"LECTURE#01"

Date:

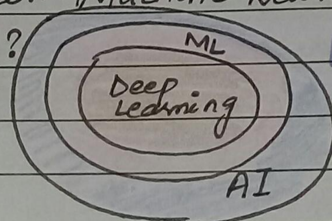
→ Internal Work of Machine Learning
→ Machine Learning on Data Science
→ 80% Feature Engineering, 20% Modelling
* Why machine learning is different from Conventional Programming?

Ans: You take user's input as well as output Machine Learning provide you code.

→ Differentiate between Machine Learning, Artificial Intelligence, Deep Learning?

کونسا ایسا کام ہے جسے ہم ML نہیں کیتے

- AI is superset

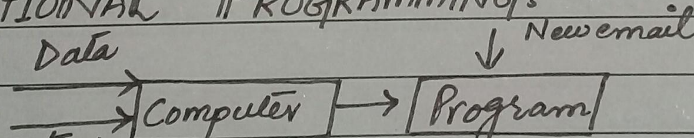


→ LEARNING: "The ability ~~of~~ to improve the behaviour ~~of~~ based on Experience"

→ By TOM MITCHAELL 1998

- A Computer Program is said to learn from Experience E with respect to some Task T & some performance measure P, if its performance on Tas measured P, improves with Experience E.

TRADITIONAL PROGRAMMING:



Parameters

i.e: Area, Size cost output

- Model is program / Code
- Nearest Best result given by Model
- Collection of data is important

→ Predictive Analysis

Date:

→ Self Learning اپنی غلطیوں سے سیکھنا
• Types of Learning Algorithm:
* Supervise Learning → LABEL لگا ہوا ہو
* Unsupervise Learning → MIX DATA BEFORE CLASSIFICATION
~~BEFORE LABELLING~~

→ Same data Behaviour → GROUPED
⇒ SUPERVISED LEARNING:
اگر ہم output دے رہے ہیں تو وہ سپروائزڈ لرننگ ہے۔
→ UNSUPERVISED LEARNING:
There are 2 types وہ جس کو نیو ڈیٹا کو specific Category میں divide کرتے ہیں
• Classification Problem } supervised Category
• Regression Problem
وہ جس کو پیرامیٹرز کے through 1 ویلیو calculate ہو کر آئے
BINARY CLASSIFICATION
کرے گا یا نہیں کرے گا۔
• Classification is Discrete Mapping
• Regression Problem is Continuos
ایسی چیز جو measure نہیں ہو سکتی اور count ہو سکتی ہو وہ discrete ہوتی ہے۔
(Countable things → Discrete)
جہاں پوائنٹ کا concept آجائے وہ regression Problem ہوتی ہے۔
• The things which are send in the form of Numbers are called features
• Set Threshhold Calculation
• By Tunning Threshold you can control your system
• attributes/Features → Parameters
Kmean method for Clustering

Date:

→ What is Machine Learning?

Arthur Samul (1959): Machine Learning: Feild of study that gives computers ability to learn without being explicitly programmed.

Tom Mitchell (1998): Well Posed learning problem:
A Computer program is said to learn from Experience E with respect to some task T & some performance measure P, if its performance on T, as measured by P, improves with experience E.

Suppose your email program watches which emails you do or do not mark as spam, & based on the learns how to better filter spam. What is the task T in this setting?

→ Classifying emails as spam or not spam **T**
→ Watching you label emails as spam or not spam **E**
→ The number (of fraction) of emails correctly specified as spam/not spam. **P** **M**

Machine Learning algorithms:

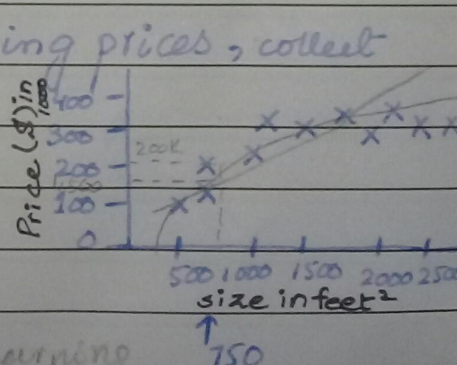
→ Supervised Learning (we are going to teach Computer how to do something)
→ Unsupervised Learning (We going to learn by itself)

Others: Reinforcement Learning, recommender system

Also talk about: Practical advice for applying learning algo's

SUPERVISED LEARNING:

Let say you are going to predict Housing prices, collect data sets from a city



→ The Question is how to decide that you wanna fit striaght line or a Quadratic function to the data

→ This example belongs to Supervised Learning

In Supervise learning each data is labelled

PRODUCT OF H·P

Date:

→ In supervised learning we give input as well as output. That enforce the algorithm to produce more accurate results

REGRESSION PROBLEM:

→ Predict continuos valued output (Price)

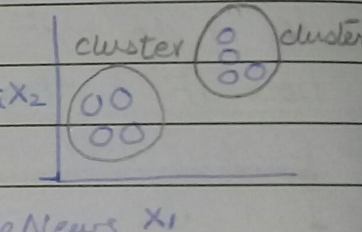
→ Prices can be round of so, that they are discrete values.

UNSUPERVISED LEARNING:

→ In Unsupervised learning there are no labels exist on the data

→ Here is a data set can you find some structure in the data

→ Same behaviour data's are clustered for the purpose of separation or breaking data



→ Clustering is used in Google New's

→ One Topic same stories are clustered on googleNews by using crawler.

Date:

What is Machine Learning?

Machine Learning is the scientific study of algorithms & statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns & inference instead. It is a subset of AI.

What is Deep Learning?

Deep Learning is part of broader family of machine learning methods based on artificial neural with representation learning

What is Artificial Intelligence?

A branch of computer Science, examine how we can achieve intelligent behaviour through computation

[the automation of] activities that we associate with human thinking, activities such as decission making, problem solving, learning (Bellman 1978)

'The art of creating machines that perform functions that require intelligence when performed by people

(Kurzweil 1990)

First, we have accuracy test
→ straight line define connected values.

