INTRODUCTION

Deep Neural Networks
Session 01

Pramod Sharma
pramod.sharma@prasami.com

2

Agenda

- Introduction
- AI vs ML vs Deep Learning
- MP Neuron

11/8/2025

pra-sâmi

3

References

- Deep Learning using Python, S Lovelyn Rose,L Ashok Kumar, D Karthika R/ Wiley India, 1st Edition
- Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville
- Neural Networks and Learning Machines, Simon Haykin
- Pattern Recognition and Machine Learning, Christopher M. Bishop
- Deep Learning with Python - François Chollet
- Hands-On Machine Learning with Scikit-Learn and TensorFlow
- TensorFlow Deep Learning Cookbook

11/8/2025

pra-sâmi

4

"*Theory exam– 40% weightage*

Lab exam – 40% weightage

Internal exam – 20% weightage"

Evaluation method

11/8/2025

pra-sâmi

5

Logistics

"The man who asks question is a fool for a minute, a man who doesn't ask is a fool for life."
~ Confucius

11/8/2025

pra-sâmi

6

Logistics

- We encourage 'discussion groups':
  - Study groups
  - Whatsapp groups
- Expect you to complete your assignment individually!
- No group assignments unless stated otherwise
- Code is small part of it
- Pay special attention to inline comments
  - Comments should focus on what you were trying to implement

11/8/2025

pra-sâmi

7

Agent In Uncertain Environment

- Agents don't have complete knowledge about the world.
- Agents need to make (informed) decisions given their uncertainty.
- It isn't enough to assume what the world is like.
  - Example: wearing a seat belt.
- An agent needs to reason about its uncertainty.
- When an agent takes an action under uncertainty, it is gambling ⇒ probability

pra-sâmi

8

Overview

- Nature is a continuum where as math is discreet values
  - Old film based images were continuous painting of colors ... are pixels
- Brain works differently than our mathem...
- Brain is highly complex, non... ...puter
- Neural networ... ...spired from brain
- Highly ... a Neural Network is modeled to simulate manner in which brain performs a task

All models are wrong... some models are useful!

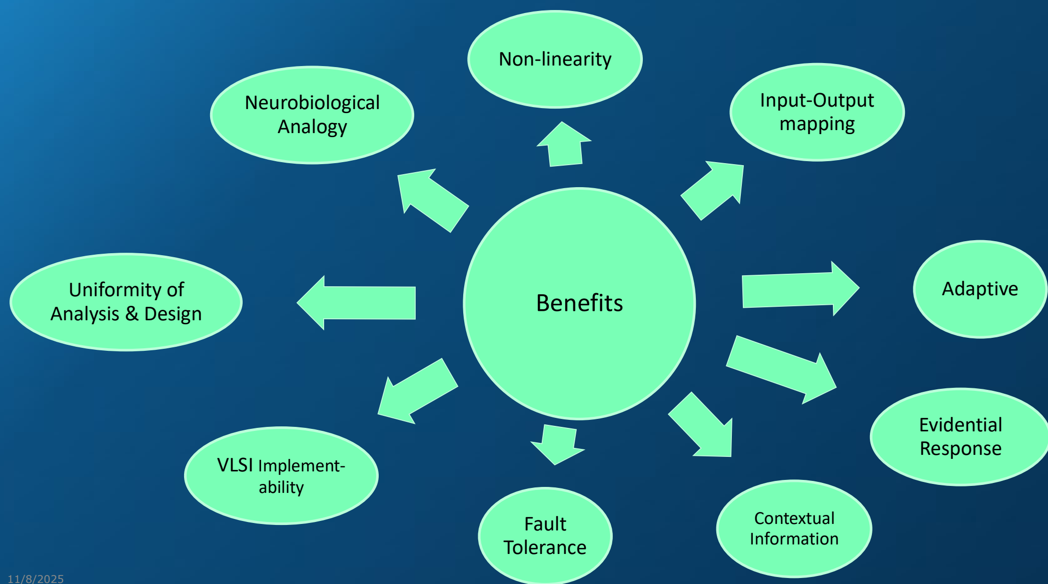
S...ulus → Receptors ⇄ Neural Net ⇄ Effectors → Response

Is this how our brain works? Really!!

pra-sâmi

9

Benefits of Neural Networks

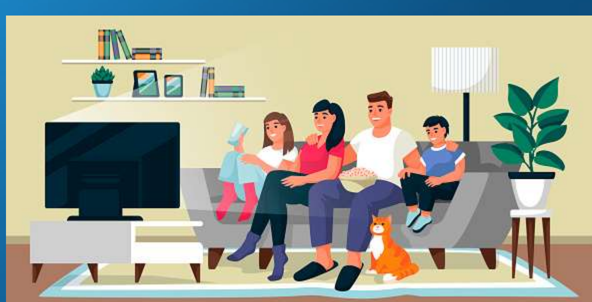


11/8/2025

pra-sâmi

10

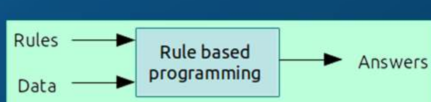
What's Being Played...



