Interview Prep Session 1

[Classical Machine Learning]

Warm Up Questions

Q: If you are training a LR to predict Farhite using Celcius as independent var. what will be the value of the intercept? (Quiz)

→ Automated | Easy

Soln → We know →

$$F = \frac{9}{5} \times C + 32$$

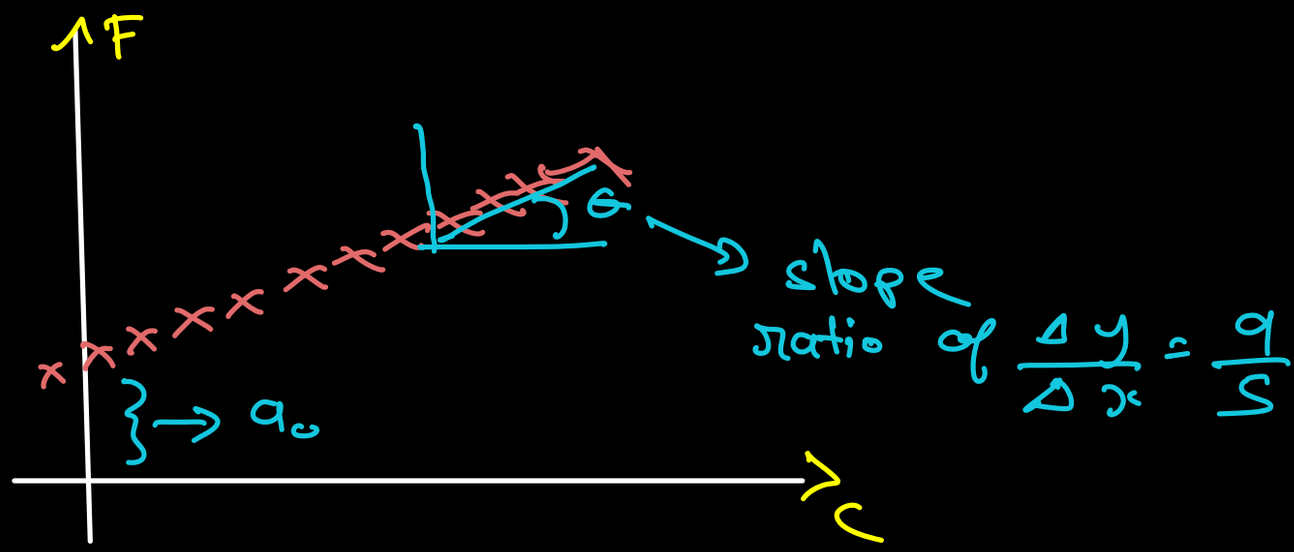
Train a LR, st.

$X = [-, -, -, -, -, -]$ Readings in C°

$y = \{\ldots\}$ Readings in F

We need $f(x)$, s.t

$$X \to f \to \hat{y} \sim y = \frac{9}{5} \cdot x + 32$$

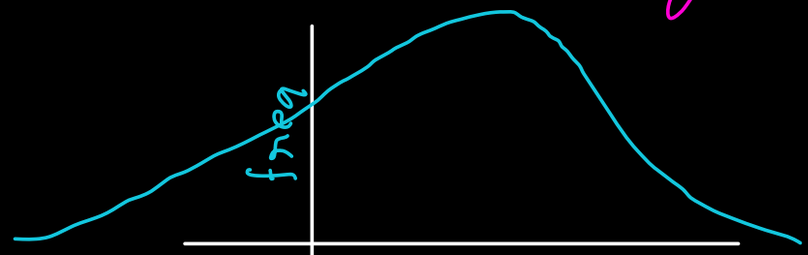


$$\therefore \hat{y} = \underbrace{a_1}_{\frac{9}{5}} C + \underbrace{a_0}_{32}$$ → intercept

Hence answer = $\boxed{32.}$

Q: After training a LR model, the residual error has the following distribution

Should this model go into production? (Quiz)



