Date
Page

Hypothesis and Cost-function of Linear Regression

Hypothesis

$h_\theta(x) = \theta_0 + \theta_1 x_1$ — θ_0 comes first = intercept (up and down)

θ_1 multiplies X = slope (flatter / steeper)

imp

Cost-function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x) - y)^2$$

1. **Hypothesis Function**

$h\theta(x) = \theta_0 + \theta_1 x$ — this is the prediction line

$y = b + mx$

- $h\theta(x)$ = your predicted value $\hat{Y}$ same as $\hat{Y} = mX + b$
- θ_0 = y-intercept (same a b)
- θ_1 = slope (same as m)
- X = input value

So $h\theta(X) = \theta_0 + \theta_1 X$ is litterally same as $y = mX + b$

2. **Cost function**

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x) - y)^2$$

This measures how wrong your line is:

- $(h\theta(x) - y)^2$ - difference between your prediction and actual value (the error)

this is same as Error = Actual - Predicted

Error = $Y - \hat{Y}$

error = +5, $(+5)^2 \Rightarrow 25$

error = -5, $(-5)^2 \Rightarrow 25$

In the cost-function they wrote it Predicted - Actual value = $\hat{Y} - Y$

- $\sum_{i=1}^{m}$ = sum of all the squared errors for all your data points

- $\frac{1}{2m}$ = take the average and divide by 2 (the 2 makes the calculus easier later)

In simple terms : the cost function tells you "how bad is my current line" A high cost = bad line , low cost = good line

The Goal :
Find the values of θ_0 and θ_1 that make $J(\theta_0, \theta_1)$ as small as possible . When cost is minimized , you have the best fitting line .

↓

Method 1 : Use the Direct Formulas
For θ_1 (slope) :

$$\theta_1 = \Sigma[(X_i - \bar{X})(Y_i - \bar{Y})] / \Sigma[(X_i - \bar{X})^2]$$

For θ_0 (intercept) :

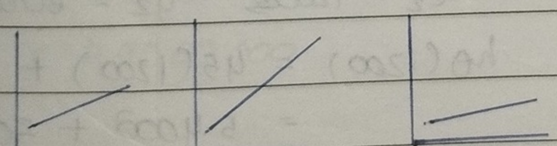
$$\theta_0 = \bar{y} - \theta_1 \cdot \bar{x}$$

This gives you the exact best answer immediately using math

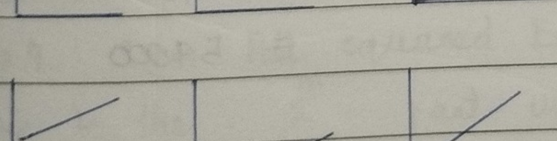
Method 2 : Gradient Descent (The ML Way)

– x – x – x — x — x — x — x — x — x — x — x — x — x —

$y = mx + b$ m :- controls the slope
b :- intercept of the line



This is m - controlling the slope (making line steeper or flatter) :- θ_1



This b - controls the intercept of the line (moving line up and down) :- θ_0

Setup:

- Total data points : 10
- Test split = 0.2 means 20% testing, 80% training
- Training points : 10 x 0.8 = 8 points
- Test points : 10 x 0.2 = 2 points

Our Training Data (8 points)

Predicting house price from size:

Point	size (x)	Actual Price (y)
1	1000	\$ 50,000
2	1200	\$ 60,000
3	1400	\$ 70,000
4	1600	\$ 80,000
5	1800	\$ 90,000
6	2000	\$ 100,000
7	2200	\$ 110,000
8	2400	\$ 120,000

So $m = 8$ (our training size)

Our current line: $h\theta(x) = 45x + 5000$

1: Calculate Predictions for Each point

Point 1:

$x_1 = 1000$, $y_1 = 50000$

$h\theta(1000) = 45(1000) + 5000$

$= 45000 + 5000$

$= 50000$

Point 2:

$x_2 = 1200$ $y_2 = 60000$

$h_\theta(1200) = 45(1200) + 5000$

$= 54000 + 5000$

$= 59000$

Point 3:

$x_3 = 1400$ $y_3 = 70000$

$h_\theta(1400) = 45(1400) + 5000$

$= 68000$

Point 4 : 77000

Point 5 : 86000

Point 6 : 95000

Point 7 : 104000

Point 8 : 113 000

Step 2: Calculate Error for Each point

Error = Predicted - Actual

= $h_\theta(x) - y$

$h_\theta(x) \Rightarrow$ Predicted value

$y \Rightarrow$ Actual Value

Point 1 : 50,000 - 50000 = 0

2 : 59000 - 60000 = -1000

3 : 68000 - 70000 = -2000

Point 4 : 77000 - 80000 = -3000

5 : 86000 - 90000 = -4000

6 : 95000 - 100000 = -5000

7 : 104000 - 110000 = -6000

8 : 113 - 120000 = -7000

Step 3: Square Each Error ~~for Each Point~~

$(h_\theta(x) - y)^2$

Point : 1 = 0

2 = 1000000

3 = 4000000

4 = 9000000

5 = 16000000

6 = 25000000

7 = 36000000

8 = 49000000

Step 4 : Sum All squared Errors

This is the $\sum_{i=1}^{m}$ part where $m = 8$

Sum = 0 + 1000000 + 4000000 + 9000000 + 16000000 + 25000000 + 36000000 + 49000000

Total sum = 140000000

imp { $1/2m \times$ predicted - actual $\rightarrow$ cost function J
$1/m \times$ predicted - actual $\rightarrow$ MSE }
$\sqrt{MSE}$ = RMSE

Step 5: Apply the $\frac{1}{2m}$ formula

$RMSE = \sqrt{MSE}$

m = 8 (num of training points)

$2m = 2 \times 8 = 16$

$MSE = \frac{1}{m} \times (\text{predicted} - \text{actual})^2$

$\frac{1}{2m} = \frac{1}{16}$

Final Cost : $J = (1/16) \times 140000000 = 8750000$
$\hookrightarrow (\text{predicted} - \text{actual})^2$

Comparison with simple error :

- old way : Just average error = Total error ÷ num of points (training points)
- ML way : Average error ÷ 2 (the ÷ 2 make calculus easier)

What this value means ?

1. How "Wrong" your line is :
- Cost = 8750000 means your line has an average error of around \$2958 per house
- Because $\sqrt{8750000} \approx \cancel{38}$ 2958 - this is called RMSE

2. Is this Good or Bad ?
- Depends on House prices !
- If houses cost \$100000 on average, the \$2958 error = pretty good (3% error)
- If house cost \$50000 on average, the \$2958 error = not so good (6% error)

3. Comparison Tool
Line A : Cost = 8750000
Line B : Cost = 1500000
Line A is better (lower cost = better fit)

4. When to stop:

- If cost keeps going down → keep training
- If cost stops changing → you found the best line.

- Multiple Linear Regression

Matrix form

$\hat{y} = X \cdot \theta$

X = design matrix $X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \end{bmatrix}$

θ = parameter vector

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$

y = output vector $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

Finding the best Parameter

1. Normal Equation

$\theta = (X^T X)^{-1} X^T y$

2. Gradient Descent

$\theta = \theta - \alpha \cdot \frac{1}{m} X^T (X\theta - y)$

Convergence theorem

$\{$

$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$

$\}$

classmate

Date

Page

multiple linear regression formula :

$\Rightarrow y = m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_nx_n + c$

Linear Regression - Indepth Maths Intution

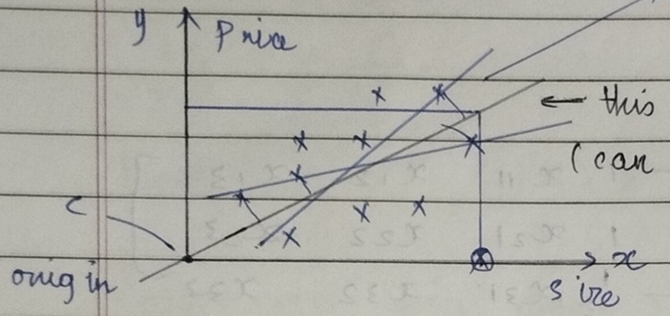
$y = mx + c$ — Best fit line (Single linear regression formula)

m = slope

c = intercept

★ sumation of all this error should be minimal

★ i should minimize the error in this



← this is the best fit line (can be used to determine the price)