بخار کتنا ہوگا؟ → continuos
بخار ہوگا یا نہیں ہوگا؟ → Discrete

Date: ____

LECTURE # 02

→ We have data & we want to predict output
→ i.e: How many runs will be make by Pakistan?
→ 1 Means = Yes or 0 Means = NO } Classification

<u>Data</u>

S#	A_1,	A_2,	A_3,	A_4	......	A_n	Output
1	✓	✓	✓	✓		✓	1

→ Temperature is continuos so it belongs to regression

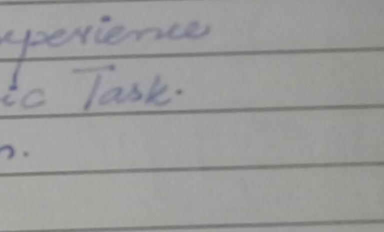
1 Input → / output → [ ] → [Program/Model]

→ The model is trained through Experience
→ A model is created for specific Task.
→ We work on existing algorithm.

→ SUPERVISED LEARNING:

→ REGRESSION:

→ We are going to calculate value → straight line
→ The model we used is : Linear Regression
→ <u>Regression</u> is a supervised learning problem
→ At given X and Y, we need to develop a model of linear regression which predict Y on new value of X



$$f: X \rightarrow Y$$

$$f(x) = Y = m\,x + c$$

m → slope
c → y-axis / y-intercepts

the value of y depends on x

For regression Y is continuos

one attribute

$X \xrightarrow{f(x)} Y$

Date:

S#	Area S.feet	Cost
1	120	1100 K
2	80	1000 K
3	20	2000 K
4	1000	500 K

$f(x) \rightarrow Y$

$f(x_1, x_2) \rightarrow Y$

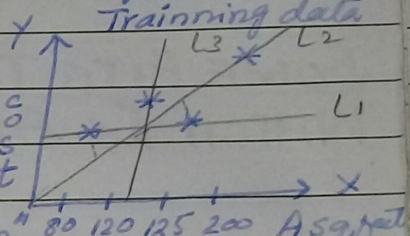
$f(x_1, x_2, x_3) \rightarrow Y$

→ If you have 3 attributes you can't represent it on Computer

→ Working is on 'n' dimensions.

→ Sum of distance must be minimum for drawing line on plotting points

→ The summission is called "Cost function."



$= d_1 + d_2 + d_3 + d_4$

$= d_1' + d_2' + d_3' + d_4'$

→ distance should be minimum

HEIRARCHIES OF REGRESSION

1 feature — Regression Models

- Simple
  - Linear
  - Non-linear
- Multiple (2+ feature)
  - Linear
  - Non-linear

LINEAR REGRESSION:

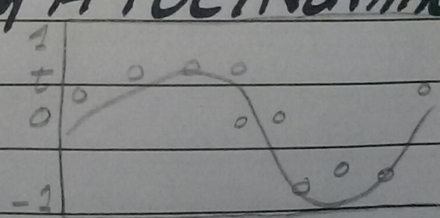
→ The Simplest function is linear function for Regression.

1- Simple regression where instance x depends on single variable.

2- Multiple regression where instance x depends on multiple variables.

A SIMPLE EXAMPLE: FITTING A POLYNOMIAL

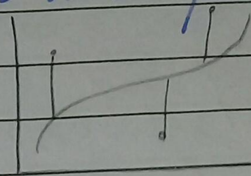
→ The green curve is a true function (which is not a polynomial)



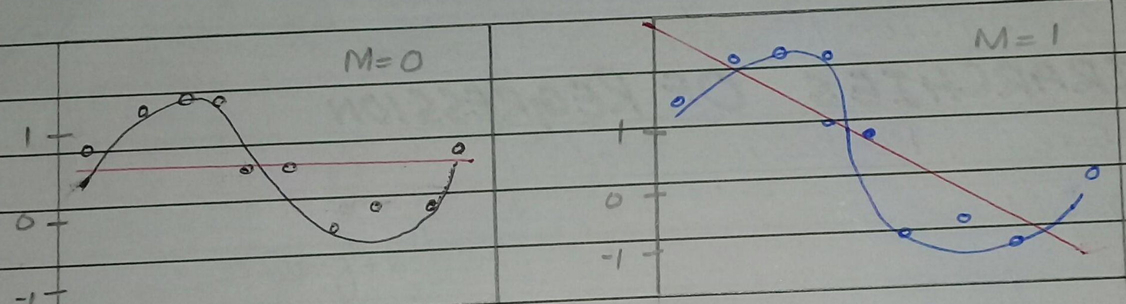
Date:

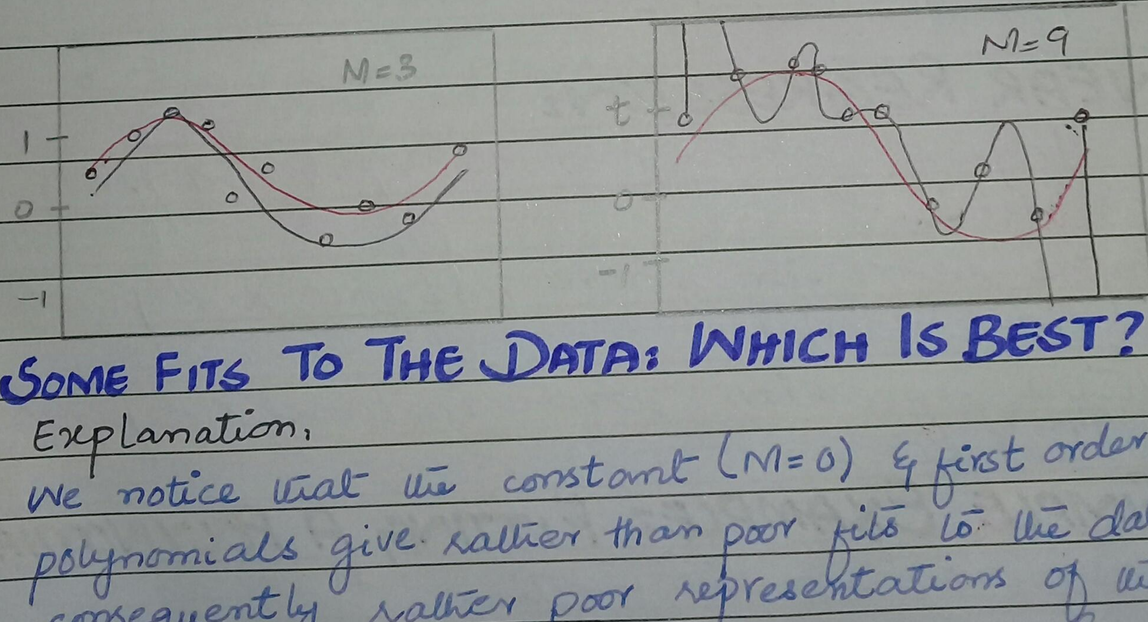
→ Plot of a trainning data of N = 10 points, shown as blue circles, each comprising an observation of the input variable x along with the corresponding target variable t.