$-\infty$ $+\infty$

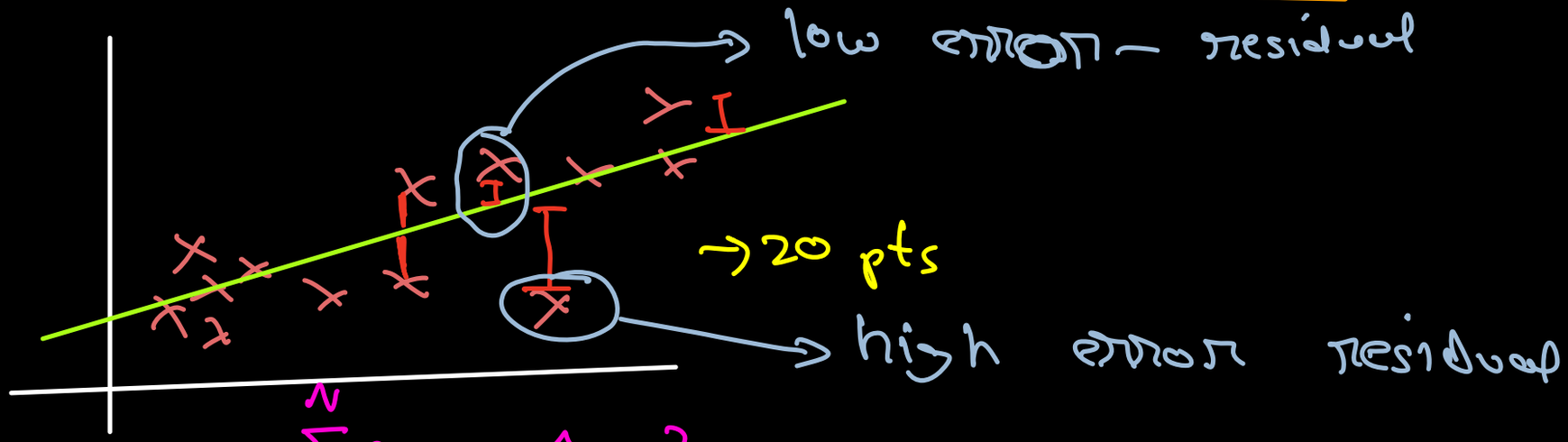
→ Automated | Easy

Sol^n → No

→ The error dist is not normal, assuptions violated

→ The mean of residual is non-zero, +ve, indicating high tendency to Over predict

Concept: Regression Residuals



$$MSE = \sum_{i=0}^{N} (y - \hat{y})^2$$

$(y - \hat{y})$ → residual

$$MSE = \sum_{i=0}^{N} r_i^2$$

$$R = y - \hat{y}$$

$$r_i = y_i - \hat{y}_i$$

Importance of Residuals

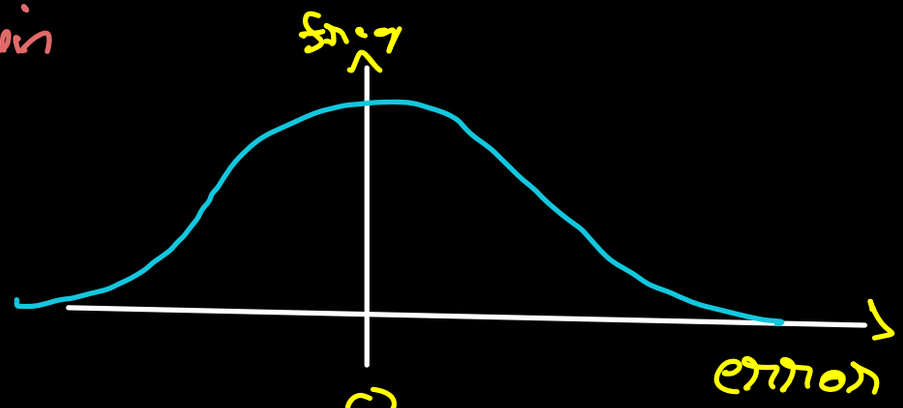
→ We just saw → all error metrics are computed on residuals

↳ **Analysis**

$X, y \xrightarrow[\text{train}]{} f(x) \rightarrow \hat{y}$; $y - \hat{y} = r$

n pts.

