ minima shift.




$\boxed{\hat{\lambda}\uparrow \; m\downarrow}$

→ Slope (m) will never become 0

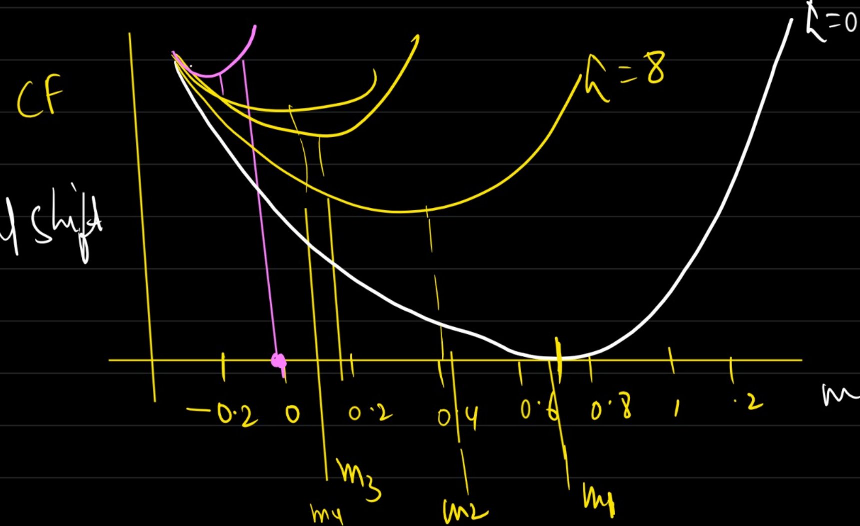
* Lasso Regression (L1 penalty, L1 norm)

$$\text{Lasso} = \frac{1}{n}\sum_{i=1}^{n}(y_i - mx_i - c)^2 + \hat{\lambda}\sum_{i=1}^{n}|\text{slope}|^{①}$$

CF vs $\theta$, $\hat{\lambda}$

$\hat{\lambda}\uparrow$ CF $\uparrow$ minimal shift

$\hat{\lambda}\uparrow$ Slope $\downarrow$




→ Slope can be 0.

## Ridge ($\lambda m^2$)

→ $L_2$ norm / penalty

→ $\lambda \uparrow$ $m \downarrow$

→ m will never become 0,
it will be close to 0.

→ Reduces the overfitting
(by reducing the coefficience)

→ effective in handling
multi collinearity.
(reducing all the coeff)


→ $X_1 \sim [X_2\ X_3]$

$X_1$ is also passed, $x_2, x_3$ is also passed ⟹ memorisation happen → overfitting

→ Ridge redues overfitting ⟹ reduce multicollinearity as well.

→ disadvantage

↳ doesnt make least important feature coefficient to 0, feature selection

$$y_{pred} = 2.3x_1 + 4.2x_2 + 0.01x_3 + 2$$

$\lambda \uparrow$ $m \downarrow$

## Lasso ($\lambda |m|$)

→ $L_1$ norm / penelty.

→ $\lambda \uparrow$ $m \downarrow$

→ m will become 0.

→ Advantage

↳ leads to feature selection.