Can a Rule based system answer based on the data provided?

We, as humans can easily guess!

Time	Day	Type
9:00 AM	Weekday	News
11:00 AM	Weekday	K-Soaps
3:00 PM	Weekday	Soaps
5:00 PM	Weekday	Soaps
6:00 PM	Weekday	Cartoons
9:00 PM	Weekday	Sports, Movies
4:00 PM	Weekday	???
1:00 PM	Weekday	???

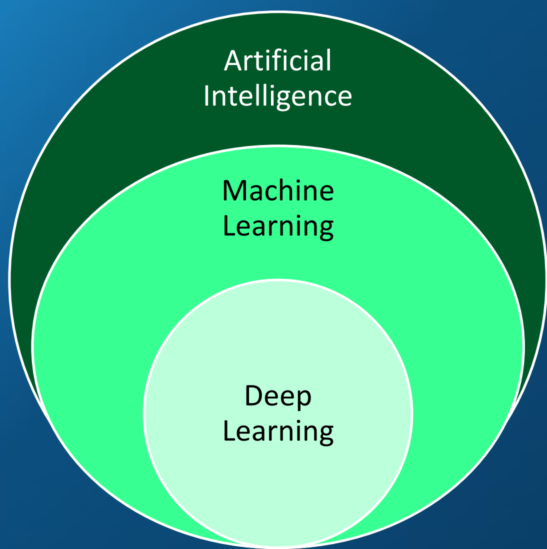


11/8/2025

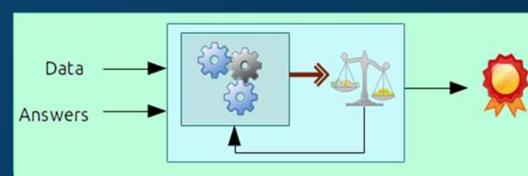
pra-sâmi

11

AI vs ML vs Deep Learning



- Used interchangeably
- AI is a broader concept, it includes basic AI to Deep learning.
- Machine learning: enabling Machines to Learn from the past incidents (available data).
- Deep Learning: One can say that it tries to copy information processing patterns found in the human brain



11/8/2025

12

Artificial Intelligence vs. Machine Learning

Artificial Intelligence

- Create intelligent machines that can simulate Human thinking capability and behavior
- A man-made thinking power
- No pre-programing needed
- Algorithms which can work with their own “intelligence”
- Algorithms such as Reinforcement learning algorithm and deep learning neural networks being used in multiple places such as Siri, Google’s AlphaGo, AI in Chess playing, etc.
- Based on capabilities, AI can be classified into three types:
  - Weak AI
  - General AI
  - Strong AI
- Currently, we are working with weak AI and general AI. The future of AI is Strong AI for which it is said that it will be more intelligent than humans (???)

Machine Learning

- An application or subset of AI
- Allows machines to learn from data without being programmed explicitly
- Uses a massive amount of structured and semi-structured data
- It can work only on data it has seen
- For unknown cases it becomes unresponsive or unreliable
- Being used for online recommender system, for Google search algorithms, Email spam filter, Facebook Auto friend tagging suggestion, etc.
- It can be divided into three types:
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning

11/8/2025

13

Deep Learning

- Large Neural Networks

"Using brain simulations, hope to:
Make learning algorithms much better and easier to use,
Make revolutionary advances in machine learning and AI,
I believe this is our best shot at progress towards real AI."

- Andrew Ng

- Learning successive layers of increasingly meaningful representations
- Modern network contain hundreds of successive layers
- Successive layers are learned via "neurons" connected via neural network

Some concepts were inspired by how our brain works
It is NOT a replica of human brain!!!

14

Deep Learning

- Why Deep Learning is more practical today?
  - Availability of large computing power
  - Availability of large datasets
- Most flavors of the old generations of learning algorithms, performance will plateau
- Deep learning that is scalable
  - Performance just keeps getting better as more and more data is fed
- Most value today is coming from supervised learning
- Eventually, we will see benefits of unsupervised learning

15

Deep Learning

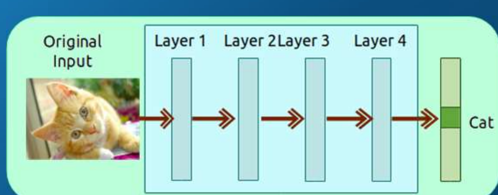
- Usually a neural network contains
  - Input Layer
  - Hidden layers [1 … n]
  - Output layer
- We may call network with 1 to 2 hidden layer as shallow
- Network with 10 or more layers as deep
  - No set demarcation!
- I guess, scientists just got excited when someone labeled them as deep network
- Intelligent software to automate routine tasks, understand speech or images, make diagnosis in medicine and support basic scientific research