house

$\Rightarrow$ With respect to the size of the house what would be the price

(Independent Variable - Input) (Dependent Variable) output

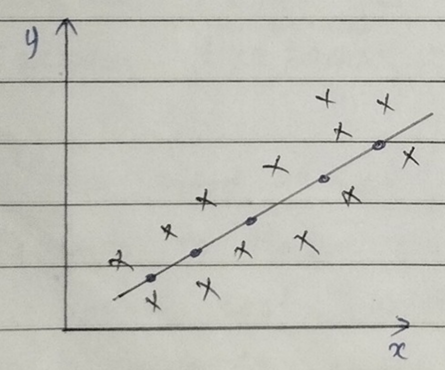
Whenever the size is → 0

↳ x is 0

if $x = 0$, equation would be $y = mx + c$

$y = m \times 0 + c$

$y = c \Rightarrow$ intercept



cost function

$$= \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$$

(we should try to minimize this error)

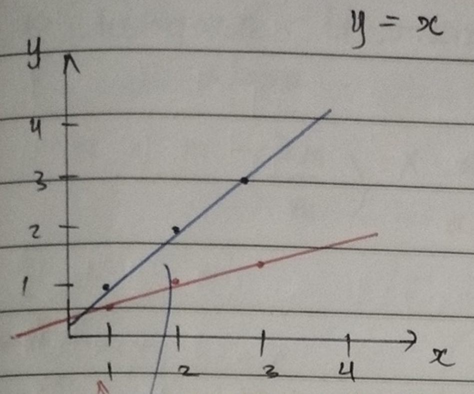
less error = best fit line

m = number of training points

$\hat{y}$ = predicted value (x)

$\hat{y}$ = Actual value (•)

Date
Page



this is real data point ⇒ find best fit line

$\hat{y} = mx + c$

for eg if $c = 0$

for $x = 1$, let's consider

$m = 1$ m :- slope

plotting this points in the graph

$\hat{y} = mx + c$

$x = 1$ 1.) $\hat{y} = 1(1) + 0 = 1$

$x = 2$ 2.) $\hat{y} = 1(2) + 0 = 2$

$x = 3$ 3.) $\hat{y} = 1(3) + 0 = 3$

→ This is the best-fit line :- when $m = 1$

· After this we'll find the cost-function ∴ m = no of data points = 3

$$\text{cost function} = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$$

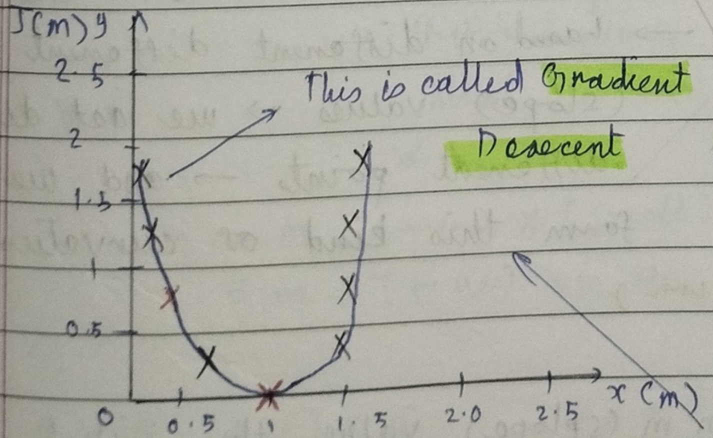
$\hat{y}$ = predicted

y = actual

$$= \frac{1}{2(3)} \left((1-1)^2 + (2-2)^2 + (3-3)^2 \right)$$

$$= \frac{1}{6} \times 0 = 0$$

⇒☆ When m → slope = 1 the cost function ~~is zero~~ as 0



In the next step we'll change the value of $m = 0.5$

⇒ $x_1 = 0.5$, $x_2 = 1$, $x_3 = 1.5$

0.58

cost function for $m = 0.5$

similarly for different values of cost function and m values we we'll be getting points which will form this kind of curvature