↳ histogram

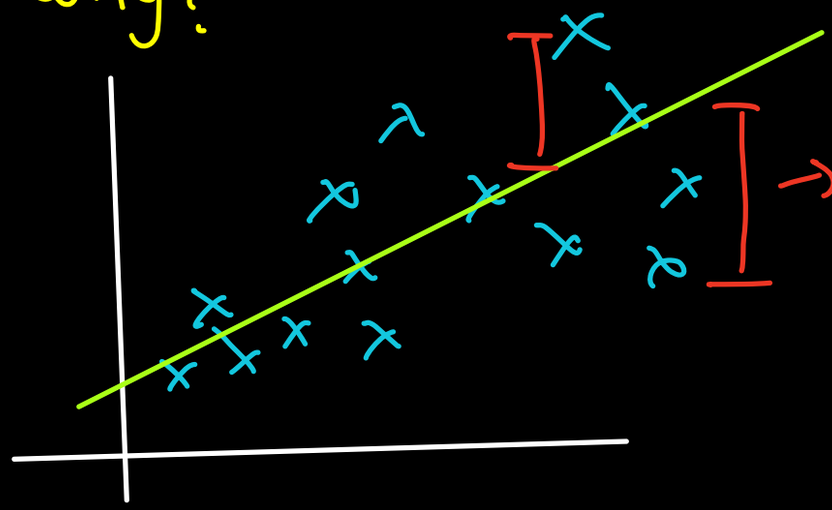


→ Ideally for LR → dist should be normal

→ Mean should be 0

→ Variance : lower the better

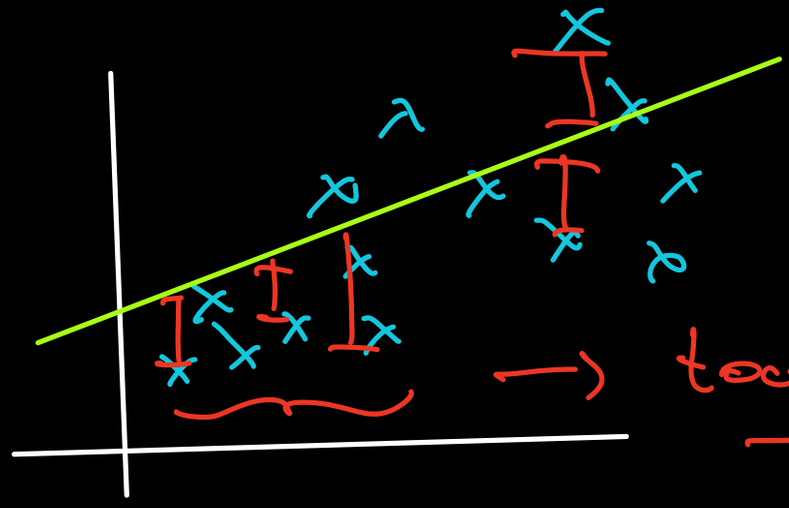
Why?