11/8/2025

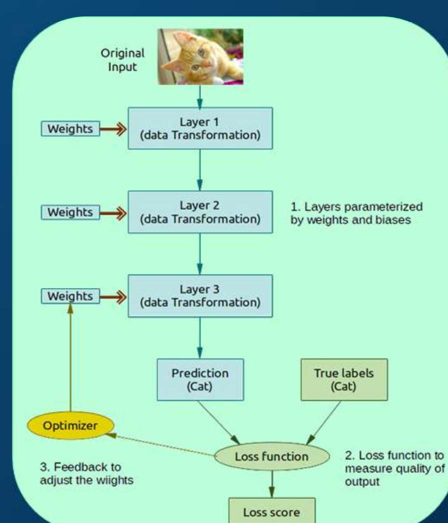
16

How deep learning works…

- As the images are processed through the layers



- The representations are increasingly filtered, purified and distilled to make them more meaningful



11/8/2025

17

What has been achieved so far

- Learn to see and hear… so natural to humans but elusive to machines earlier
- Image classification
- Speech recognition
- Handwriting recog
- Writing style recognition (who was the author)
- Improved machine translation
- Text-to-speech conversion
- Digital assistants such as Google Now and Amazon Alexa
- Little autonomous driving
- Improved ad targeting, as used by Google, Baidu, and Bing
- esults on the web
- Ability to answer natural-language questions
- Superhuman games playing: chess, go…

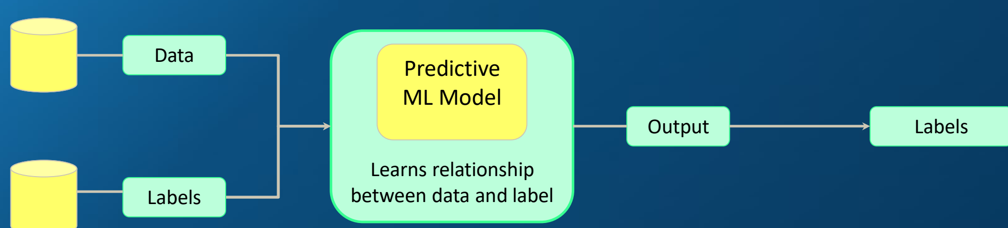
Still long way to go…
Human-level general intelligence too far away…

11/8/2025

pra-sâmi

18

Deep Learning

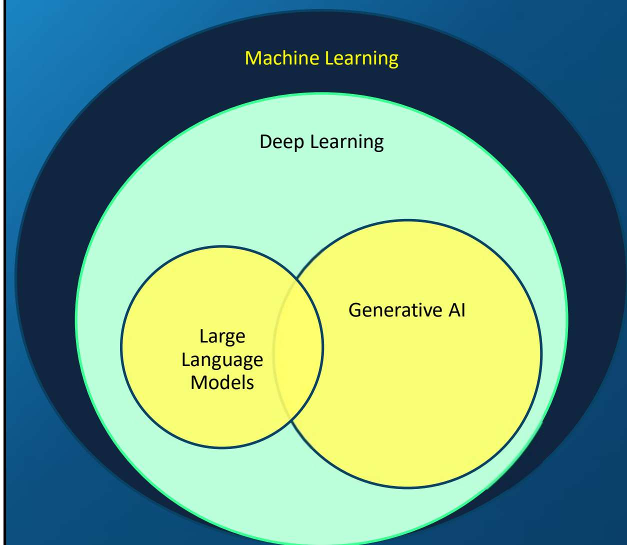


11/8/2025

pra-sâmi

19

Generative AI



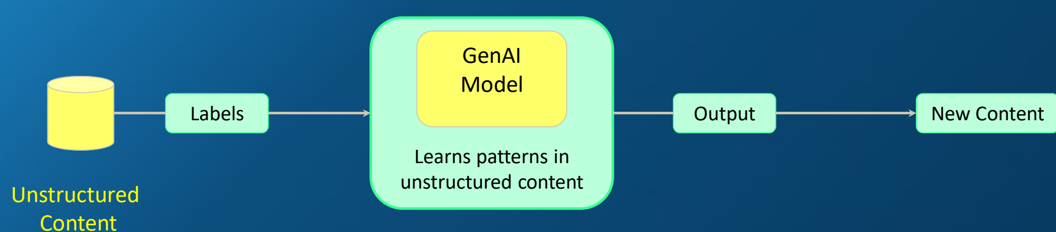
- Gen AI is a subset of Deep learning
- LLM:
  - For a given paragraph, predict its title and category
  - For a given request what should be response
- Can process:
  - Labeled data
  - Unlabeled data
- Using:
  - Supervised methods
  - Unsupervised methods
  - And semi-supervised methods

