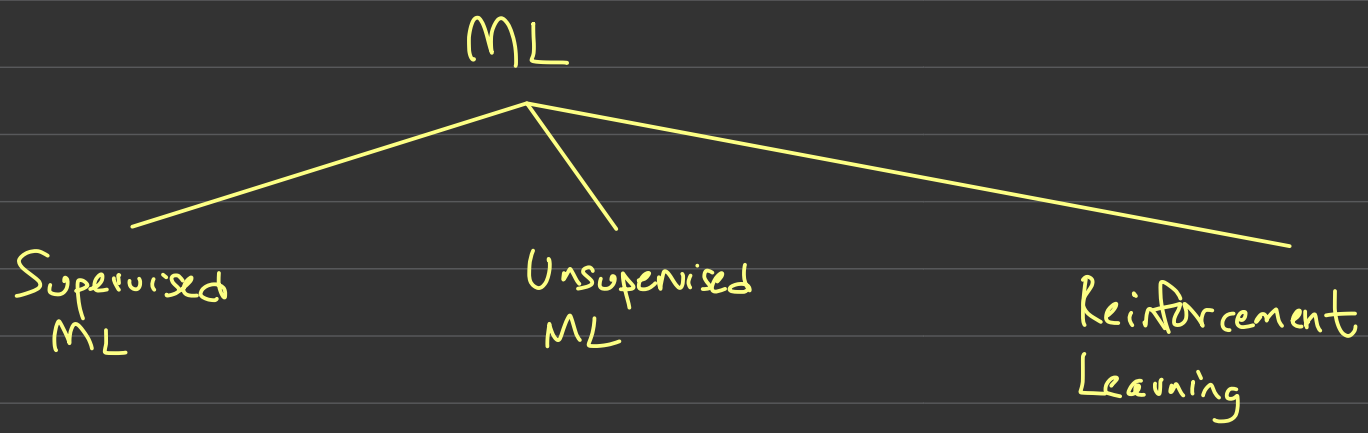
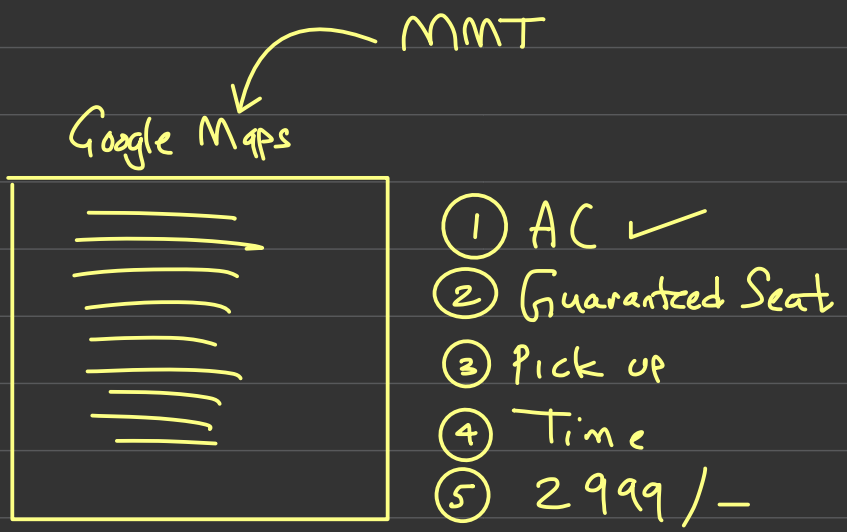
 # Agenda: AIML

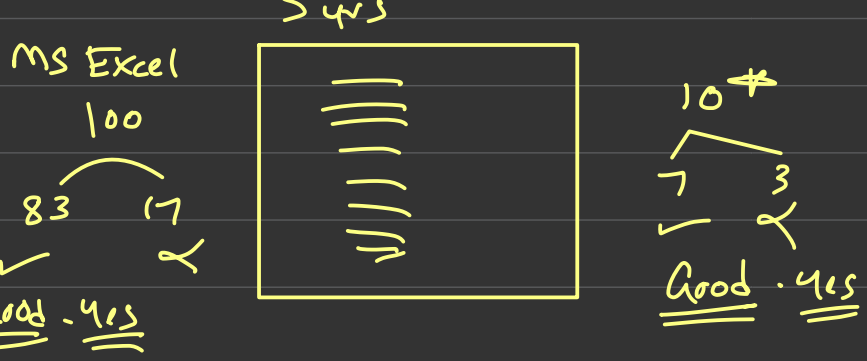
Introduction to ML
Types of ML
Linear Regression

