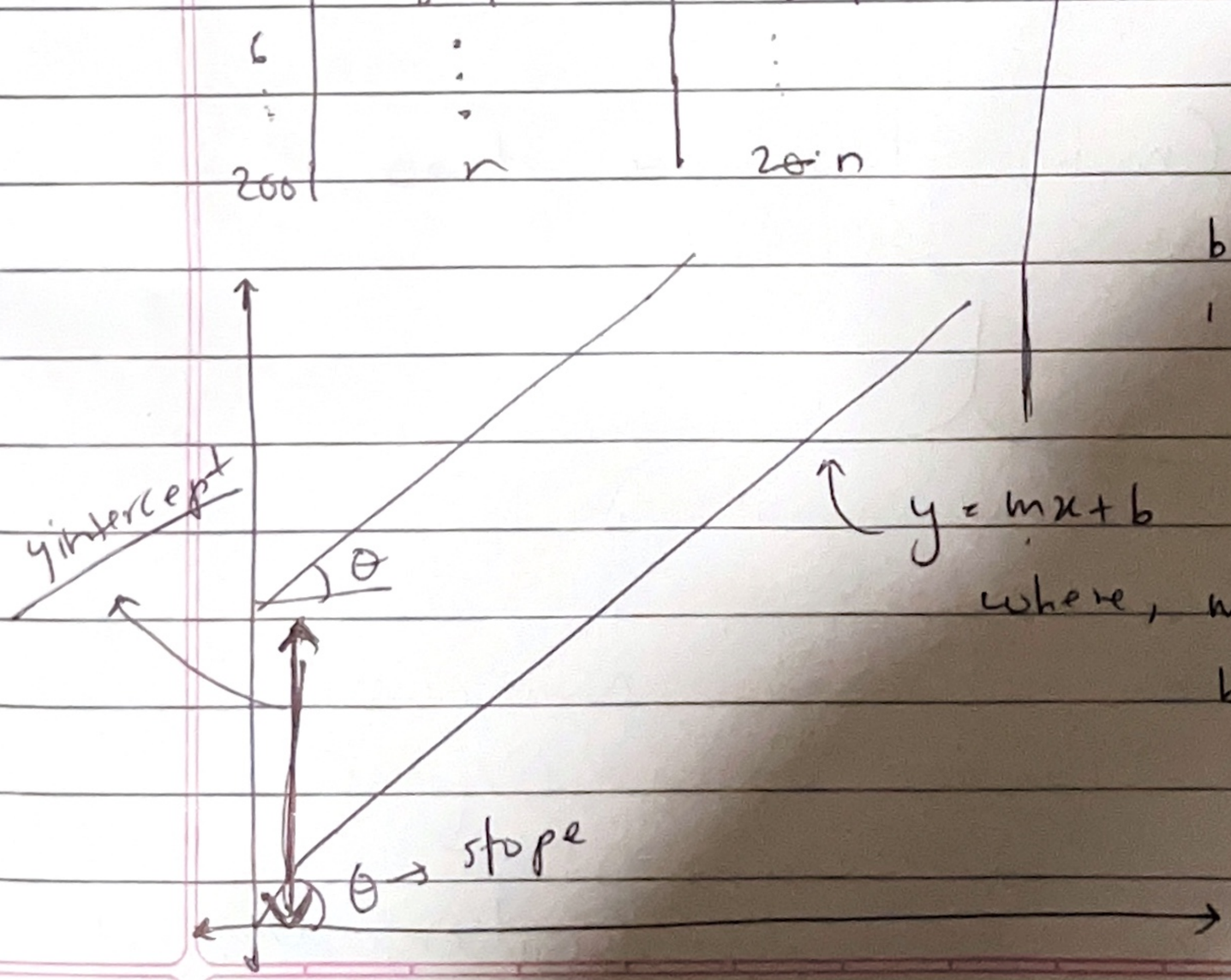
◦ Simple Linear Regression ◦

Sr No	cgpa	Package
1	6.66	3.01
2	7.1	3.5
3	4.7	1.2
4	8.9	4.2
5	8.1	3.9
6	⋮	⋮
⋮	⋮	⋮
200	n	20 n