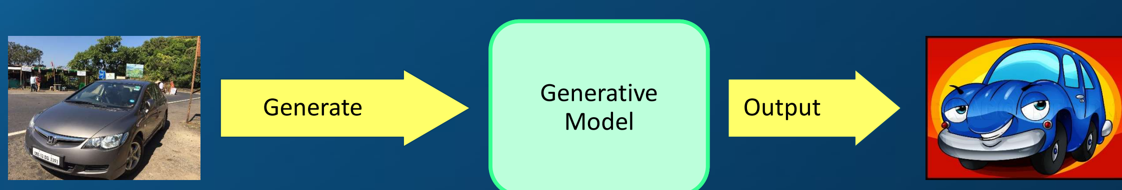
11/8/2025

pra-sâmi

20

Generative AI





11/8/2025

pra-sâmi

21

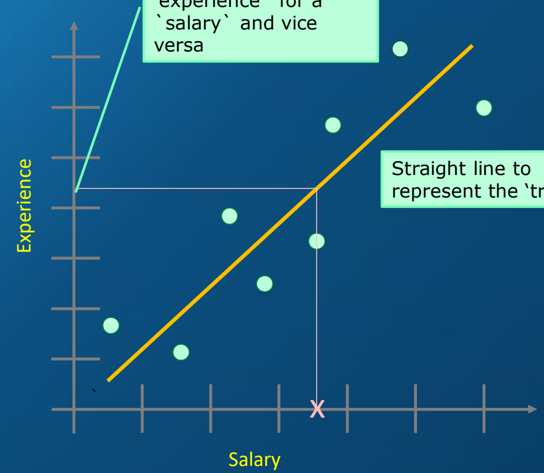
Linear Regression

11/8/2025

22

Correlation

We can guess `experience` for a `salary` and vice versa



Straight line to represent the 'trend'

At present, datapoints are somewhat away from the trend line.

11/8/2025

23

Correlation

If datapoints are closer to the trend line, our guestimates will be more accurate.

We can say, if datapoints are away from the trend line, two features have weaker relation.
If they are close together, the relationship is strong.

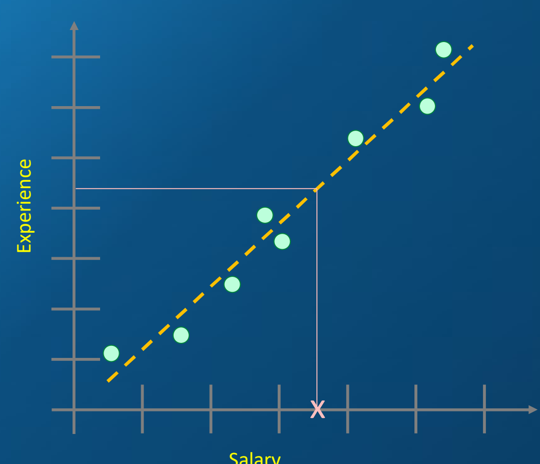
11/8/2025

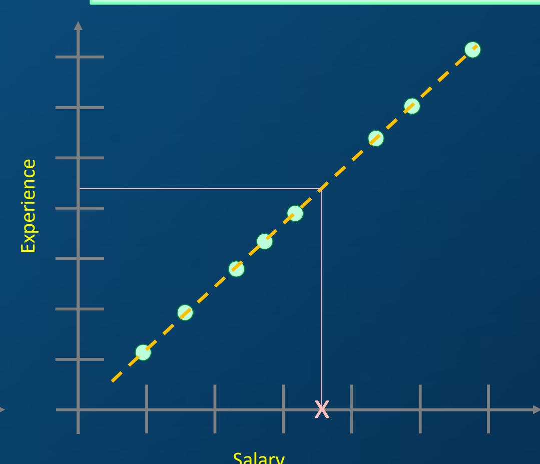
pra-sami

24

Correlation

We can hope for best fit at correlation = 1
The trendline passes through each of the point.





PS: Worst case is when correlation becomes 0.

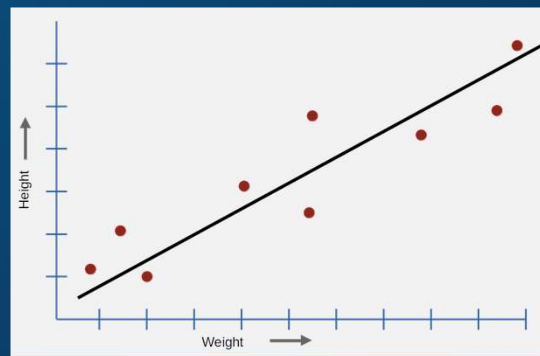
11/8/2025

pra-sami

25

Fitting a line to data

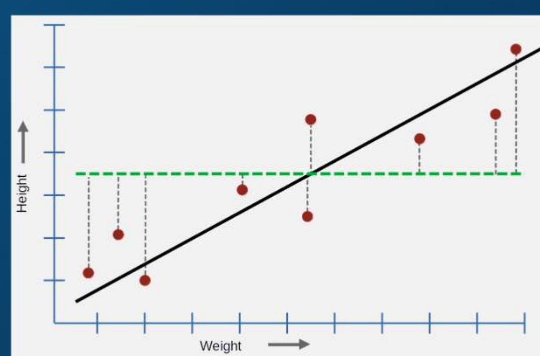
- Imagine you have Height and Weight data
- How do we make prediction given a weight?
- We fit a line through the data



26

Residuals and Squared Error

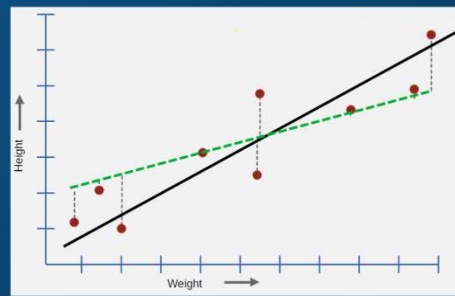
- Lets assume a horizontal line at mean height
- Calculate how far away each point is from this line.
  - The distances are called residuals
- If we simply take the difference, error will cancel out
- Better approach will be to calculate sum of squared errors
  - $\sum(h_i - pred_i)^2$

