## Last Class – May 27

1) Univariate LR ✓
2) Multi-variate LR ✓
3) R2 (Coeff of Determination) ✓
4) Adjusted R2 ✓
5) Feature Importance ✓
6) Model Interpretibility ✓

## Today's class – May 30

✓ 1) Recap / Review of Feature Importance & Model Interpretibility.

✓ 1.5) Gradient Descent Geometric Recap.

2) Assumptions (Statistical) behind Linear Regression

3) Practical Scenario.

4) Closed Form Solution of Linear Regression (if time permits)

5) AMA (11:15 – 11:20 pm)

## Feature Imp. & Model Interpretibility

$$y = w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + (2) \cdot x_3 + (5)\, x_4 + (-10) \cdot x_5 + \dots\ w_d \cdot x_d$$

$x_4 \rightarrow 5$ $\quad |-10| = 10$

$x_3 \rightarrow 2$

$x_5 \rightarrow (-10)$

$x = [x_1, x_2, \dots x_d] \rightarrow$ d-dim vector

F.I for L.R $\rightarrow |w_j|$, $j = 1, 2, \dots, d$

Cars 24 dataset $\rightarrow$ Resale value of the car

f1 $\rightarrow$ 1) Mileage $\rightarrow$ 24 km/l, 30 km/l, 12 km/litre

f2 $\rightarrow$ 2) km-driven $\rightarrow$ 30,000 km, 2,00,000 km, ...

f1 ↑ ⇒ SP ↑

f2 ↑ ⇒ SP ↓

f1 ↑ ⇒ SP ↑

+vely correlated

W1 for f1 should be +ve

W1 $\rightarrow$ +ve

f2 ↑ ⇒ SP ↓

-vely correlated

W2 $\rightarrow$ -ve

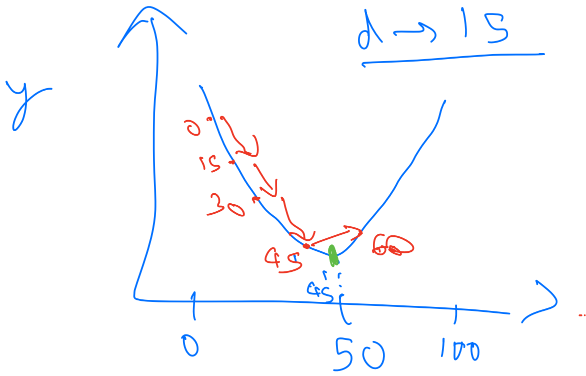
Feature Imp & Model Interpretation

$W_j = 0$ → is this possible ?

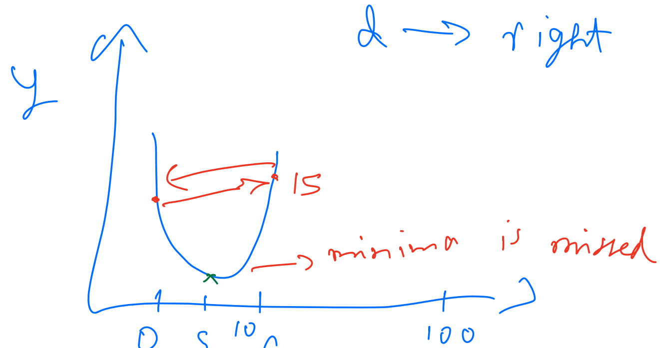
f3 → height of the car owner

↳ 5' 6", 5' 11", 5' 2"

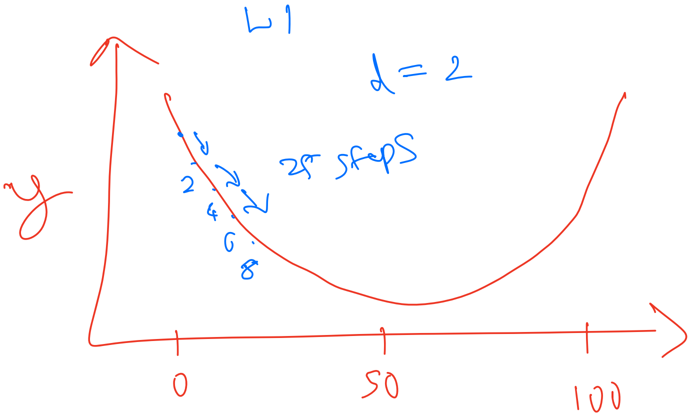
↳ irrelevant when it comes to predicting resale price