→ We may used a loss function that measures the squared error in the prediction of y(x) from x



SOME FITS TO DATA: WHICH IS BEST? (1/2)





SOME FITS TO THE DATA: WHICH IS BEST? (2/2)

Explanation,

We notice that the constant (M=0) & first order (M=1) polynomials give rather than poor fits to the data & consequently rather poor representations of the functio sin (2X). The third order M=3 polynomial seems to give the best fit to the function sin (2πX) of the examples shown in figure

PRODUCT OF H·P HUSSAIN PAPER PRODUCT

→ When we go to a much higher order polynomial (M=9), we obtain an excellent fit to the trainning data.
→ In fact, the polynomial (M=9), we obtain an excellent fit to the trainning data $E(w^*)=0$

Date:

1 Input 1 output

→ Trainning set (this is your data set)

→ m = number of trainning examples

x's = input variable target variables

y's = output variable target example

(x, y) = single trainning example

(x^i, y^j) = specific example (i^{th} trainning example)

(Trainning Set)

↓

(Learning Algorithm)

↓

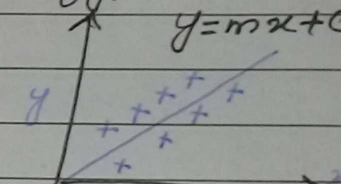
size of ~~House~~ → (h) → Estimated price

h is the function which map x's to y's

~~Hypothesis~~ Hypothesis (astimated value of y)

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

y=mx+c



constant dependant on x

$h_\theta(x) = 1.5 + 2x$

Best is that whose distance is minimum $h(x) = \theta_0 + \theta_1 x$

Actual − Hypothesis

We meant that y is a linear function of x

→ A cost function let us figure out how to fit the best straightline to our data

The cost function:

To minimize the sum of squared value

$h_\theta(1) = 1 + 1.5(1) = 2.5 \Rightarrow (2.5-1.5)^2 = 1$

$h_\theta(2) = 1 + 1.5(2) = 4 \Rightarrow (4-2)^2 = 4$

$h_\theta(3) = 1 + 1.5(3) = 5.5 \Rightarrow (5.5-2.5)^2 = 9$

14

→ Take Trainning set → Algorithm outputs a function

→ Pass into a learning algorithm (denoted h) (h = hypothesis)

• This function takes an input (e.g. size)

• Tries to output estimated value y

$$\sum_{i=1}^{i=n} ((\theta_0 + \theta_1 x) - y)^2$$

$$h_\theta(x_i) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \ldots \theta_n x_n$$

Date:

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

$$\underset{min}{J(\theta_0, \theta_1)} = \frac{1}{2M} \sum_{i=1} \underbrace{(h_\theta(x^i) - y^i)^2}_{difference}$$ $i = 1, 2, 3, 4 \ldots n$

Extra Parameters, Mean

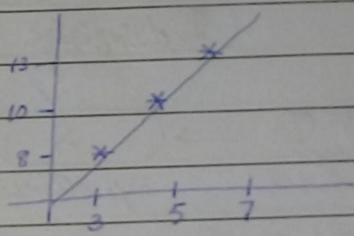
Minimize $(h_\theta(x) - y)^2$

Cost function is also called Squared Error Function

$$= \frac{1}{2 \times 47} \{ (h_\theta(x^{(1)}) - y^{(1)})^2 + (h_\theta(x^{(2)}) - y^{(2)})^2 \ldots \ldots$$

J is the notation of cost function.

#	X	Y
1	3	8
2	5	10
3	7	13



$h_\theta^{(1)}(x) = \theta_0 + \theta_1 x$

$h_\theta^{(1)}(x) = 1 + 1.5x$

$h_\theta^{(1)}(x) = 1 + 4.5$

$h_\theta^{(1)}(x) = 5.5$

Residual value = error

$h_\theta^{(2)}(x) = 1 + 7.5 \cancel{(5)}$

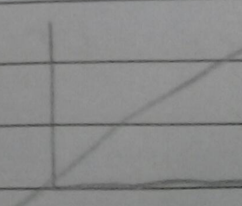
$h_\theta^{(2)}(x) = 8.5$

$h_\theta^{(3)} = 1 + 10.5 \cancel{t}$

$h_\theta^{(3)} = 11.5$

$$J(1, 1.5) = \frac{1}{2 \times 3} \left[(5.5 - 8)^2 + (8.5 - 10)^2 + (11.5 - 13)^2 \right]$$

$$\underset{min}{J(1, 1.5)} = 1.79$$



direction of slantness changes due to slope

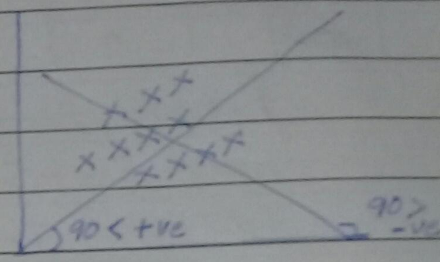
اگر 90 سے چھوٹا ہو تو +ve اگر 90 سے بڑا ہو تو -ve

slope means $= \theta_1$

Date:

Why there is 2 Parameters → because it is single feature

$$h_\theta^{(2)}(x) = 1.5 + 0.5x \quad \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}$$



→ Either θ negative or not?