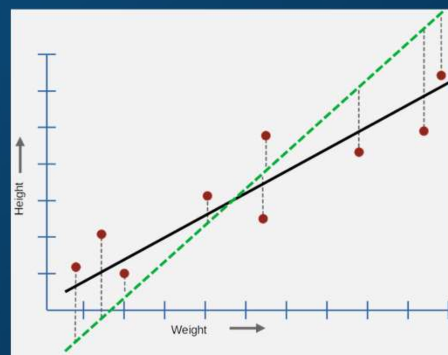


27

Least Squared Error

- If we rotate the line counter clock wise, errors will reduce
- If we keep rotating further the errors will start increasing
- Lets plot residual error w. r. t. rotation
- Method of fitting line is called "Least Squares"



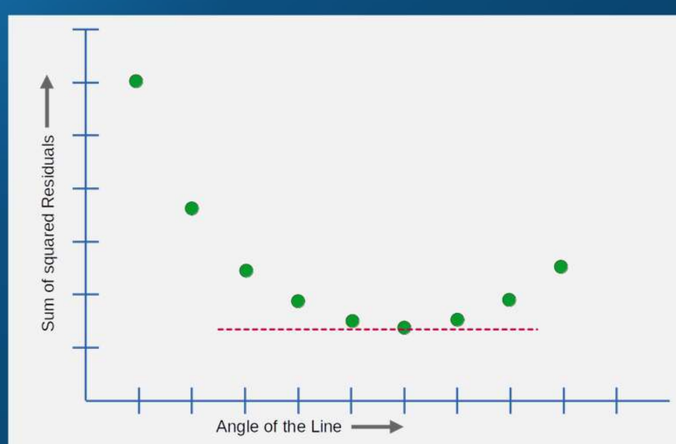


11/8/2025

pra-sâmi

28

Least Squared Error



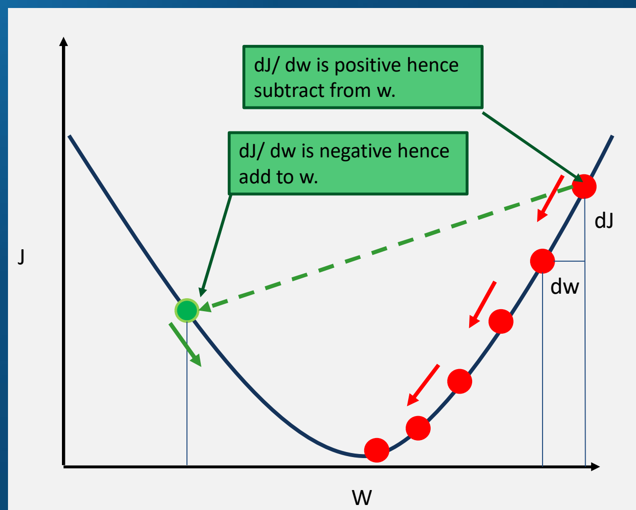
For one particular rotation, the sum of squared errors will be minimum and that's the rotation we are looking for.

11/8/2025

pra-sâmi

29

Gradient Descent

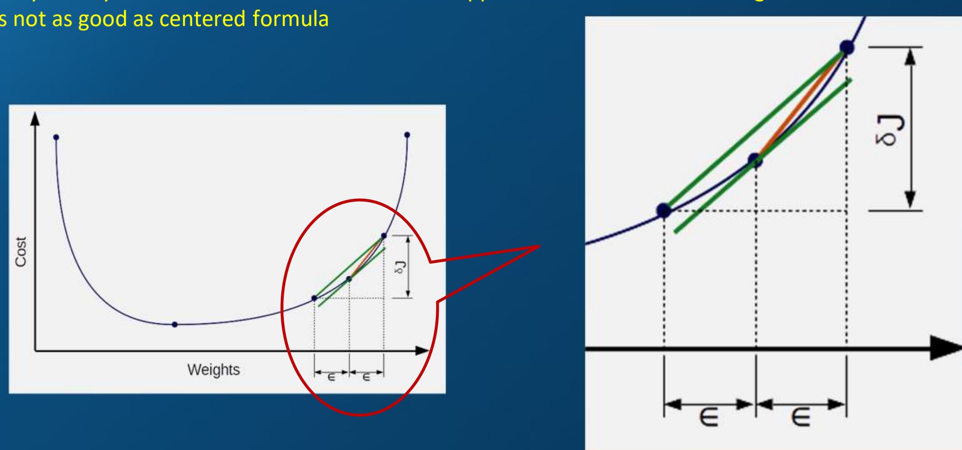


pra-sâmi

30

Calculation of Derivative

- Use the centered formula
  - The formula you may have seen for the finite difference approximation when evaluating the numerical gradient is not as good as centered formula