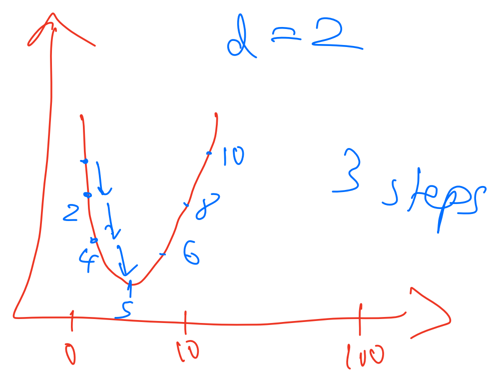

f1 →
↓
w1




f2 →
↓
w2




f1 →




f2 →

grad f1 → +1 to -1

grad f2 → +50 -50

Normalize $f1, f2, f3 \dots f_d$

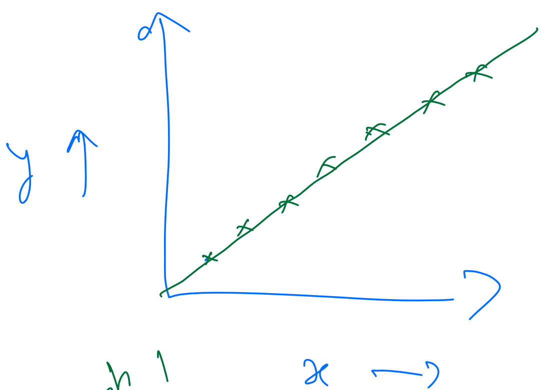
→ Standard Scaler $\quad f_j \rightarrow +1$ to $-1$

→ Min Max Scaler

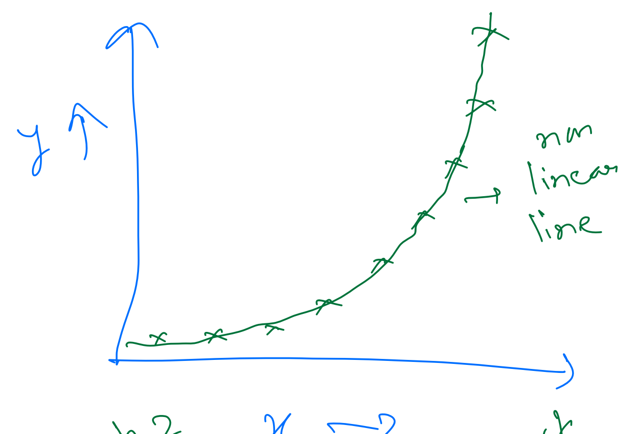
## Assumption (Statistical)

1) Dependant variable(y) is linearly related to independent variable (x)