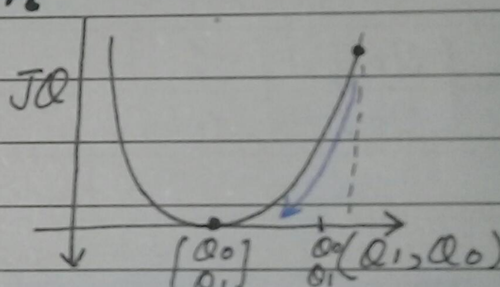
→ θ can be negative

کوئی ایسا method جو ثابت کر دے کہ یہ best ہے

$$h_\theta^{(3)}(x) = -1 + 1.5x \quad \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1.5 \end{bmatrix}$$

$$J(\theta_0, \theta_1) = \{J_1(\theta_0, \theta_1), J_2(\theta_0, \theta_2), \dots, J_n(\theta_0, \theta_2)\}$$

GRADIENT DECENT ALGORITHM:



It becomes near to converging points

Date:

Assignment By Home

S.No	X	Y
#1	3	8
#2	5	10
#3	7	13

$M = 3$

$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}$

$h_\theta(x_1) = \theta_0 + \theta_1 x$

$h_\theta(x_1) = 1.5 + 2(3)$

$h_\theta(x_1) = 7.5$

$h_\theta(x_2) = \theta_0 + \theta_1 x$

$h_\theta(x_2) = 1.5 + 2(5)$

$h_\theta(x_2) = 11.5$

$h_\theta(x_3) = \theta_0 + \theta_1 x$

$h_\theta(x_3) = 1.5 + 2(7)$

$h_\theta(x_3) = 15.5$

$J(\theta_0, \theta_1) = \frac{1}{2M} \sum_{i=0}^{i=m} (h_\theta(x_i) - y_i)^2$

$J(1.5, 2) = \frac{1}{2 \times 3} (7.5 - (8))^2 + (11.5 - 10)^2 + (15.5 - 13)^2$

$J(1.5, 2) =$

$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1.5 \end{bmatrix}$

$h_\theta(x_1) = \theta_0 + \theta_1 x$

$h_\theta(x_1) = -1 + 1.5(3)$

$h_\theta(x_1) = 3.5$

$h_\theta(x_2) = \theta_0 + \theta_1 x$

$h_\theta(x_2) = -1 + 1.5(5)$

$h_\theta(x_2) = 6.5$

$h_\theta(x_3) = \theta_0 + \theta_1 x$

$h_\theta(x_3) = -1 + 1.5(7)$

$h_\theta(x_3) = 9.5$

$J(\theta_0, \theta_1) = \frac{1}{2M} \sum_{i=0}^{i=m} (h_\theta(x_i) - y_i)^2$

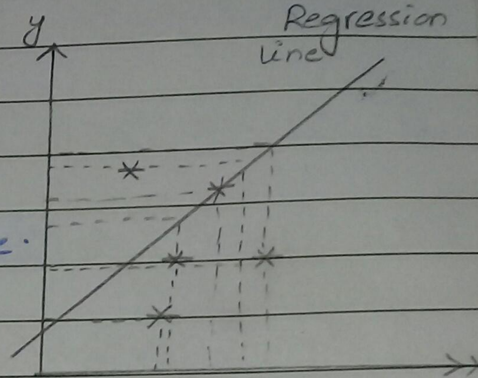
$J(-1, 1.5) = \frac{1}{2 \times 3} (3.5 - 8)^2 + (6.5 - 10)^2 + (9.5 - 13)^2$

$J(-1, 1.5) =$

Date:

LECTURE # 03

LINEAR REGRESSION:



→ How our line show with one feature in graph?

Ans = Linear line.

- Line depends on Y-intercept & slope.
- Gradient distance
- Least Square method.
- Theta is data dependance
- The distance of over all summi-sion should be minimum.
- For selection of best thetā we apply gradient descent

- Single Feature
- $h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$
Hypothesis
- unknown elements are θ_0 & θ_1
- $\underset{min}{J(\theta)} = \frac{1}{2M} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)})$
- Cost function

GRADIENT DESCENT:

Formula: $\boxed{\theta_0 = \underbrace{\theta_0}_{\text{Assume value of } \theta_0} - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})}$

$\boxed{\theta_1 = \theta_1 - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) * x^{(i)}}$

Formula of gradient Descent:

$\boxed{\theta_0 = \theta_0 - \underset{\text{learning rate}}{\alpha} \frac{\partial J}{\partial} \underbrace{J(\theta)}_{\text{cost function} \rightarrow (A)}}$ → General formula

First $\theta = \theta_0$

(A) ⇒ $\theta_0 = \theta_0 - \alpha \left[\partial/\partial\theta \ 1/2M \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \right]$

$\theta_0 = \theta_0 - \alpha/2M \left[\partial/\partial\theta \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)}) - y^{(i)})^2 \right]$

$= \theta_0 - \alpha/\cancel{2}M \times \cancel{2} \left[\sum_{i=1}^{m} ((\theta_0 + \theta_1 x^{(i)}) - y^{(i)} (1 + 0 - 0) \right]$

$\theta_0 = \theta_0 - \alpha/m \left[\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)} \right]$

Slope become parallel to x axis of Q_0

Sequential Method
Independant Method

Second $Q = Q_1$

$$Q_1 = Q_1 - \alpha \, \partial/\partial Q_1 \left[1/2M \sum_{i=1}^{m} (h_Q(x^{(i)}) - y^{(i)})^2 \right]$$

$$Q_1 = Q_1 - \alpha \, \partial/\partial Q_1 \left[1/2M \sum_{i=1}^{m} ((Q_0 + Q_1 x^{(i)}) - y^{(i)})^2 \right]$$

$$Q_1 = Q_1 - \alpha/\cancel{2}m \times \cancel{2} \left[\sum_{i=1}^{m} ((Q_0 + Q_1 x^{(i)}) - y^{(i)}) (0 + x^{(i)} - 0) \right]$$