If slope > 0 (positive slope) → decrease the value to reach global minima

If slope < 0 (negative slope) → increase the value to reach

If $m = 0.5$ $\quad \hat{y} = mx + c$

Input :- 1, 2, 3 $\quad = 0.5(1) + 0$

$= 0.5$

$= 0.5(2) + 0$

$= 1$

$= 0.5(3) + 0$

$= 1.5$

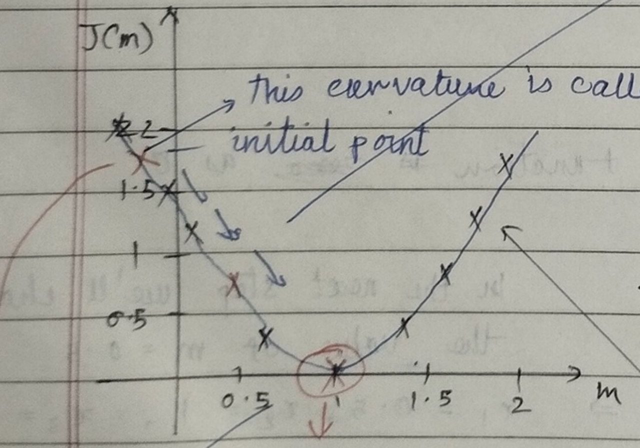
$$\frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$$

Cost function for $m = 1.5$

$$CF = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$$

$$= \frac{1}{2(3)} ((0.5-1)^2 + (1-2)^2 + (1.5-3)^2)$$

$= 0.58$



in order to move downwards we write :- convergence theorem

→ When should we stop find values of m ? for best fit line ?

→ based on different different m (slope) values → we get different different points → and we can form this kind of curvature

(How do we arrive to this particular region)

This is called global minima

→ considering based on m (slope) value the initial point is somewhere overhere

+ve :- $m := m - \alpha(+ve)$ ⇒ This will subtract (decrease)

-ve :- $m := m - \alpha(-ve)$ ⇒ This will addition (increases)

Convergence theorem

{

$$m = m - \left(\frac{\partial m}{m}\right) \times \alpha$$

slope

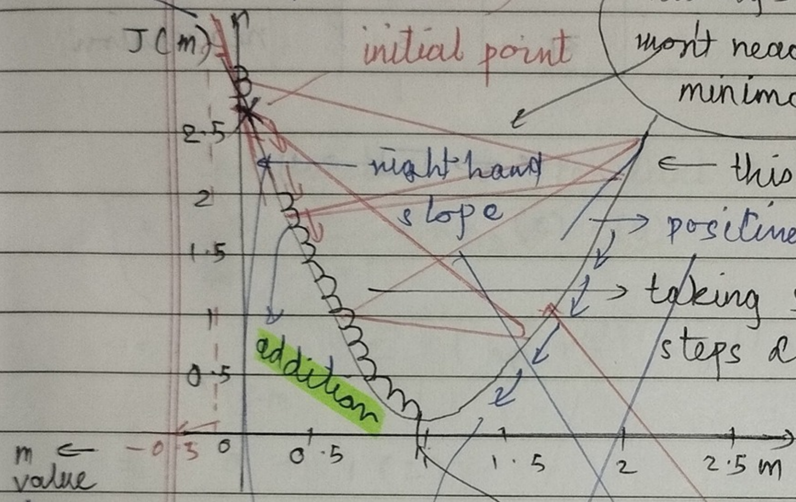
this is learning rate

for this reason always select lower learning rate $\alpha = 0.001$

applying this theorem

}

left hand slope



initial point

This big steps won't reach global minima $\alpha = 1.00$

multiple iteraction

right hand slope

← this is gradient descent

positive slope

taking small steps $\alpha = 0.001$

negative slope derivative → we got negative value also we need to select α learning rate smaller value $\alpha = 0.001$

This is global minima

This will subtract

How to find this +ve or -ve slope

This is pointing downwards this a - negative slope

If it is having a negative slope!

What we need to do is :

$$m = m - (-ve) \times (\text{small})$$
$$= m + (+ve) \text{ smaller}$$

If we change this value to (1.00) this is don't take small steps instead it will directly jump

$m = m - (+ve) \times \alpha$ $\quad m = \theta_1$

$= m -$

Positive slope :- $\theta_1 = \theta_1 - \alpha(+ve)$

Negative slope :- $\theta_1 = \theta_1 - \alpha(-ve)$