19/3/21 Lean In ML Circle

Linear Regression Model

- Mathematical Understanding
- Sklearn implementation

Terms

1) Dependent variable
2) Independent variable

Example:- Annual Sales of a company.

depends on:- No. of employees, Profit, Product sale } Independent variable

~~And~~ ~~as~~ as annual sales depends on these factors, it is Dependent variable

Factors affecting annual sales variable are your independent variable.

* When we try to establish a relationship between dependent and independent variable. that is called Regression ~~ana~~ analysis.

Regression → Linear Regression
→ Logistic Regression.

Linear Regression

Why Linear?
The effect that independent variable has on dependent varible is proportional.

Formal Definition.

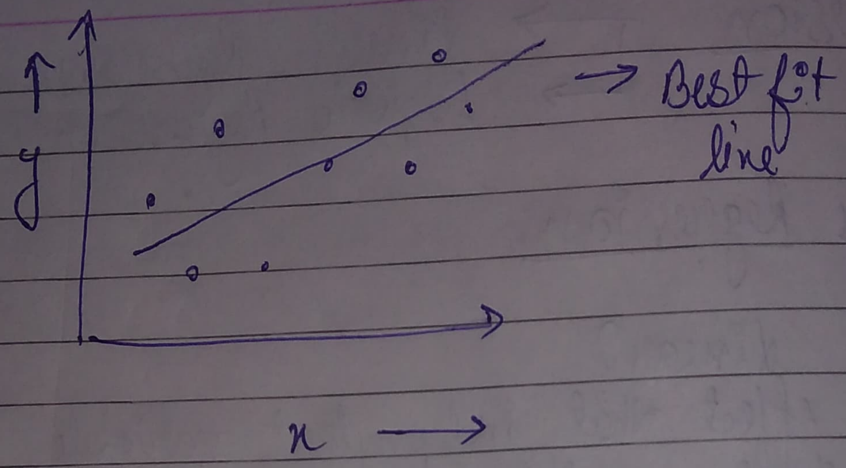
It is a technique which is used to model and analyse the relationship ~~att~~ between variables and often times how they contribute and are related to a particular outcome.

Nature of independent :- Numeric and continuous

Difference b/w discrete and continuous

Countable/finite

When we have only one independent variable it is called Simple linear regression and when we have more than one independent variable it is called Multiple Linear regression.



$$y = mx + c$$

slope — intercept.

What values of m and c will give us best fit line?

Approach

1. Gradient descent algo.
2. Direct formula (least sq. method)

Gradient Descent Algorithm

↳ optimal m & b values
  ↳ slope ↳ intercept

m, b → machine learning → weights
b → θ_0 : intercept
m → θ_1 : slope

imp
$$y = b + mx$$
$$y = \theta_0 + \theta_1 x$$

→ line equation
↳ linear regression
↳ Relationship is proportional in nature

In ML, hypothesis → dependent variable / Target value i.e. y in our case.

$$y = \theta_0 + \theta_1 x$$
↳ $h_\theta(x) = \theta_0 + \theta_1 (x)$

$h(x) / h_\theta(x) = y$

$h_\theta(x) \rightarrow y$ but this y is not the actual value in our dataset.
It is the predicted y.

Example

x	y
10	5
12	6.6
13	1

$\theta_0 = 1$ intercept
$\theta_1 = 0.5$ slope

$h_\theta(x) = \theta_0 + \theta_1(x)$
for data point $x = 10$,

predicted $\hat{y} = h_\theta(x) = 1 + (0.5) \times 10$
$= 6$

$y = 5$

Actual value from data set

Error = (predicted value − original value)^2
mean square error
$= (6-5)^2$
$= 1$

(White board ! Cost function)

$$\sum_{i=1}^{m} \frac{(\text{predicted} - \text{original})^2}{2m}$$

$J(\theta_0, \theta_1)$
↳ cost function

Our Goal is to minimize the cost function, we want our error close to zero

Error
⟶ Reduce it to zero.
↳ minimize my cost funct
⇓
Best fit line

How Gradient descent will help in getting those optimal m & b values?

Maths!

⇒ m → current position of a person
D → steepness sloop
L = Speed with which he should move

m' → next position
L × D → size of the step

$J(\theta)$ = cost function

$$= \frac{1}{2m} \sum_{i=1}^{m} [h_\theta(x) - y]^2$$

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\frac{1}{2m} \sum_{i=1}^{m} [h_\theta(x) - y]^2\right]$$

$$= \frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta} [h_\theta(x) - y]^2$$

$$\boxed{\frac{\partial}{\partial x} x^n = n\, x^{n-1}}$$

$$= \frac{1}{2m} \sum_{i=1}^{m} 2\,[h_\theta(x) - y]\, \frac{\partial}{\partial \theta}\,[h_\theta(x) - y]$$

$$\Downarrow$$

$$?$$

$$= \frac{1}{m} \sum_{i=1}^{m} [h_\theta(x) - y]$$

partial derivative wrt θ_1, θ_0

$$\frac{\partial}{\partial \theta_0}[h_\theta(x) - y] = \frac{\partial}{\partial \theta_0}[\theta_0 + \theta_1 x - y]$$

$$= [1 + 0 - 0]$$

$$= 1$$

$$\frac{\partial}{\partial \theta_1}[h_\theta(x) - y] = \frac{\partial}{\partial \theta_1}[\theta_0 + \theta_1 x - y_1] = x$$

minimize cost function :

$$\frac{\partial}{\partial \theta_0}(J(\theta_0, \theta_1)) = \frac{1}{m}\left[\sum_{i=1}^{m} h_\theta(x) - y\right]$$

$$\frac{\partial}{\partial \theta_1}(J(\theta_0, \theta_1)) = \frac{1}{m}\left[\sum_{i=1}^{m} h_\theta(x) - y\right] . x$$

New values of $m \rightarrow \theta_1$

$b \rightarrow \theta_0$

$$\begin{cases} \theta_0 = \theta_0 - \dfrac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x) - y) \\ \theta_1 = \theta_1 - \dfrac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x) - y) \cdot x \end{cases}$$

α = learning rate (length of step)

↳ 0.0001

— x — x — x — x —