$y = mx + b$

where, m = slope

b = yintercept

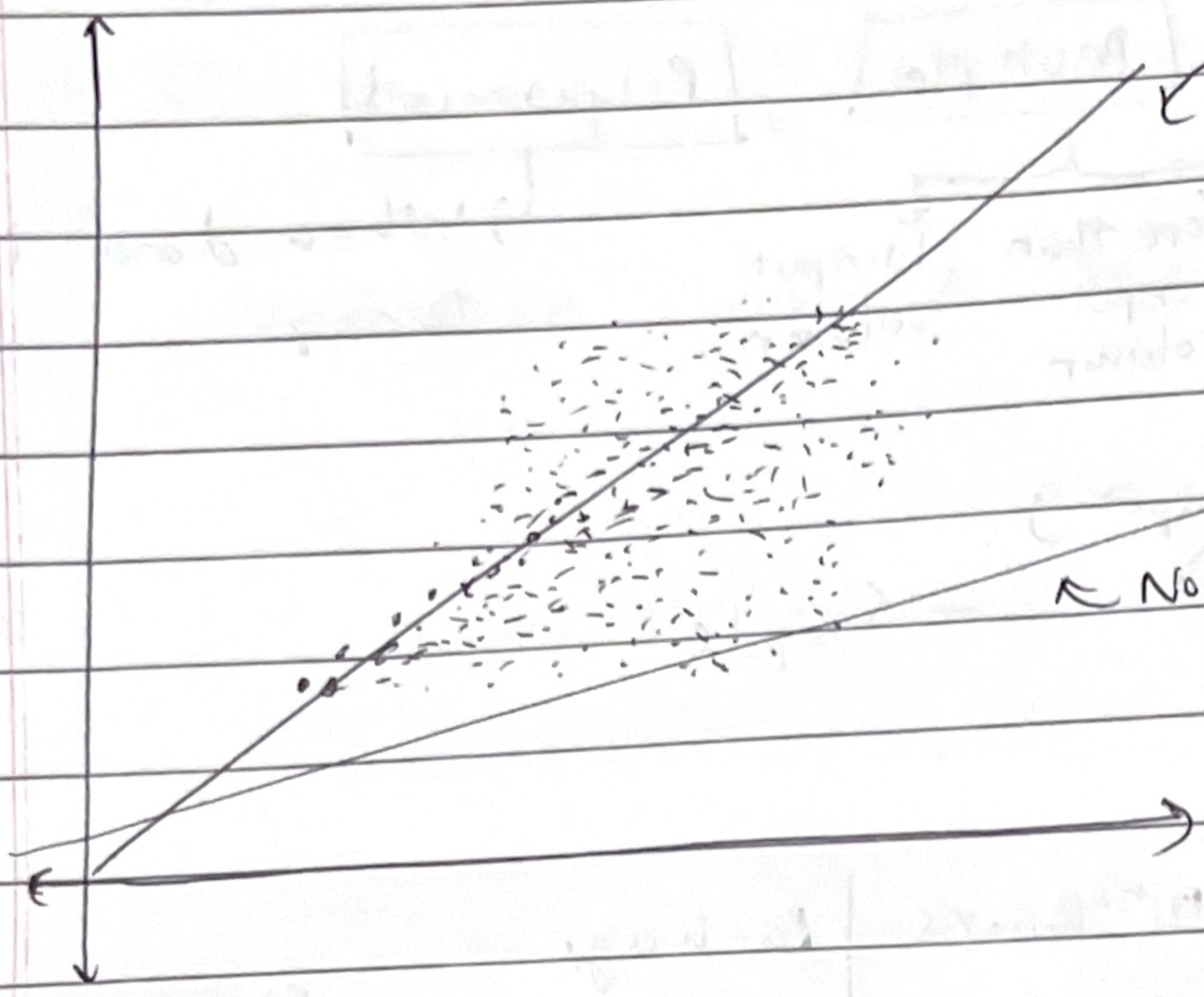
① Plot the dataset

Que. Why dataset in not linear ?

→ • Real world dataset it is.

• The factors which cannot be determined are called 'Stochastic Errors'.

Eg कितना खराब गया Interview ?