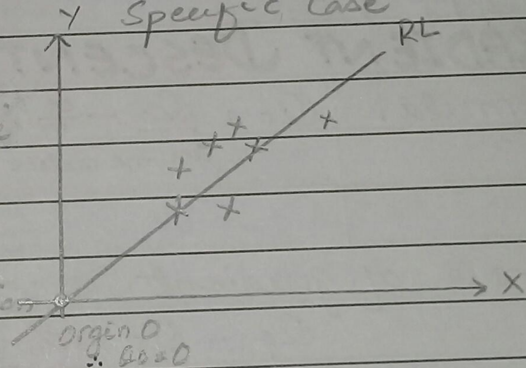
$$Q_1 = Q_1 - \alpha/m \left[\sum_{i=1}^{m} (h_Q(x^{(i)}) - y^{(i)}) * x^{(i)} \right]$$

GRAPHICAL PRESENTATION:

→ By defeault agr apnay right side per value lay li Hai bol function ki value bari Hogi

→ Agr apnay left side per value lay li Hai bol function ki value choti Hoti Hai

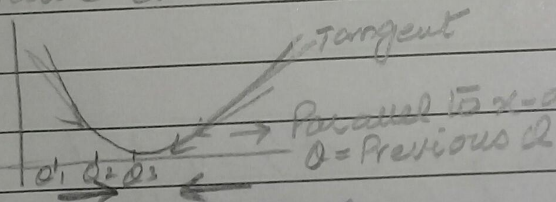
Specific Case



$h_Q(x) = Q_0 + Q_1 x$

$h_Q(x) = Q_1 x$

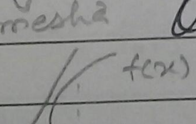
$$J_{Q = Q_1 \atop min} = \frac{1}{2M} \sum_{i=1}^{m} (h_Q(x^{(i)}) - y^{(i)})^2$$



→ Jis line ka slope 90 say chota Hota Hai wo Hamesha Positive Hota Hai

slope = Gradient

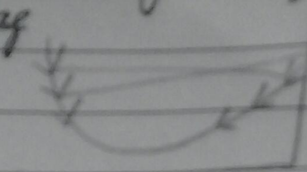
$$Q_1 = Q_1 - \frac{\alpha}{m} (h_Q(x^{(i)}) - y^{(i)}) * x^{(i)}$$



$$Q_j = Q_j - \frac{\partial}{\partial Q_j} J(Q_j)$$

Date

→ Learning Rate is used to tunning the formula
→ Learning rate is for zig zag or fluctuation of your



→ If you have low or High learning rate so, their is zig zag or

Example:

Create a Regression line for following set of points
$\{(-2,-1),(1,1),(3,2)\}$

Solution:

$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$ Let $\alpha = 0.01$

$\theta_0 = \theta_0 - \alpha/m \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$ Given $m = 3$

X	Y
-2	-1
1	1
3	2

Suppose $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

For First Iteration:

$\theta_0 = 0$ & $\theta_1 = 0$

$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$

$= 0 + 0x^{(i)} = 0$

$h_\theta(x^{(1)}) = 0$

Similarly $h_\theta(x^{(2)}) = 0$

$h_\theta(x^{(3)}) = 0$

Now, $h_\theta(x^{(1)}) - y^{(1)} = 0 - (-1) = 1$

$h_\theta(x^{(2)}) - y^{(2)} = 0 - 1 = -1$

$h_\theta(x^{(3)}) - y^{(3)} = 0 - 2 = -2$

AND,

$(h_\theta(x^{(1)}) - y^{(1)}) * x^{(1)} = 1 \times (-2)$

$(h_\theta(x^{(2)}) - y^{(2)}) * x^{(2)} = -1(1) = -1$

$(h_\theta(x^{(3)}) - y^{(3)}) * x^{(3)} = (-2)(3) = -6$

Now,

$$Q_0 = Q_0 - \alpha/3 \sum_{i=1}^{3} (h_Q(x^{(i)}) - y^{(i)})$$

$$Q_0 = 0 - 0.01/3\,[1 + (-1) + (-2)]$$

$$Q_0 = 0.007$$

& $$Q_1 = Q_1 - \alpha/3 \sum_{i=1} (h_Q(x^{(i)} - y^{(i)}) * x^{(i)}$$

$$Q_1 = 0 - 0.01/3\,[-2 + (-1) + (-6)]$$

$$Q_1 = 0.030$$

$$Q = \begin{bmatrix} Q_0 \\ Q_1 \end{bmatrix} = \begin{bmatrix} 0.007 \\ 0.030 \end{bmatrix}$$

Suppose $Q = \begin{bmatrix} Q_0 \\ Q_1 \end{bmatrix} = \begin{bmatrix} 0.007 \\ 0.030 \end{bmatrix}$

Second Iteration:

$$h_Q(x^{(i)}) = Q_0 + Q_1 x^{(i)}$$

$$h_Q(x^{(i)}) = 0.007 + 0.030(-2)$$

$$h_Q(x^{(i)}) = -0.053$$

$$h_Q(x^{(2)}) = 0.007 + 0.030(1)$$

$$h_Q(x^{(2)}) = 0.037$$

$$h_Q(x^{(3)}) = 0.007 + 0.030(3)$$

$$h_Q(x^{(3)}) = 0.097$$

Now,

$$h_Q(x^{(i)}) - y^{(i)} = -0.053 - (-1) = 0.947$$

$$h_Q(x^{(2)}) - y^{(2)} = 0.037 - 1 = -0.963$$

$$h_Q(x^{(3)}) - y^{(3)} = 0.097 - 2 = -1.903$$

AND:

$$(h_Q(x^{(i)}) - y^{(i)}) \times x^{(i)} = 0.947 \times (-2) = -1.894$$

$$(h_Q(x^{(2)}) - y^{(2)}) * x^{(2)} = -0.963 \times (1) = -0.963$$

Linear regression is not uncertain

0.2667

Date:

$(h_\theta(x^{(3)} - y^{(3)})) * x^{(3)} = -1.903 * (3) = -5.709$

Now,

$\theta_0 = \theta_0 - \alpha/3 \sum_{i=1}^{3} (h_\theta(x^{(i)}) - y^{(i)})$

$\theta_0 = 0.007 - \frac{0.01}{3} [0.947 - 0.963 - 1.903]$

$\theta_0 = 0.013$

& $\theta_1 = \theta_1 - \alpha/3 \sum_{i=1}^{3} (h_\theta(x^{(i)} - y^{(i)})) * x^{(i)}$

$\theta_1 = 0.03 - \frac{0.01}{3} [-1.894 - 0.968 - 5.909]$

$\theta_1 = 0.059$

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 0.013 \\ 0.059 \end{bmatrix}$

Now, $h_\theta(x) = 0.013 + 0.059x$

Predict $x = 4.3$?