$$y = mx + c$$

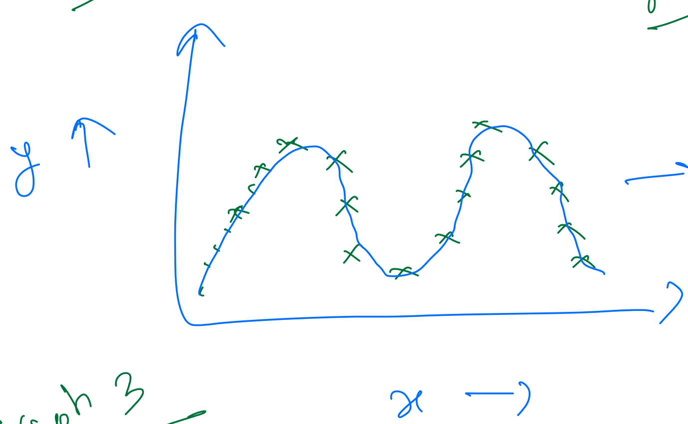



graph 1




graph 2

→ LR won't work




→ sine curve

→ LR won't work

graph 3

## 2) Multi-Collinearity Shouldn't exist

$f_1, \; f_2, \; f_3$

$$\boxed{w^* = (0, 1, 2, 3)}$$

$$y = 0 + 1 \times f1 + 2 \times f2 + 3 \times f3$$

Let's say → $f1$ & $f2$ are highly correlated

$\boxed{\text{Let's assume, } f2 = 1.5 f1}$

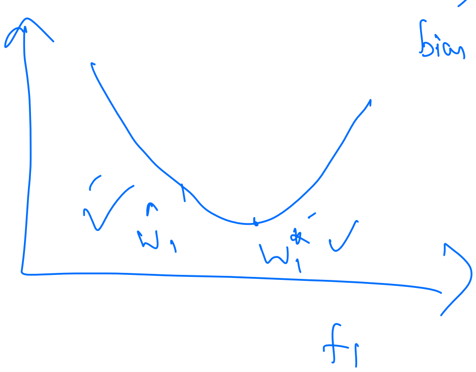
$$y = 0 + 1 \times f1 + 2 \times (1.5 f1) + 3 \times f3$$

$$= 0 + \underbrace{4 \times f1}_{} + 0 \times f2 + 3 \times f3$$

$$\rightarrow \hat{w} = \begin{array}{cccc} w_0 & w_1 & w_2 & w_3 \\ (0, & 4, & 0, & 3) \end{array}$$

$$w^* = \begin{array}{cccc} (0, & 1, & 2, & 3) \\ \text{bias} & f1 & f2 & f3 \end{array}$$

Multiple possibilities of weights




1) Model is Confused which is the optimum minima

→ Replication of results is not possible

2) Feature imp. is completely messed up ⇒ model interpretation is messed up

$$f2 = 1.5 f1 \Rightarrow \underline{f1} = \underline{0.67 f2}$$

$$y = 0 + 1 \times f1 + 2 \times f2 + 3 \times f3$$

$$= 0 + 1 \times (0.67 f2) + 2 \times f2 + 3 \times f3$$

$$= 0 + 0 \times f1 + 2.67 f2 + 3 \times f3$$

$$\begin{array}{cccccc} & & w_0 & w_1 & w_2 & w_3 \\ \bar{w} & = & (0, & 0, & 2.67, & 3) \end{array}$$

$$\hat{w} = (0, 4, 0, 3)$$

$$\begin{array}{cccccc} w^* & = & (0, & 1, & 2, & 3) \\ & & \text{bias} & f1 & f2 & f3 \end{array}$$

$$\boxed{y} = w_0 + w_1 f1 + w_2 f2 + w_3 f3$$

LR 1: $\boxed{f1 = \bar{w}_{21} f2 + \bar{w}_{31} f3}$ LR

$\boxed{y}$ $\boxed{R2 \rightarrow f1} \rightarrow \approx 1$

with $f1$ & my output & $f2, f3$ as my input

LR 2: $f2 = \bar{w}_{12} f1 + \bar{w}_{32} f3$

$R2 \rightarrow f2$

LR 3: $f3 = \bar{w}_{13} f1 + \bar{w}_{23} f2$ $R2 \rightarrow f3$

→ f1_R2 = 1 ⇒ f1 can be learnt from f2 & f3 perfectly

→ f1_ R2
→ f2_ R2
→ f3_ R2

VIF → Variance Inflation Factor (for each feature)

$$VIF = \frac{1}{1 - R2}$$ ✓

→ R2 = 1 ⇒ $\boxed{VIF = \frac{1}{1-1} \to \infty}$ →

R2 = 0 ⇒ $VIF = \frac{1}{1-0} \to 1$ ✓

R2 = −ve(−1) $VIF = \frac{1}{1-(-1)} = \frac{1}{2} = 0.5$ ✓

If R2 ↑, VIF ↑ higher value of R2, higher value of VIF

5 ≤ VIF < 10 → high multi-collinearity

VIF ≥ 10 → very high multi-collinearity

VIF < 5 → acceptable MC

f1, f2, ... $f_{10}$ → VIF for all f1, ... $f_{10}$

check which $f_j$ has VIF ≥ 10 → drop them

$f_1, f_2, \dots \dots f_{10} \rightarrow$ $f_8, f_9$ have VIF = 12

↓ dropped

$f_1, f_2, f_3, \dots f_7, f_{10} \rightarrow$ VIF for rest

1) which have VIF $\geq 10$ ↓

2) which have VIF $\geq 5$ ↓

What should remain is features with

$0 \leq \text{VIF} < 5$ ✓

10 features $\rightarrow$ 10 R 2 $\rightarrow$ 10 VIF

10 features $\rightarrow$ $^{10}C_2$ $\rightarrow$ Correlations

$$\frac{10 \times 9}{2} = 45$$

But using 45 correlations, we don't know which features to drop first.