- Steps to work on any dataset:-

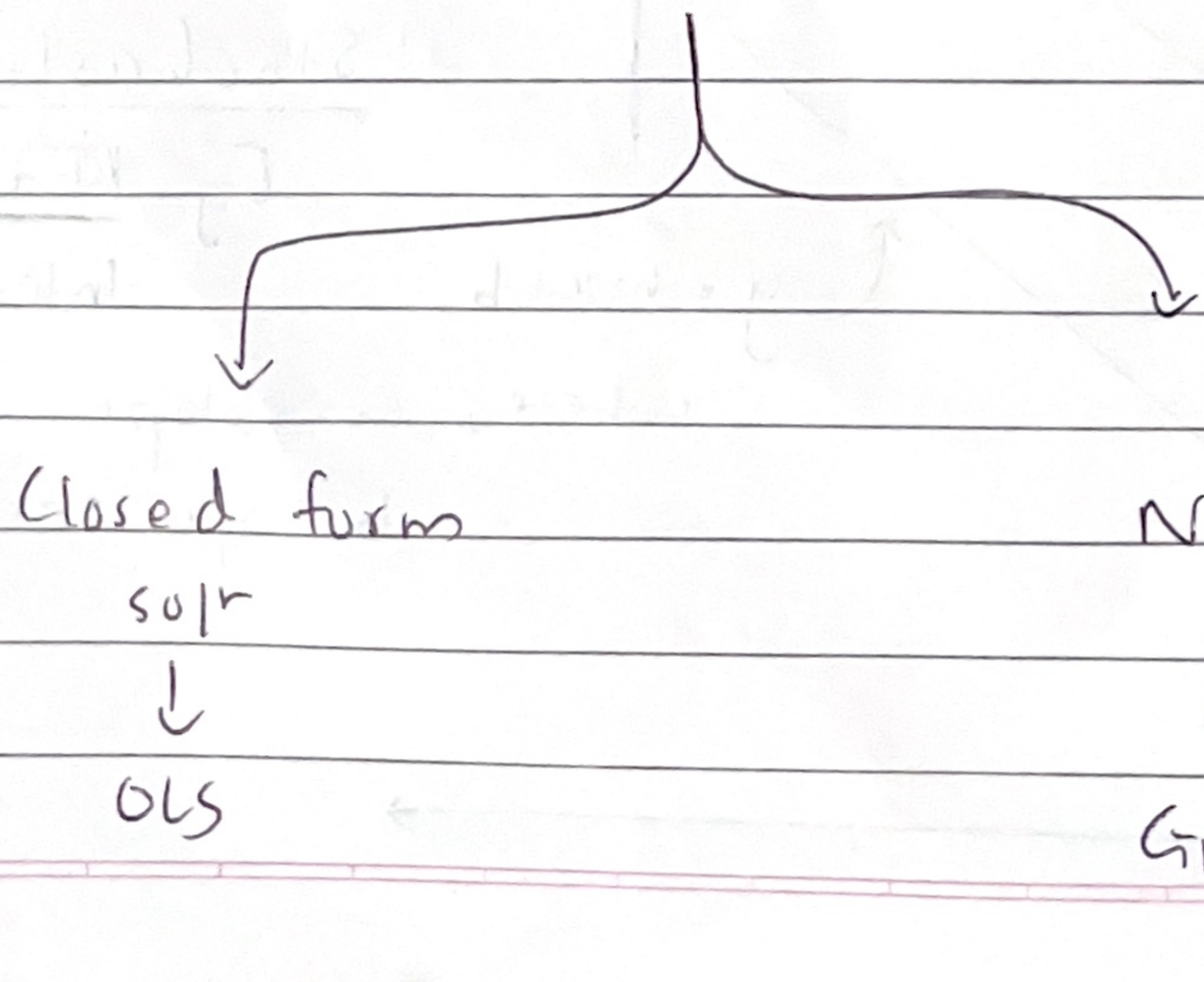
(1) Plot the dataset and analyse.
(2) Seperate x and y columns.
(3) Divide data in train and test

X_train	X_test
y_train	y_test

(4) Import linear regression model.
(5) Use fit to train data
(6) Now test the model.

Que
What is input shape?
↳ Size of dataset.
Eg. 64x64
or
29x2

To find (m,b) there are two methods:



(Direct formula)

Closed form solution → +, −, ×, ÷ use karke jis bhi chiz ka formula bun sakta hai but no differentiation and integration.

Eg. quadratic eqns.

↳ OLS - Ordinary least squares

Non-closed form sol^n ⇒ Ek approximation method use karke (with using differentiation & integration) (if needed) sol^n dhundhna.

↳ Gradient Descent.