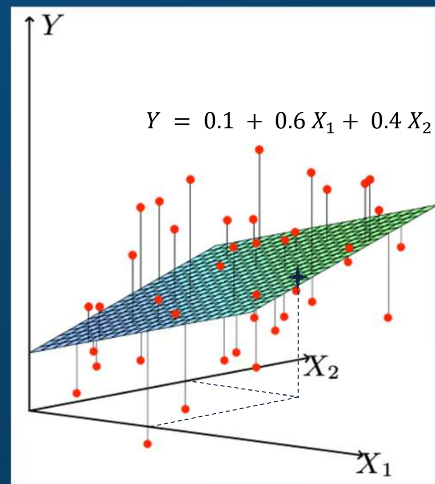


pra-sâmi

31

Linear Regression with Multiple Features

- Imagine that we have one more feature
  - Zodiac Sign
- All our calculations will be same
  - This time we will be fitting a plane instead of line
- If Zodiac sign has no bearing on Height
  - Probably slope in that direction will be zero
- Note: adding extra parameter will not make predictions any worst!
  - Keep this in mind when we talk about Deep Networks

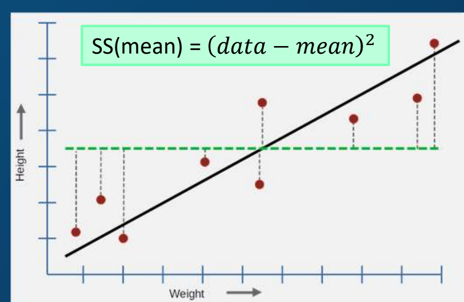


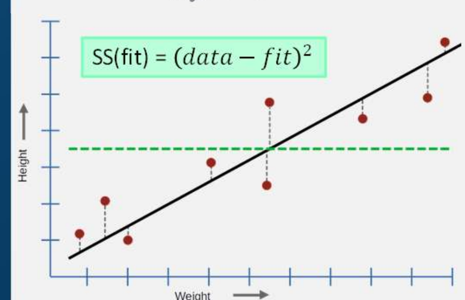
11/8/2025

32

How Good is Fitted Line? - Calculate R^2

- Straight line has two parameters
  - Slope and intercept
- Calculating R^2
  - Sum of squared residuals around mean = SS(mean)
    - Variation around mean = SS(mean) /n
  - Sum of squared residuals around fit = SS(fit)
    - Variation around fit = SS(fit) /n
- R^2 tells us how much of variation in Height can be explained by Weight
  - $R^2 = \frac{Var(mean)-Va\ \ (fit)}{Var(mean)}$ or
  - $R^2 = \frac{SS(mean)-SS(fit)}{SS(mean)}$ in other words
  - $R^2 = \frac{Variation\ in\ Height\ explained\ by\ Weight}{Overall\ variation\ in\ Height}$





11/8/2025

33

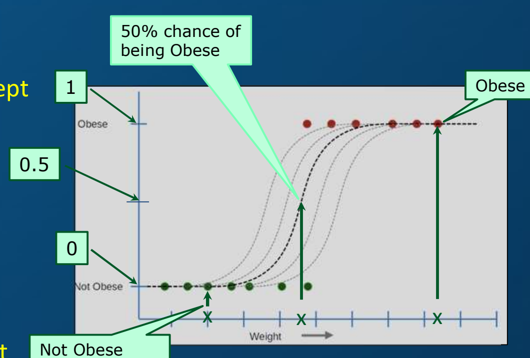
Logistic Regression

pra-sâmi

34

Logistic Regression

- Logistic regression is similar to linear regression, except
  - It's a classification task
  - Predicts class, rather than continuous value
- Logistic function curve goes from 0 to 1
  - Curve tells you if some one is obese based on its weight
- The manner in which this line is fitted is also different
  - Linear Regression → Least squares
  - Logistic Regression → Maximum Likelihood



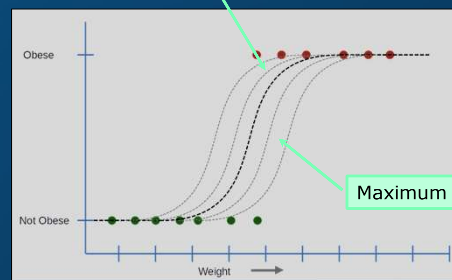
pra-sâmi

35

Maximum Likelihood

- Unlike linear regression, logistic regression fits an ‘S’ shaped “Logistic Function”
  - Curve is for probability that a person is obese based on its weight
- Calculate the likelihood of the observed data - obese or not
- Multiply likelihood of all data
  - Likelihood of the data given this distribution
- Keep shifting the line and calculate likelihood
- Look for the distribution resulting in maximum likelihood

Pick a probability scaled by weight



Maximum likelihood

36

What are Odds anyway!