ML — predictive analytics



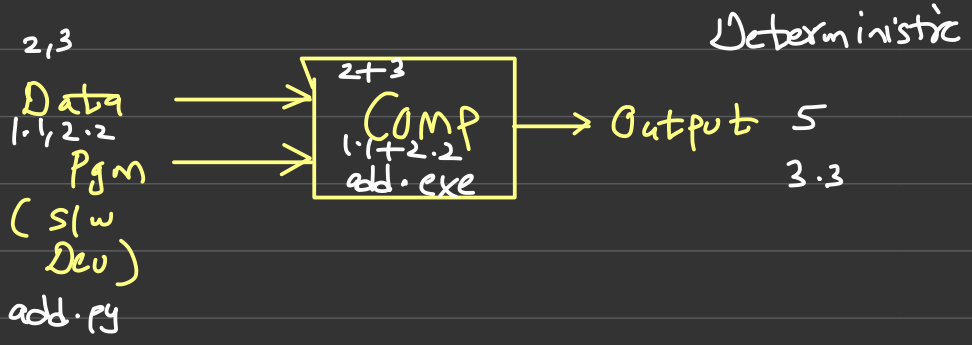


Will you lend me Rs. 5000/- ?

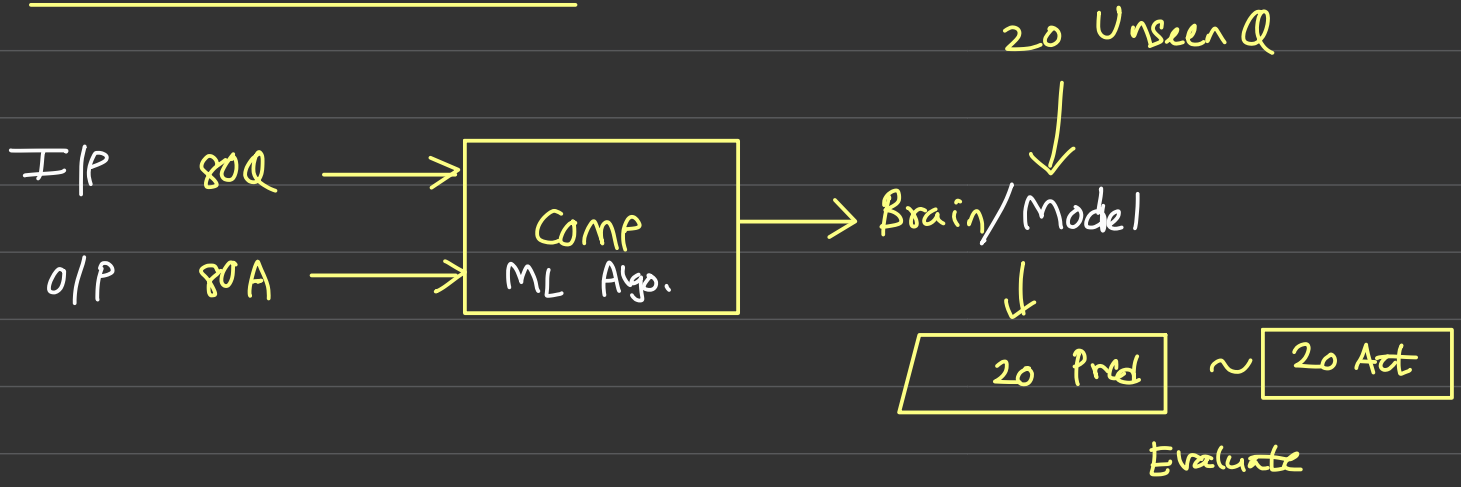


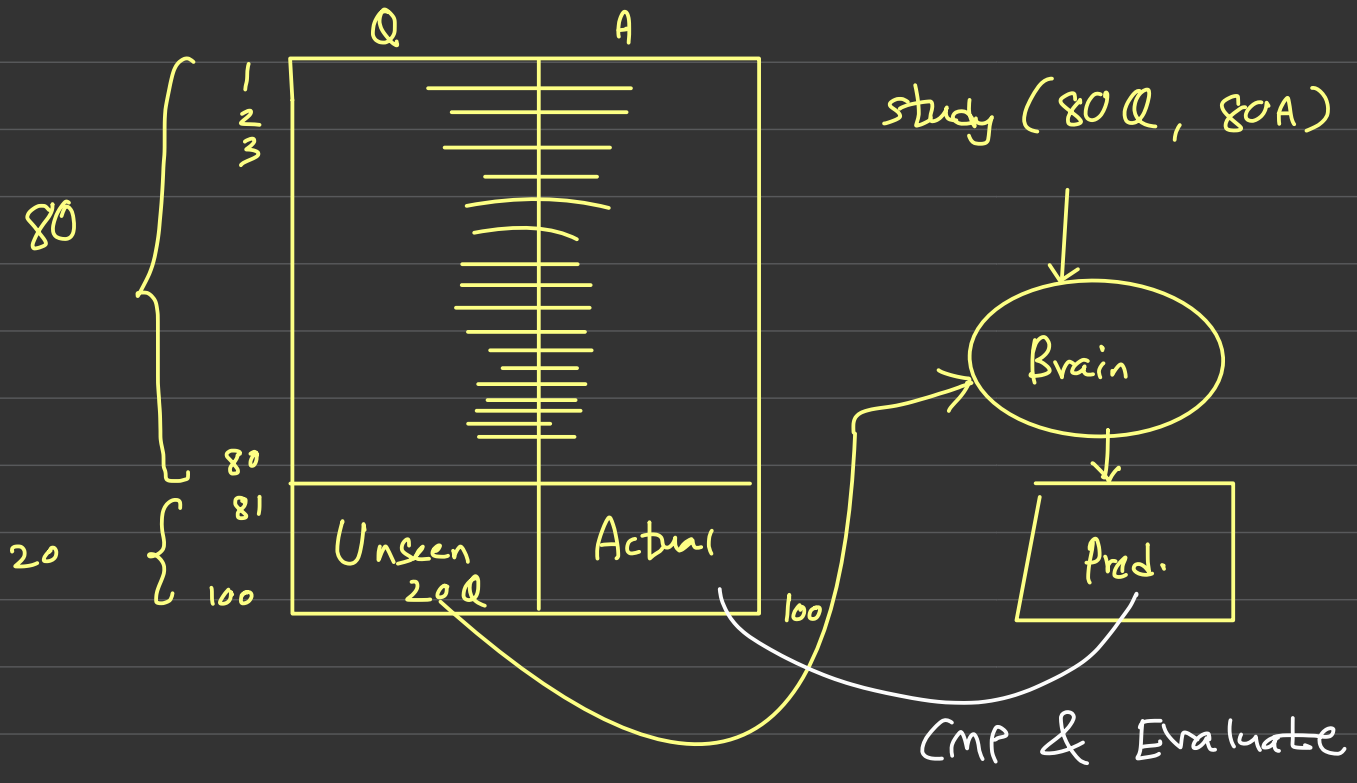
Traditional Programming Paradigm

HDFC



ML Prog. Paradigm





Dependent variable — Independent variables

① Price of House ~ Sqft area + Location + Gym + Pkg + Audi + Club + Pool + # Bathrooms + # Floor + Hospital + School + Mkt + Transportn.

$y \sim X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7$

② Price of Phone ~ Brand + Battery mA + RAM + ROM + Camera + Display + Fastch. + Screen Size + Processor