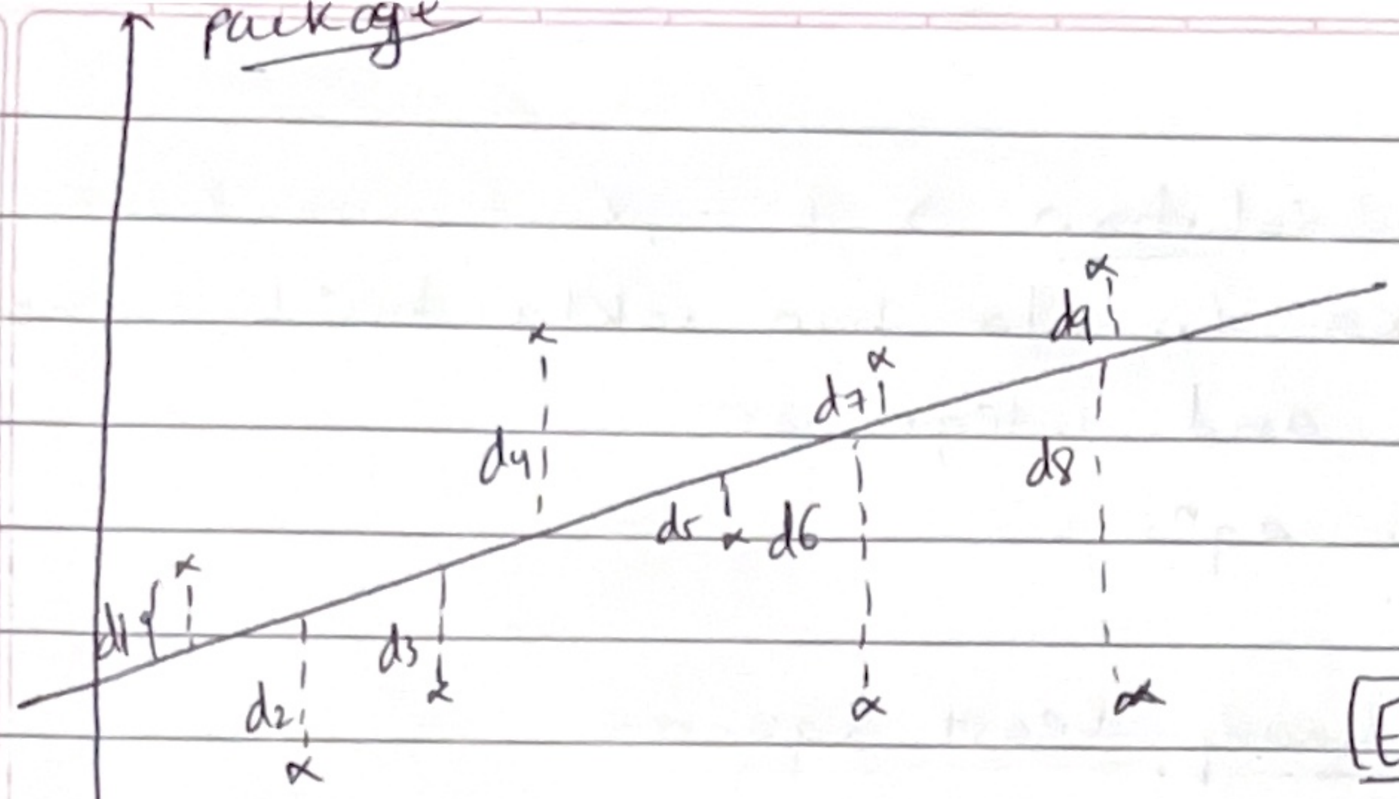
Que. If we have a fixed formula to find answer than why use GD?

→ In higher dimensions, using a formula is not enough to find the solution.

OLS (Ordinary least squares) $\Rightarrow \boxed{b = \bar{y} - m\bar{x}}$

Slope → (m) = $$\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $\bar{x}$ and $\bar{y}$ are means



Now to find total distance / error,

$$E = d_1 + d_2 + \dots d_n$$

What is penalizing?

→ A technique i.e used for reducing (shrinking a large amount of features to a manageable set and for making good predictions in a variety of large data sets.

OR

Giving some weightage for deviating from the objective.

Now, use square the distances and then add

$$E = (d_1)^2 + (d_2)^2 + \dots (d_n)^2$$

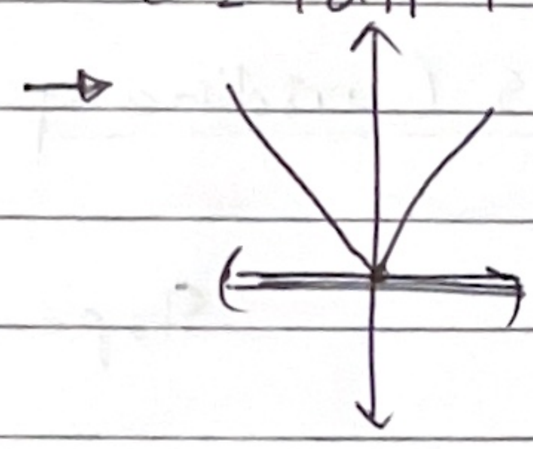
REASON: To cancel out positive and negative error

Therefore, our Error function looks like this,

$$E = \sum_{i=0}^{n} (d_i)^2$$ ← loss function

Que. Why not mod the error?