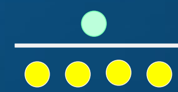
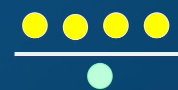
- Odds in favor of team winning are 1 to 4
- Mathematically : $\frac{1}{4}$ ➔
- If the odds are 1 to 100 its even smaller
  - It can go from 1 to 0
- Odds are ratio of some thing happening to something not happening

$$\frac{\text{Something happening}}{\text{Something Not happening}}$$

- Probabilities are ratio of something happening to total events : $\frac{1}{5}$ ➔

37

Odds and Probabilities

- Odds in favor of team winning are 1 to 4 : $\frac{1}{4}$ ➔ 
- Odds in favor of team losing are 4 to 1 : $\frac{4}{1}$ ➔ 
- Probability of my team winning : $\frac{1}{5}$
- Probability of my team losing : $\frac{4}{5}$
- Thus: $Probability_{winning} = 1 - Probability_{losing}$ ➔ $\frac{1}{5} = 1 - \frac{4}{5}$
- And: $\frac{P_{Wining}}{P_{Losing}} = \frac{\frac{1}{5}}{\frac{4}{5}} = \frac{1}{4}$ = Odds in favor of team winning

38

Coefficients

- Odds can be calculated from probabilities : $\frac{p}{(1-p)}$
- At present Y axis varies from 0 to 1
- Let's transform this axis to range from -∞ to +∞
  - From likelihood scale to log(odds of obesity) = log($\frac{p}{1-p}$)
- In short it varies from -∞ to -∞
- With log (odds) on y-axis 'S' curve becomes straight line

The log of probabilities are called Logit Function

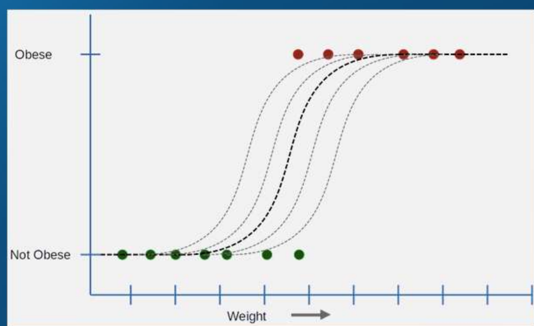
#	p	$Log_e(\frac{p}{1-p})$
1	0.5	0
2	0.731	1
3	0.88	2
4	0.95	3
5	1	∞

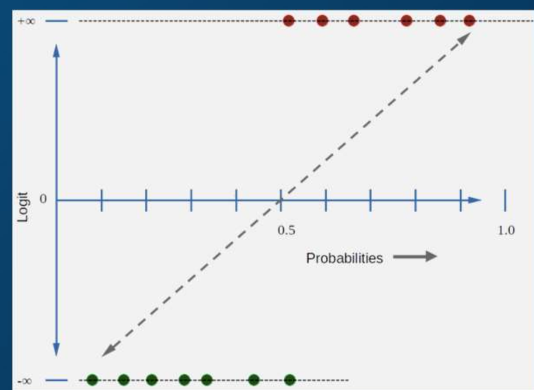
39

Coefficients

With log (odds) on y-axis ‘S’ curve becomes straight line

- Coefficients are presented in terms of log (odds).
- Coefficients are similar to Linear Regression (y = -4.78 + 0.89 * Weight)





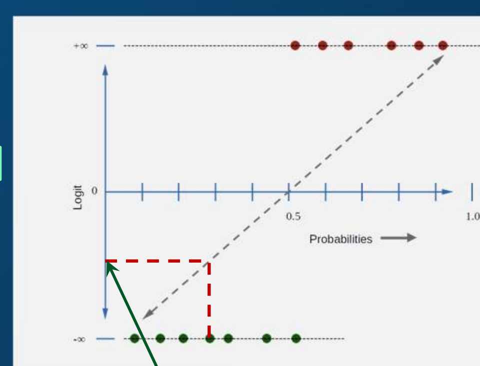
11/8/2025

40

Fitting a Logistic Regression Model via Maximum Likelihood

- Define the Candidate Line
  - Start with a candidate line that represents log(odds) on the y-axis
- Project Data Points:
  - Project each data point onto this candidate line to get its candidate log(odds) value.
- Transform to Probability:
  - Convert each candidate log(odds) into a candidate probability using the logistic function: p = $\frac{e^{\log(odd)}}{1+e^{\log(odd)}}$

‘Sigmoid’



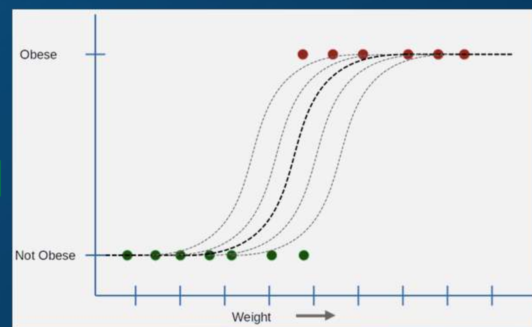
Candidate log (odds) value

11/8/2025

41

Overall Likelihood

- Map to S-Curve:
  - Plot these probabilities on the logistic S-curve. The y-value for each point is its estimated likelihood.
- Calculate Total Likelihood:
  - The probability for "Not Obese" is $1-p_{no}$.
- The likelihood of the entire dataset is the product of these individual probabilities.
  - p_o^1 x p_o^2 x p_o^3 x … p_o^n x (1-p_{no}^1) x (1- p_{no}^2) x (1- p_{no}^3) x … x (1−p_{no}^m)
  - Note : o – obese; no – not obese
- Its better if we calculate log of likelihood
  - log(p_o^1) + log(p_o^2) + log(p_o^3) + … + log(p_o^n) + log(1-p_{no}^1) + log(1- p_{no}^2) + log(1- p_{no}^3) + … + log(1- p_{no}^m)
  - = -2.85 (say)
- Optimize: The best-fitting line is the one that maximizes this total likelihood.