$+ X_8 + X_9$

③ Apple vs Orange ~ Color + Taste + Shape + Texture + Weight + Smell

④ Iris S vs Ver vs Vir ~ PL + PW + SL + SW

ML

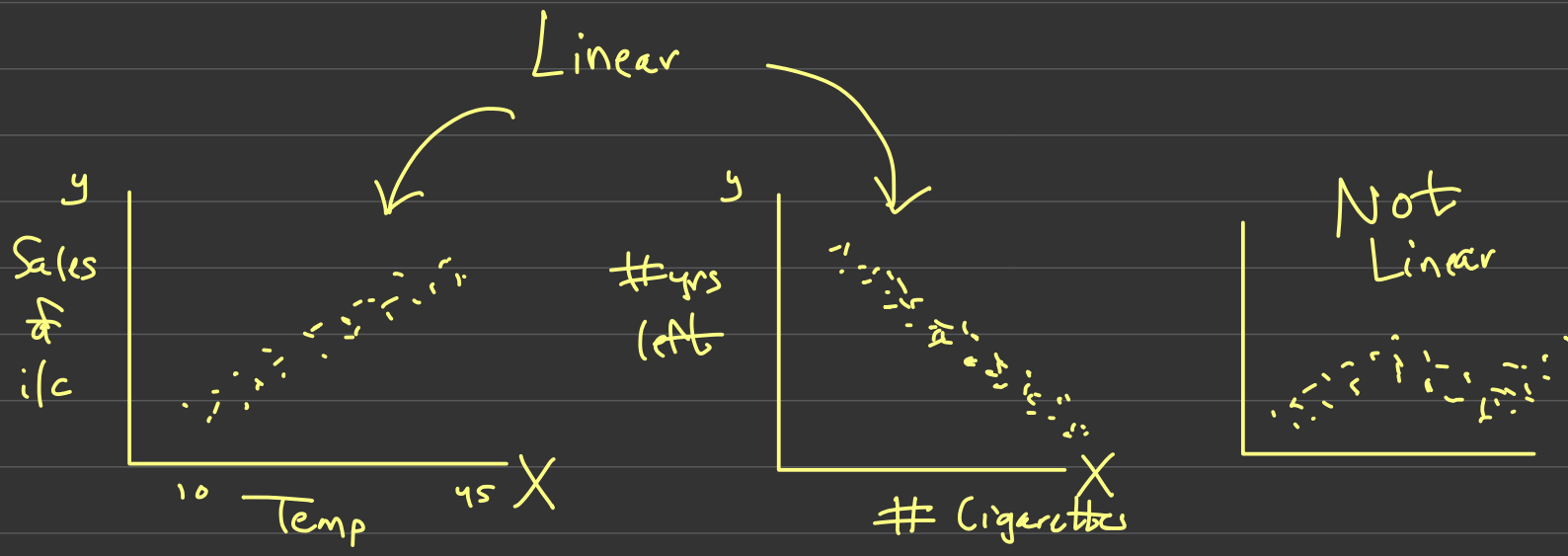
① Supervised ML — labeled data

$X \sim y$

- Regression — y – Continuous — eg: ① & ②
- Classification — y – Categorical — eg: ③ & ④