→ removing insignificant feature

↓

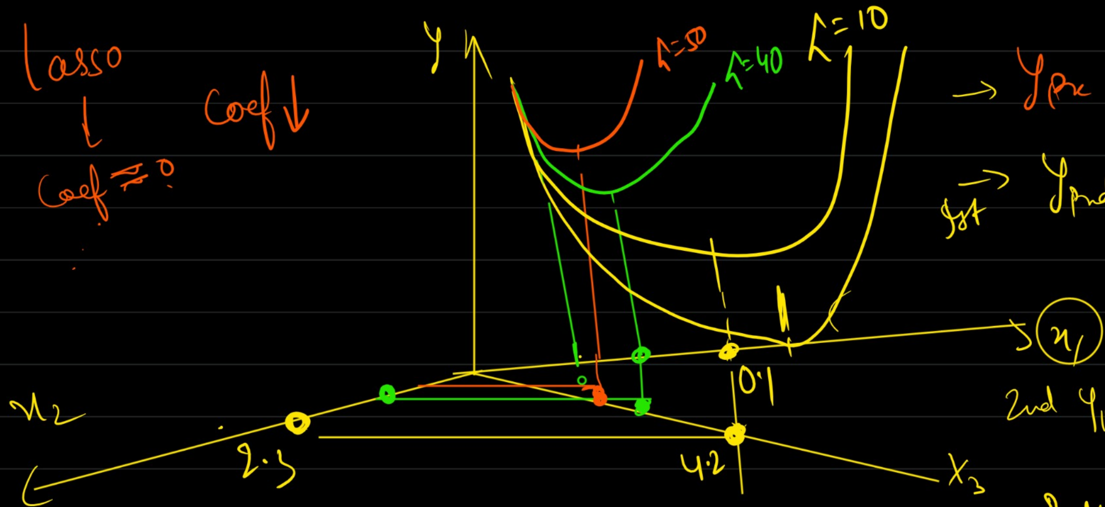
Automatic feature selection / Sparsity

→ multicollinearity

→ overfitting

Lasso → Coef ↓

Coef ≈ 0




$\rightarrow y_{pre} = 2.3x_1 + 4.2x_2 + 0.1x_3 + 2$

1st $\rightarrow y_{pred} = 2.1x_1 + 3.9x_2 + 0.08x_3 + \dots$

2nd $y_{pre} = 1.8x_1 + 3x_2 + 0.01x_3 \rightarrow 0$

3rd $y_{pr} = 1.5x_1 + 2.5x +$

As you $\lambda \uparrow$ the less significant feature will be removed in Lasso. ⇒ feature Selection method

③ Elastic Net
- Reducing overfitting ($L_2$)
- feature Selection ($L_1$)

$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{pred})^2 + \underbrace{\lambda_1 \sum_{i=1}^{n}(\text{slope}^2)}_{\text{Ridge, } L_2 \text{ Reduce overfitting}} + \underbrace{\lambda_2 \sum_{i=1}^{n}|\text{slope}|}_{L_1\text{, Lasso, feature selection}}$$

$\lambda_1, \lambda_2 \Rightarrow$ hyperparameter.

$(0, \infty)$

general formulae : Ridge: $\frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x)_i)^2 + \lambda(\theta_1^2 + \theta_2^2 + \theta_3^2 + \dots \theta_n^2)$

Lasso = " $+ \lambda(|\theta_1| + |\theta_2| + \dots |\theta_n|)$

H/W Why Lasso makes Coeff 0, Ridge near to 0 ?

Ridge $c=0$

①

②

$$CF = (y - mx_i)^2 + \lambda m^2$$

$$(a-b)^2 = a^2 - 2ab + b^2$$

$$CF = y^2 - 2ymx_i + m^2x_i^2 + \lambda m^2$$

$$\frac{\partial CF}{\partial m} = 0 - 2xy + 2mx^2 + 2\lambda m = 0$$

$$\Rightarrow m(x^2 + \lambda) = xy$$

$2x \to 2$

$$\boxed{m = \frac{xy}{x^2 + \lambda}}$$

$\lambda \to \alpha$ $\quad \frac{xy}{\alpha} = 0$

for larger $\lambda$, m can be 0.

$\lambda m$, $\lambda = 2$

Lasso

$$(y_i - mx)^2 + \lambda |m|$$

$$= y^2 - 2xm + m^2x^2 + \lambda m$$

$$= -2xy + 2xm + \lambda = 0$$

$$= m = \frac{2xy - \lambda}{2x^2}$$

$\lambda = 2$

for different $\lambda$, there can be infinite possibilities that Lasso can be 0.