## 3) Errors are Normally Distributed

→ $y_i$ → ground truth $\qquad e_i = y_i - \hat{y}_i$ → train data

→ $\hat{y}_i$ → predictions $\qquad$ ↓ error

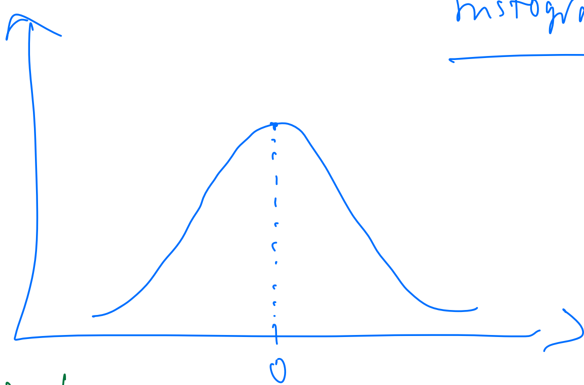
$e_i \rightarrow N(0, \sigma)$ $\qquad \sigma = K$

histogram of errors

↓

gaussian distribution check?

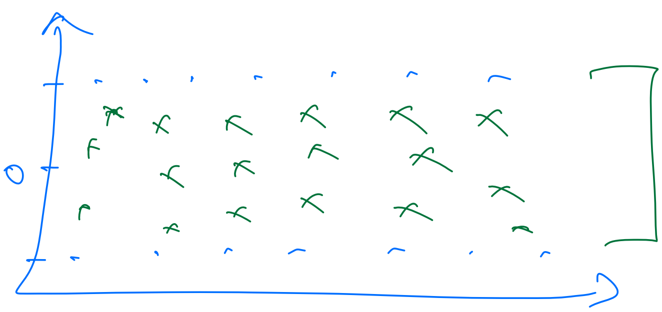
cnt ↑



$e_i$ →

KS - test (1) or AD - test (2)

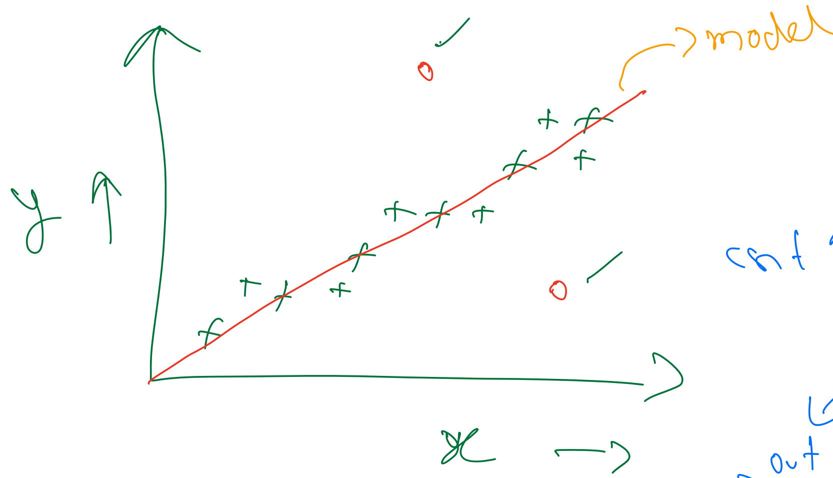
$e_i$ ↑



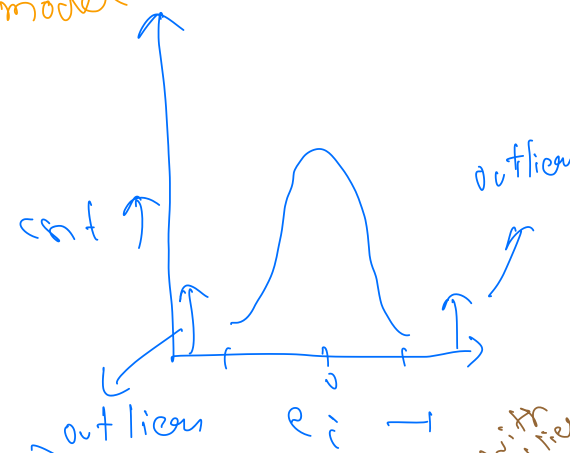
All my errors are within a certain range