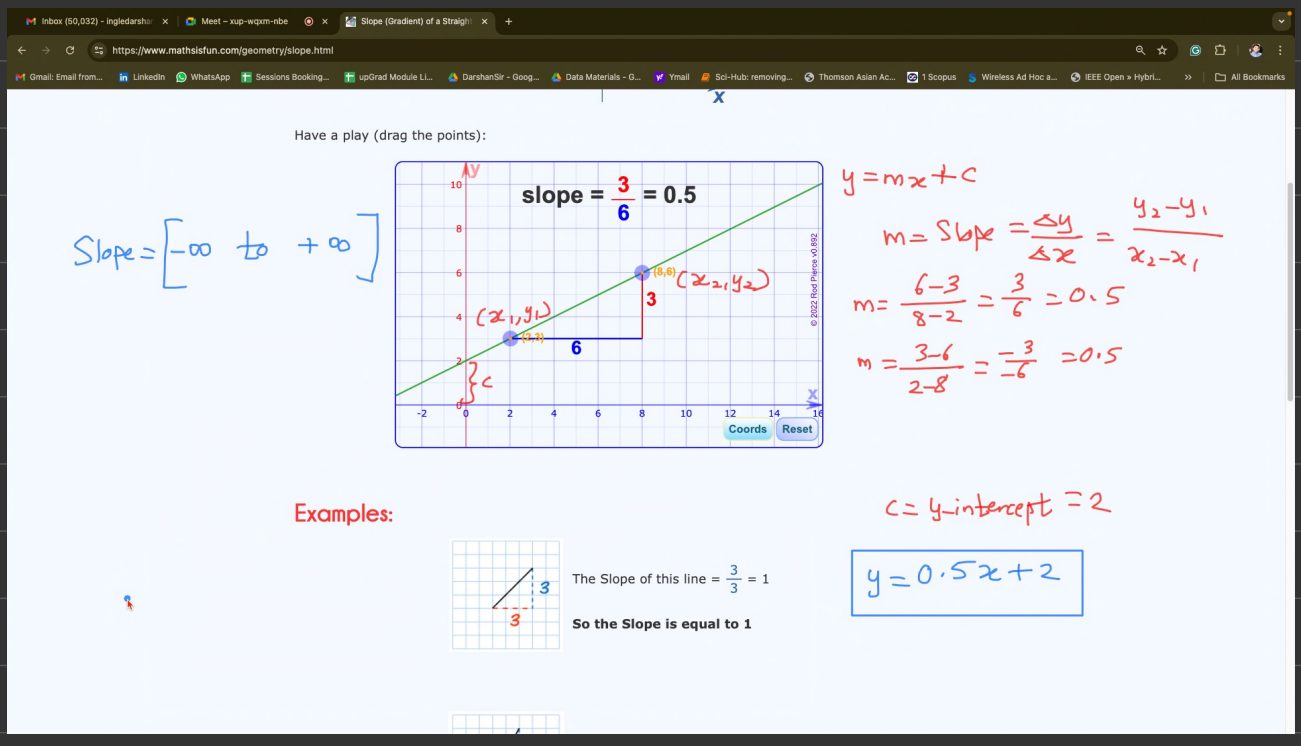
Linear Regression: Regression

X ~ y



https://www.linkedin.com/posts/drdarshaningle-instructor_i-bet-u-still-dont-know-this-activity-7168450493587951616-866h?utm_source=share&utm_medium=member_desktop

https://www.linkedin.com/posts/drdarshaningle-instructor_want-to-know-choosing-the-right-regression-activity-7159395083413274624-8DN3?utm_source=share&utm_medium=member_desktop



Slope = [−∞ to +∞]

$y = mx + c$

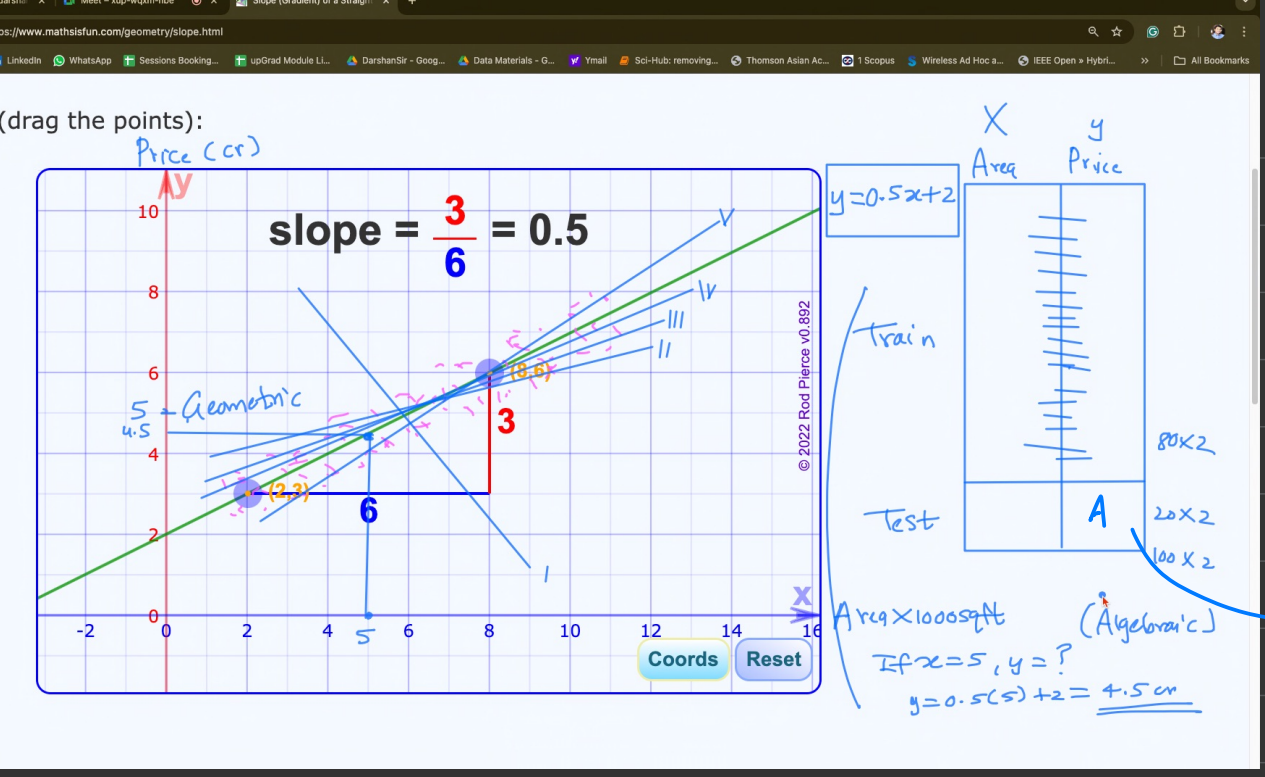
$m = \text{Slope} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$