$h_\theta(x = 4.3) = 0.013 + 0.059(4.3)$

Date:

→ co relation
→ featured Engineering } dependant on Human

→ import pandas as pd → read data from csv file (imp)
→ import numpy as np → third person array
→ import matplotlib.pyplot as plt → scattered value, for graph
→ import seaborn as sns
→ from sklearn.model_selection import (imp) train_test_split
→ from sklearn.linear_model import Linear regression, Lasso, Ridge
→ Read csv file
→ store it into variable
Boston_Train = pd.read_csv ("boston-train.csv")
Print (Boston-Train.Head())
action in Memory
→ Drop Id
Boston_Train.drop ('ID', axis = 1, inplace = True)
x = Boston Train [[features/attributes]]
y = Booston Train[medv]
X-train, Xtest

Regression →
Classification

Date:

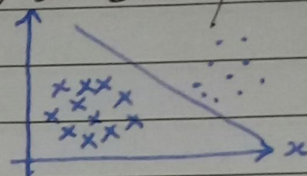
LINEAR REGRESSION
(LECTURE # 04)

Methametical Regression

classification

LOGISTIC REGRESSION:

→ We are going to handle classification problem
→ Multiclass problem also exist, same as binary class.
→ Classification Problem



→ Threshold or decission boundary will be created between the two types of dissease

SNo						target
1						0
2						1
Test	-	-	-	-	-	?

→ Agr line sy bari ay to above the line
→ Agr value line sy choti ho to below the line
→ If their is 2 class problem, we will create boundary (either polynomial line, or linear line)
→ Decission boundary is sometime called Threshold
→ Sigmoid function has the property for creating Boundary.
→ We need separation between data

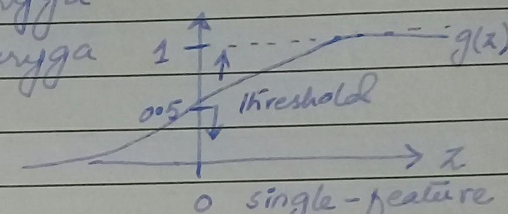
What issues of linear regression? in Binary classification?
→ Threshold

Date:

→ We are disscussing Binary classification → sigmoid

→ **Logistic Regression method →**

→ 0 or 1 k darmiyan asymptotic Hota Hai
Kbhi bhi 1 ko meet nahi Karyga
Kbhi bhi 0 ko meet nahi Karyga



→ What does Sigmoid look like?
→ crosses 0.5 at the origin, then flattens out
- Asymptotes At 0 & 1

→ Given this we need to fit θ to our data

$$h_\theta = \theta_0 + \theta_1 x$$
$$= \begin{bmatrix}\theta_0\\ \theta_1\end{bmatrix}^T \begin{bmatrix}1\\ x\end{bmatrix}$$
$$= \begin{bmatrix}\theta_0 & \theta_1\end{bmatrix}\begin{bmatrix}1\\ x\end{bmatrix}$$
$$= \theta_0 + \theta_1 x$$

→ Our Hypothesis becomes function

$$h_\theta(x) = g((\theta^T x))$$

x ka koi function Hoga
$f(x)$
Part of function
$f(x) = x^2 + x + 1$
$\theta^T x = z$
$\underbrace{g(z) = g(\theta^T x)}_{\text{sigmoid}}$

$$g(z) = \frac{1}{(1+e^{-z})}$$

$$g(z) = \frac{1}{1+e^{-\theta^T x}}$$

→ Their exist multiple functions for logistic Regression

$$z = \theta^T x = \theta_0 + \theta_1 x$$
$$z = \theta^T x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$
$$z = \theta^T x = \theta_0 + \theta_1 x_1 + \theta_1 x_1 + \theta_1 x^2 + \theta_3 x^3 + \dots$$

Regularization → formula

seaborn → density graph

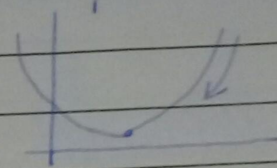
Date:

Issues

→ We don't get convex curve, if we put (bowl function)
$\frac{1}{1+e^{\theta t x}}$ in our Hypothesis

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

اس سے bol نہیں آتی ہے



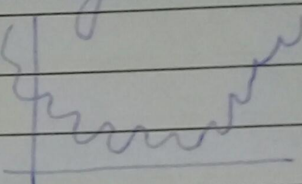
→ We are basically finding θ

→ On which basis θ will be milyga? Imp
→ On the basis of X → is the training data

→ There are several local minimas

→ We need single local minima

→ There should be no local minimas



→ Multiple local minimas that's why we are not using linear regression. (Imp)

→ We need a function which should give single bol eq

Cost function → training data ka X Hona Chahiyay
→ Output / Target ka Y Hona Chahiyay
→ θ values → for developing function

→ Optimized → minimum
→ maximum

minimum optimized value

Statistical model of ML (Gareth James)

جس کو ڈھونڈ رہے ہیں وہ Hypothesis

LECTURE #05

Date:

→ Logistic Regression → categorial dalā ko simplified krta hai

→ We create class boundary & the line is our Hypothesis.

