@) Supervised vs unsupervised.
↓   ↓
labelled dataset   unlabelled
↓
i.e some data
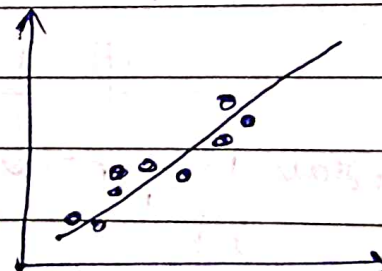is already tagged
with correct
ans.

Regression : predict continuous outcome
variable (y) based on the value
of one or multiple predictor variables (x)
Linear Regression :- it is used for
establishing linear relationship b/w
the i/p val/pred val (x) with outcome
var (y).
     indep.   target
                              dep

→ predicts a dep var (y) based on a given
indep · val (x)

(coeff of)
q_x

$$y = \theta_1 + \theta_2 x$$ (univariate)

constant   ↓
        intercept

supervised

So best fit line is predicted
with the the best θ_1 & θ_2 values
To update θ_1 & θ_2 to get best fit line

1) Cost function : By achieving best fit
       (J)   regression line

model aims to predict y value such that
the error diff b/w predicted & true val
is min, so we need to update θ_1 & θ_2

$$\min \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$\therefore \boxed{J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2}$$

[Root Mean Squared error]

Gradient Descent :- In order to to min cost-f^n θ,
to update θ_1 & θ_2 & achieve best fit line,
the model uses gradient descent, the
idea is to randomly choose θ_1 & θ_2
values & then iteratively updating the
values, reaching min cost.

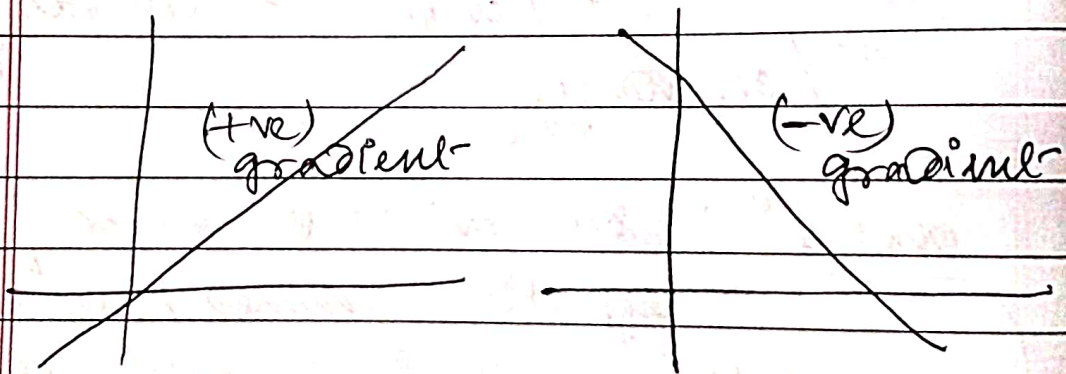Disadv : Overfitting prob., outliers
from sklearn. linear_model → LinearRegre
ssion

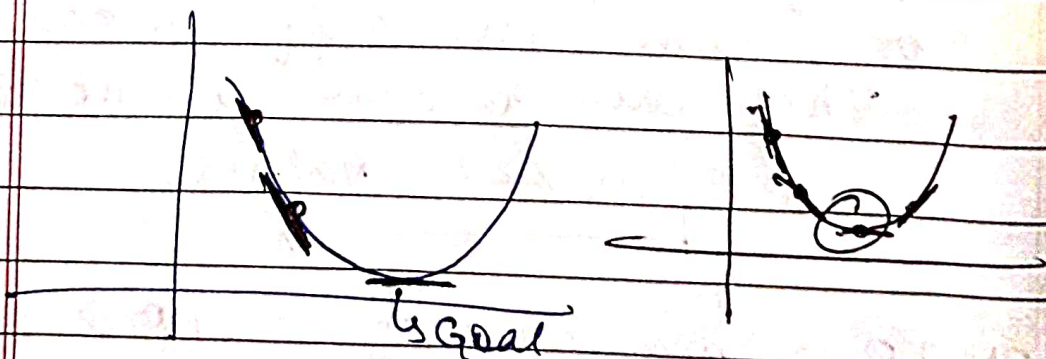This minimising technique to get the best m & c values is done by "Gradient descent" (GD)

↓

optimisation technique
(thus min. cost $f(x)$)

Gradient : rate of inclination/declination of a slope
(i.e how steep a slope is & in which dir^n it is going)

Descent = act of moving downwards

(+ve) gradient   (-ve) gradient

At the start the m & c values are randomly selected /initialized /can be 0.

↳ GOAL

Hyperparameters : parameters which determine the model's performance (or how well we are able to train it)

→ These are the parameters outside of the model to (tune) the model so that it fits the training data for the given hypothesis
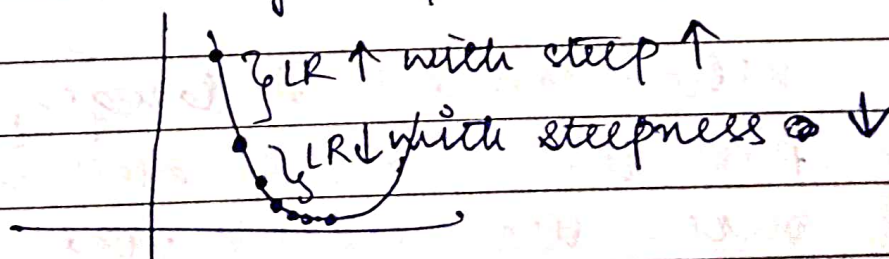
3 imp hyperparameters :-

- Learning Rate (LR)
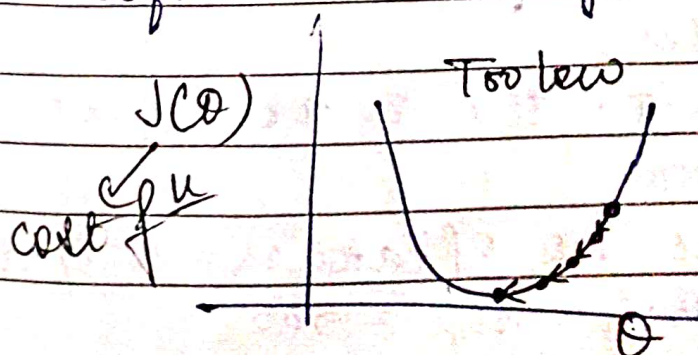- Epoch
- Batch Size

LR : help in getting optimal weights

→ It is the rate of ↑gd ↓gi.e updating the values of parameters (m,c) during GD.

initially steps will be bigger

{LR ↑ with steep ↑
{LR↓ with steepness ∞ ↓

Note :

A small LR requires many updates before reaching the min pt.

J(θ)

Too low

& too high takes drastic changes updates & can lead to divergent behav

cost ↑

θ

(Too high)