$m = \frac{6-3}{8-2} = \frac{3}{6} = 0.5$

$m = \frac{3-6}{2-8} = \frac{-3}{-6} = 0.5$

c = y-intercept = 2

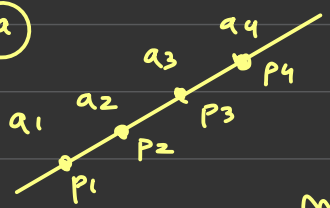
$y = 0.5x + 2$



Regression Metrics:

$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (a_i - p_i)}{n}$$

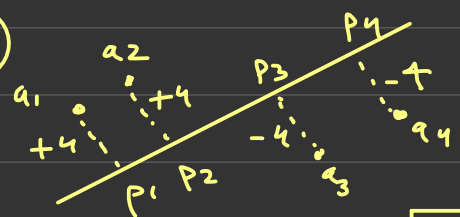
(a)



$$ME = \frac{(a_1-p_1)+(a_2-p_2)+(a_3-p_3)+(a_4-p_4)}{4}$$

$$= \frac{0+0+0+0}{4} = 0$$

(b)



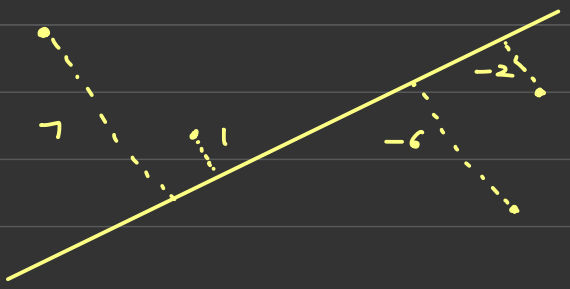
$$ME = \frac{+4+4+(-4)+(-4)}{4} = 0$$

$$MAE = \frac{\sum_{i=1}^{n} |a_i - p_i|}{n}$$

(a) $MAE = \frac{|0|+|0|+|0|+|0|}{4} = 0$

(b) $MAE = \frac{|+4|+|+4|+|-4|+|-4|}{4} = \frac{16}{4} = 4$

(c)



$$MAE = \frac{|7|+|1|+|-6|+|-2|}{4} = 4$$

$$MSE = \frac{\sum_{i=1}^{n} (a_i - p_i)^2}{n}$$

(b) $MSE = \frac{4^2+4^2+(-4)^2+(-4)^2}{4} = \frac{64}{4} = 16$

(c) $MSE = \frac{7^2+1^2+(-6)^2+(-2)^2}{4} = \frac{90}{4} = 22.5$

$$RMSE = \sqrt{MSE}$$

(b) $RMSE = \sqrt{16} = 4$

(c) $RMSE = \sqrt{22.5} = 4.734$

R^2:

It is called as Coefficient of Determination
Its value is in the range [0-1], higher the better.
Industry Accepted value: min 0.8

It quantifies the proportion of variance caused in the DV due to the IV.

X ~ y =: Temp ~ Sales
If R2=0.8
It means that Temp is causing 80% variance in the sales of icecream i.e. it is 80% important for predicting the sales of icecream.

Good Model :-

① Test Acc $\geq 80\%$.

② Train Acc $-$ Test Acc $\leq 5\%$.

Train Acc	Test Acc	Scenario
100	98	✓
100	95	✓
100	93	Overfitting
85	80	✓
83	79	Need to work (Close to good)
75	70	Underfitting
85	87