$y_i$ →

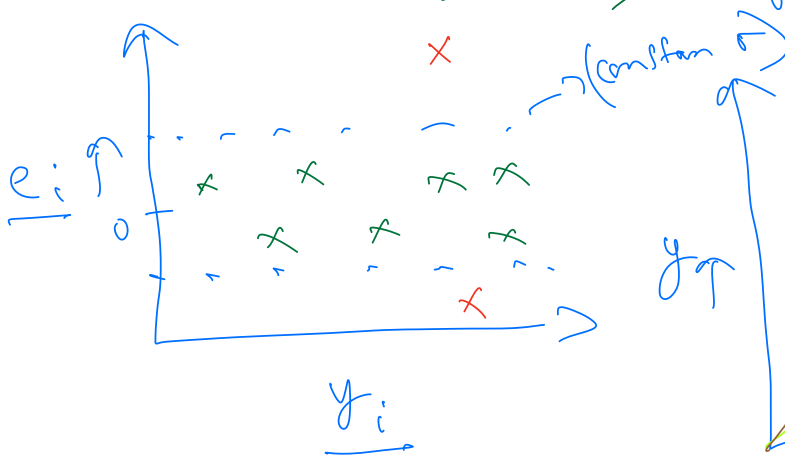
If my data has outliers, will my errors follow a gaussian distribution?