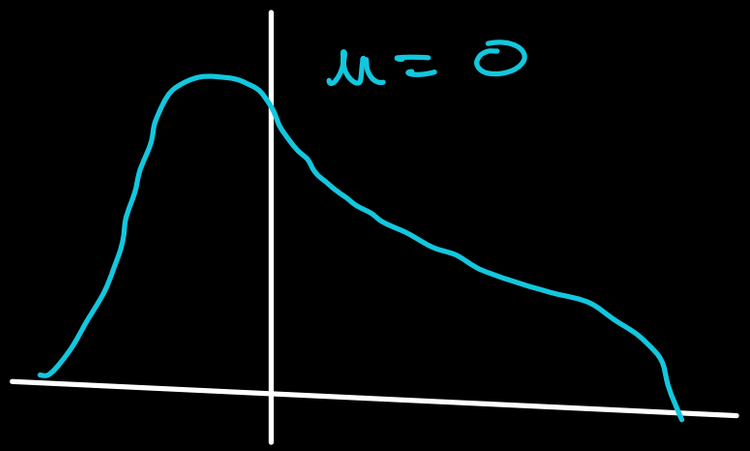
→ error!!

But overall
-ve error
≈ +ve error

→ We want symmetric errors.



→ too much -ve error } → over prediction



→ We want to predict "error rate" in test set.

→ In this model

→ if error > 0, error is low

if $x < 0$, error is high

⇒ assymetry !!

To make decisions we need confidence

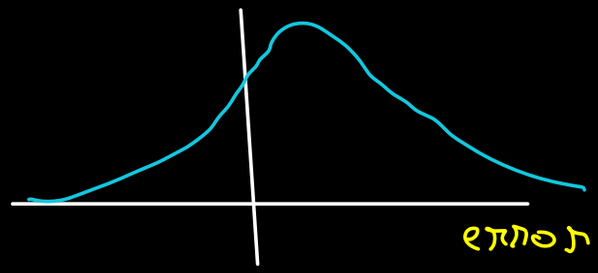
-> my prediction is 53.5 ± 3

w.p. $1 - \alpha$.

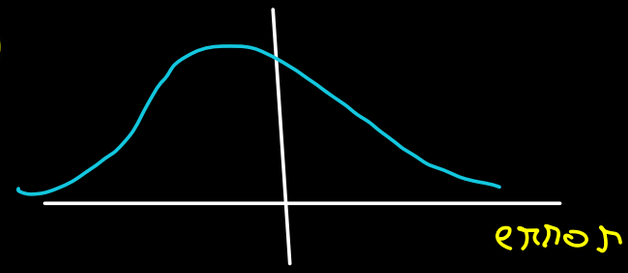
Q If a retail company has 50% profit margin for each sale, and 80% salvage price for unsold goods, while predicting Order Quantity, which of the following models seem best? (Quiz)

→ Automated / Live | Medium

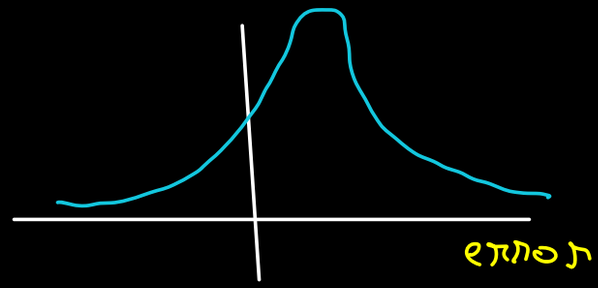
a)



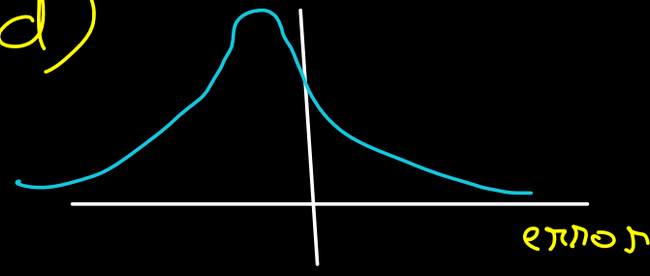
b)



c)



d)



Salvage price = Discounted price @ unsold goods are sold (Company looses money)

The company wants max profits

if over-produce → need to sell at salvage price
↳ Loss = 20%

if under-produce → Loosing potential profit
↳ Loss = 50%

Here, better to overpredict → $y - \hat{y} \leq 0$
→ Next, choose one with low variance (more accurate)

Hence Ans → d