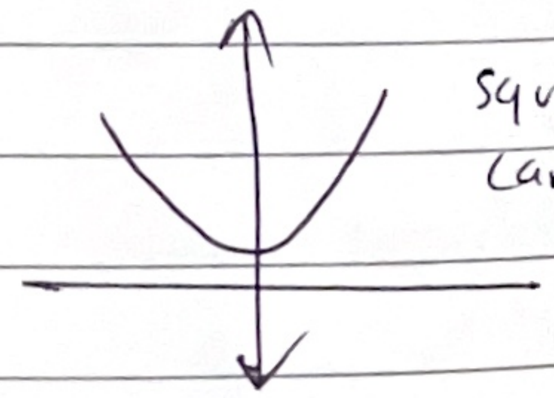
$$E = |d_1| + |d_2| + \dots |d_n| ?$$



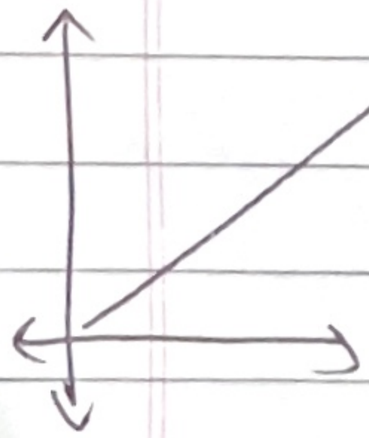
(We cannot mod differentiate mod values at zero)

NOTE: Also represented as 'J'

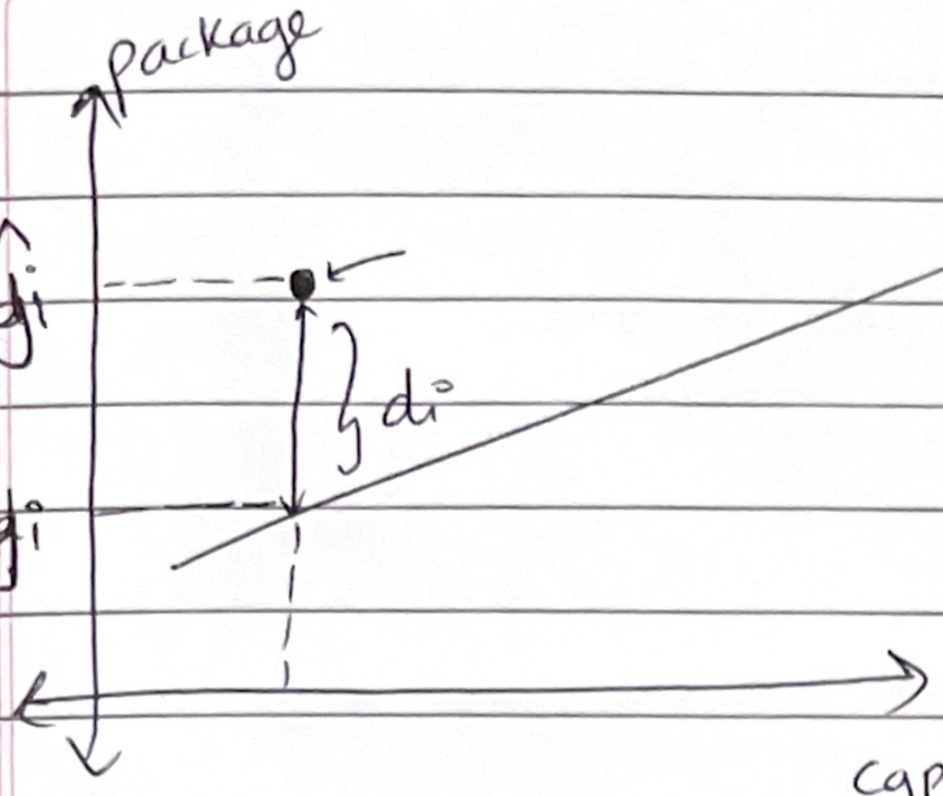
whereas,



Squared values can be differentiated



← This predicted line is represented by $\hat{}$ (HAT)



So,

(Error for one student)

$$d_i = (y_i - \hat{y_i})$$

y_i → Original Package

$\hat{y_i}$ → Predicted package

Hence,

(Loss / Error function)

$$E|J = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

i.e

$$E|J = \sum_{i=1}^{n} (d_i)^2$$

Also,

$$E|J \text{ or } E(m, b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$