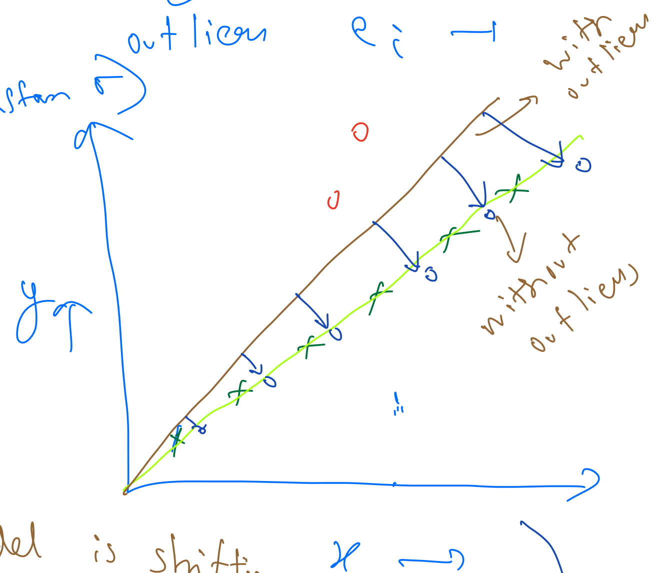










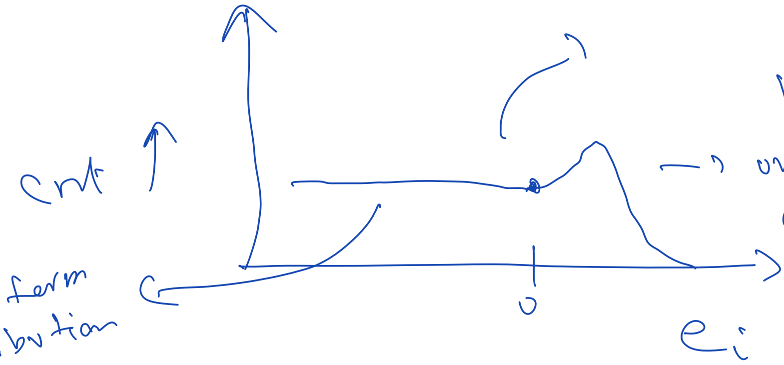




Model is shifting towards outliers

errors are getting impacted

errors will shift away from a gaussian distribution

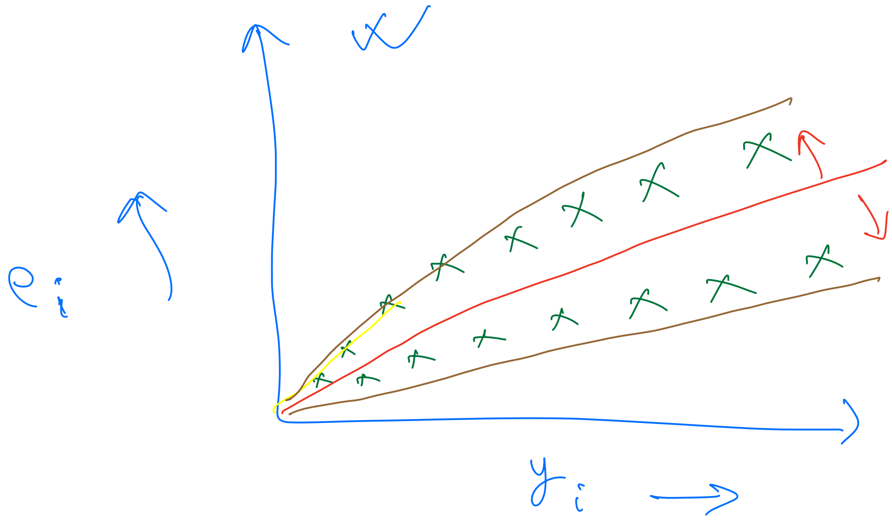




→ overall not gaussian

-ve values have a uniform distribution

→ 4) Heteroskedasticity shouldn't exist

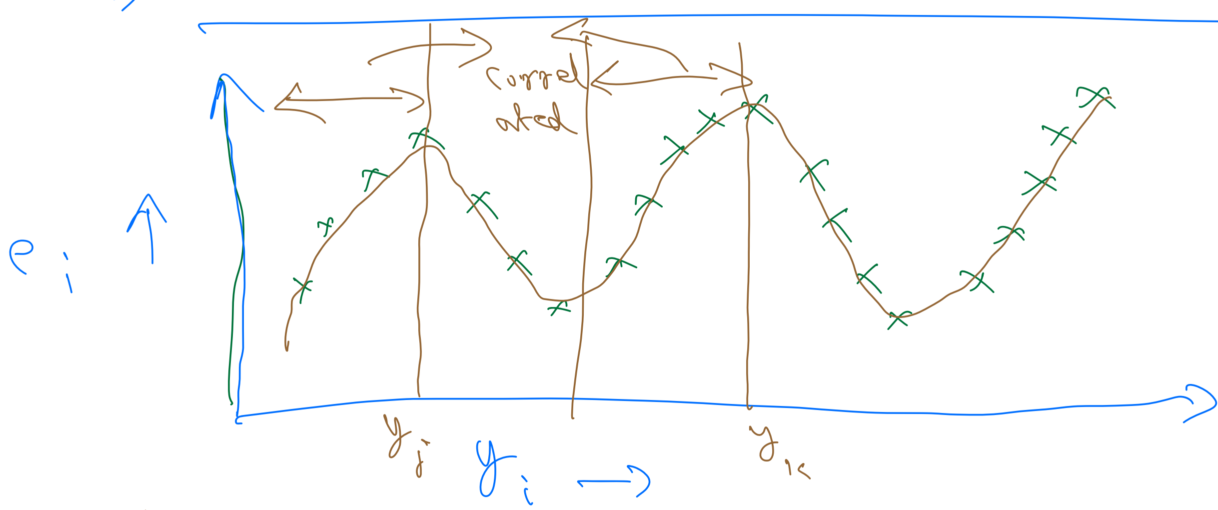



$e_i \to N(0, \sigma)$

std is no longer constant

$\sigma \to$ fn of time / $y_i$

5) Auto-correlation in the errors




$y_k$ & $y_j$ are same

$y = w_0 + w_1 x \longrightarrow LR$

$y = w_0 + w_1 \sin x + w_2 \cos x$

not a LR model

model is a great fit to a sinusoidal data

Training Dataset

- 60% Training Data
- 20% Validation Data (Proxy Test Data)
- 20% Test Data (Proxy for the real world)

V Data performance will determine the model complexity

→ meet min. criteria or not