→ We put dependancy of x on 'Q'

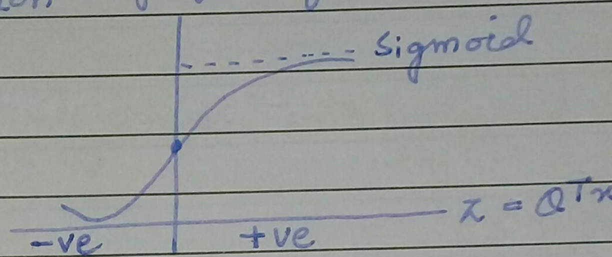
→ Q nikalnay k lijay Gradient Descent

$g(z) \rightarrow z = Q^T x = Q_0 + Q_1 x$

$$h_Q(x) = \overbrace{g(x)}^{\text{Function}}$$

$$\Downarrow$$

$$\frac{1}{1+e^{-Q^T x}}$$



→ Sigmoid function is suitable for two

→ Threshold = 0.5

→ Agr 0.5 sy bara ho tō 1 agr chota hota tō 0

→ Boundary can be of any type

→ After train 'Q' will be obtained

$$z = (Q^T x)$$

So, when

$$Q^T x \geq 0$$

$$h_Q \geq 0.5$$

→ Input is always on x-axis

→ Output is always on Y-axis

→ Sigmoid function is Logistic Function

Date:

We design cost function in ML

$h_\theta(x) = g(z) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

$= g(-3 + x_1 + x_2)$

$-3 + x_1 + x_2 \geq 0$

$x_1 + x_2 = 3$

$1 + 1 = 3 \leq$

$3 + 1 = 3 \geq$

Q- How can we say that

$x_1 + x_2 = 3$

$(x_1, 0)$ & $(0, x_2)$

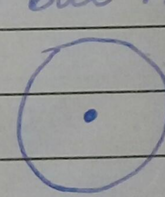
$(3, 0)$ & $(0, 3)$

Take 2 points & find cordinates

→ Get logistic regression to fit a complex

$h_\theta(x) g(\theta_0 + \theta_1 x_1 + \theta_3 x_1^2 + \theta_4 x_2^2)$

To find θ, we use gradient descent but now we assume $\theta = [-1; 0; 0, 1, 1]$



$g(z) = g(-1 + x_1^2 + x_2^2)$

$0 <$ ≥ 0

$y = 0$ $y = 1$

$-1 + x_1^2 + x_2^2 >= 0$

or $x_1^2 + x_2^2 = 1$

→ Why cost function is used? Imp

→ We are interested for the best min θ for Classification. We are seeking θ for best Hypothesis.

Date:

→ Train data means 'θ' nikalna

Cost Function:

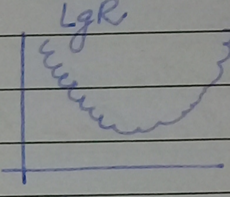
$z = \theta^T x$ (θ: Parameter, x: Feature)

optimize

$$\underset{min}{J(\theta)} = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 \quad \text{LR}$$

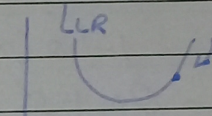
$\theta_0 + \theta_1 x^{(i)}$

For Logistic Regression → $\dfrac{1}{1+e^{-\theta^T x}}$

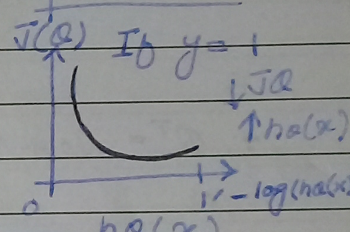


If $y = 1$, $J(\theta)$ should be minimum

→ Jab ap hypothesis sy target value tk phnchy to result minimum ho



جیسے جیسے ہم اپنے ٹارگٹ کے قریب ہوں گے تو cost function کم سے کم ہو گی۔

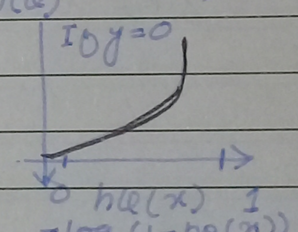


→ We need 2 functions for finding θ

جیسے جیسے ہم ٹارگٹ سے دور ہوں گے cost function بڑھے گا۔

→ θ nikalnay k liyay jo theta is timal $J(\theta)$ hoga

then, we combine both functions



Gradient Descent

→ Cost Function Formula:

$$\text{cost}(h_\theta(x), y) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

This equation is a more compact of the two cases above. Task

(Gradient Descent ka formula

Gradient Descent $= \dfrac{\partial}{\partial\theta} J(\theta)$

Linear Regression, Logistic Regression

• Cost function ka kam sirf result nikalna Hai
• Cost function yay decide nahi krta k konsa Q best hy

Date:

LECTURE # 06

There are 2 constraints

→ Multi class (only 2 class)

→ Using Sigmoid Function

جو بھی data آپ کو دیا ہوا ہو اسمیں سے نیا آنے والا ڈیٹا کس کلاس میں جانا
کرے گا -

→ We have to create threshold for making boundary

$$h_Q(x) = g(x) = \frac{1}{1+e^{-Q^T x}}$$

$$z = Q^T X \Rightarrow ?$$

↳ Linear
↳ Non - Linear

→ Hypothesis khud ak boundary line ہے

$$z = Q_0 + Q_1 x \longrightarrow \text{linear}$$

$$z = Q_0 + Q_1 x^2 \rightarrow \text{non-linear}$$

$$h_Q(x) = \frac{1}{1+e^{-(Q_0+Q_1 x)}}$$

$$h_Q(x) = \frac{1}{1+e^{-(Q_0+Q_1 x^2)}}$$

→ We develop cost function for best Q finding

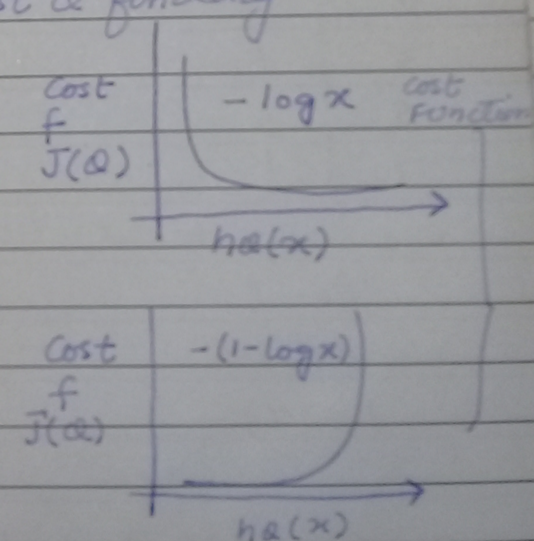
→ Cost function Q

For finding minimum Q we use Gradient Descent.

→ We find Q that become unchange

$Q_j =$

Gradient Descent provide standard

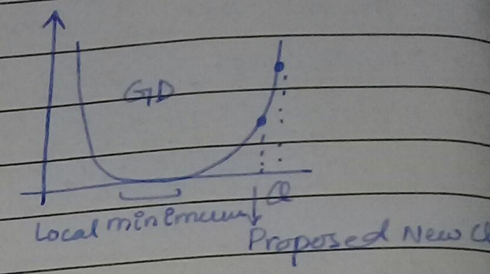


PRODUCT OF H·P

Cost function & GD are not boundary function.
They are supporting factors.

Date:

→ Data Noise } Feature Engineering
Data Biased } Problems.



•

→ COST FUNCTION:

FORMULA:

$$J\theta = -\frac{1}{m}\Big[\underbrace{\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)})}_{\text{For 1}} + \underbrace{(1-y^{(i)}) \log(1-h_\theta(x^{(i)}}_{\text{FOR 0}}$$

→ Binary Class

GRADIENT DESCENT:

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)}$$

Lin Reg: $\theta_0 + \theta_1 x$

log Reg: $\frac{1}{1+e^{-\theta^T x}}$

Sawal:

X	Y
3	1
5	0
7	1

We have to calculate Cost Function

imp

$\theta_0 = 1$

$\theta_1 = 1.5$

$$h_\theta(x) = \frac{1}{1+e^{-z}}$$

$$z = \theta^T X = \theta_0 + \theta_1 x$$

Cost Function is not iterative

$z_1 = 1 + 1.5(3) = 5.5$

$z_2 = 1 + 1.5(5) = 8.5$

$z_3 = 1 + 1.5(7) = 11.5$

Now, calculate Hypothesis

$$h_\theta(x^i) = \frac{1}{1+e^{-z}}$$

$$h_\theta(x^i) = \frac{1}{1+e^{-(5.5)}}$$

$$h_\theta(x^1) = 0.9959$$

$$h_\theta(x^{(2)}) = \frac{1}{1+e^{-8.5}}$$

$$h_\theta(x^{(2)}) = 0.9997$$

$$h_\theta(x^{(3)}) = \frac{1}{1+e^{-11.5}}$$

$$h_\theta(x^{(3)}) = 0.9999$$

Now, Applying Cost function

$$J(\theta) = -\frac{1}{3}\{y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}) + y^{(i)}\log h_\theta(x^{(i)})\}$$

$$J\theta = -\frac{1}{3}\{1\log(0.9959) + (1-0)\log(1-0.99 + 1\log(0.9999))\}$$

$$J\theta = -\frac{1}{3}\{-0.0017 - 3.5228 - 0.0017\}$$

$$J\theta = -\frac{1}{3}\cancel{(-7.8236)}(-3.5262)$$

$$J\theta = 1.1754$$

- sign shows convergence

Date

Gradient Descent is used to adjust θ

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Gradient Descent

Data: Need's initial θ

X	Y
3	1
5	0
7	1

Need's h_θ

Need's data

Initial θ's $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$z = \theta^T x = \theta_0 + \theta_1 x$

$\alpha = 0.1$

$z_1 = 0$

$z_2 = 0$

$z_3 = 0$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$h_{\theta_1}(x) = \frac{1}{1+e^{-0}} = 0.5$$

$$h_{\theta_2}(x) = \frac{1}{1+e^{-0}} = 0.5$$

$$h_{\theta_3}(x) = \frac{1}{1+e^{-0}} = 0.5$$

$X_0 = 1$

Induce Error

Date:

$J = 0$

$Q_0 = Q_0 - 0.1\{(0.5 - 1)\, x_0^{(1)} + (0.5 - 0)\, x_0^{(2)} +$
$(0.5 - 1)\, x_0^{(3)}\}$

On the basis of formula, we consider $X_0 = 1$

$Q_0 = 0 - 0.1\{-0.5 + 0.5 - 0.5\}$

$Q_0 = -0.1 + (-0.5)$

$Q_0 = 0.05$

$Q_1 = Q_1 - 0.1\{(0.5 - 1)\, x_1^1 + (0.5 - 0)\, x_1^2 + (0.5 - 1)\, x_1^3\}$

$Q_1 = 0 - 0.1\,(-0.5)\,3 + (0.5)\,5 \quad + (0.5)\,7$

$Q_1 = 0 - 0.1\,(-2.5)$

$Q_1 = 0.25$

New $Q = \begin{bmatrix} 0.05 \\ 0.25 \end{bmatrix}$

$z_1 = 0.05 + 0.25 * 3 = 0.8$

$z_2 = 0.05 + 0.25 * 5 = 1.3$

$z_3 = 0.05 + 0.25 * 7 = 1.8$

$h_Q(x) = \dfrac{1}{1 + e^{-Q^T x}}$

$h_{Q_1}(x) = \dfrac{1}{1 + e^{-0.8}} = 0.6899$

$h_{Q_2}(x) = \dfrac{1}{1 + e^{-1.3}} = \cancel{0.8528}\ 0.785$

$h_{Q_3}(x) = \dfrac{1}{1 + e^{-1.8}} = 0.8582$

$Q_0 = 0.05 - 0.1\{(0.69 - 1) + (0.79 - 0) + (0.85 - 1)\}$

$Q_0 = 0.05 - 0.1\,(0.34)$

$Q_0 = 0.05 - 0.034$

$Q_0 = 0.016$

$Q_1 = 0.25 - 0.1\{0.69 - 1)\,3 + (0.79 - 0)\,5 + (0.85 - 1)\,7\}$

$Q_1 = 0.046$

Sigmoid → Probablity

Now,

$\theta's = \begin{bmatrix} 0.016 \\ 0.046 \end{bmatrix}$ → Finalize (Best)

$z = 0.016 + 0.046x$

$h_\theta(x) = \dfrac{1}{1+e^{-(0.016+0.046x)}}$

Test data

$x = 4$

$h_\theta(x) = \dfrac{1}{1+e^{-(0.016+0.046(4))}}$

$= 0.2 > 0$