11/8/2025

pra-sâmi

42

McFadden's Pseudo R^2

- Log(likelihood of data given 'S' curve)
  - We need an estimate for something analogous to SS(mean) ➔ without using weight
  - log(p_o^1) + log(p_o^2) + log(p_o^3) + … + log(p_o^n) + log(1-p_{no}^1) + log(1- p_{no}^2) + log(1- p_{no}^3) + … + log(1- p_{no}^m)
  - Call this LL(fit); substitute for SS(fit)
- LL(Overall) is simple
  - Number of sample marked as obese divided by total number of samples: $p^{overall} = \frac{Num_{obese}}{Total}$
  - LL(overall) = log($p^{overall}$) + log($p^{overall}$) + log($p^{overall}$) + … + log($p^{overall}$) + log(1-$p^{overall}$) + log(1- $p^{overall}$) + log(1-$p^{overall}$) + … + log(1-$p^{overall}$)
- For sample size of 9 with 5 obese records; $p^{overall}$ = 5/9 = 0.56
  - LL(overall) = log(0.56) + log(0.56) + log(0.56) + log(0.56) + log(0.56) + log(1-0.56) + log(1-0.56) + log(1-0.56) + log(1-0.56)
  - = -6.18
- $R^2 = \frac{LL(overall) - LL(fit)}{LL(overall)}$

11/8/2025

pra-sâmi

43

Log (likelihood)

- Log(likelihood) =
  - log(p_o^1) + log(p_o^2) + log(p_o^3) + … + log(p_o^n) + log(1-p_{no}^1) + log(1- p_{no}^2) + log(1- p_{no}^3) + … + log(1- p_{no}^m)

Simple log where ground truth is 1

Two different treatment

Log(1-p) where ground truth is 0

- What if we replace
  - log(p_o^1) = y_o^1 * log(p_o^1) + $(1 - \ y_o^1)$ * log$(1 - p_o^1$) log(1-p_{no}^1) = y_{no}^1 * log(1- p_{no}^1) + (1- y_{no}^1) * log(1- p_{no}^1)

Note: ground truth y_o^1 for obese is 1

Note: : ground truth y_{no}^1 for not obese 0

- Both have similar form, hence in general
  - Log(likelihood) = $\sum y^i * \log(p^i) + (1 - y^i) * \log(1 - p^i)$

11/8/2025

pra-sâmi

44

Neurons

11/8/2025

pra-sâmi

45

To play or not to play!

id	Dry Weather	Low Temp	Homework Done	Team Members	Equipment	Ground	Played
1	1	1	1	1	0	1	1
2	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1
4	0	1	0	1	1	1	0
5	0	0	1	1	1	0	0
6	0	0	0	0	0	1	0

11/8/2025

pra-sâmi

46

Neurons

- Features:
  - Is it raining?
  - Is it too hot?
  - Have I completed my homework?
  - Are sufficient players ready?
  - Is cricket equipment ready?
  - Is ground available?
- Depending on the feature values, you may get to play or not
- Features like homework and availability of ground can be considered as ‘inhibitory’.

id	Dry Weather	Low Temp	Homework Done	Team Members	Equipment	Ground	Played
1	1	1	1	1	0	1	1
2	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1
4	0	1	0	1	1	1	0
5	0	0	1	1	1	0	0
6	0	0	0	0	0	1	0

- Notes :
  - Aggregator function is sum and threshold can be 3.
  - Assign 0 or 1 if a parameter is in favor or not

Given sufficient data point, we can train an algorithm to make such simple decisions for us.

11/8/2025

pra-sâmi

47

MP Neuron

- In 1943 Warren S. McCulloch, a neuroscientist, and Walter Pitts, a logician, published "A logical calculus of the ideas immanent in nervous activity" in the Bulletin of Mathematical Biophysics
- In this paper McCulloch and Pitts tried to understand how the brain could produce highly complex patterns by using many basic cells that are connected together
- These basic brain cells are called neurons, and McCulloch and Pitts gave a highly simplified model of a neuron in their paper
- The McCulloch and Pitts model of a neuron, which we will call an MCP neuron for short, has made an important contribution to the development of artificial neural networks -- which model key features of biological neurons

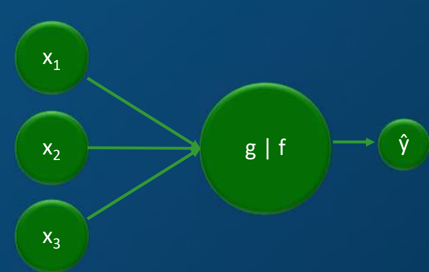
11/8/2025

pra-sâmi

48

MP Neuron

- Neurons receive signals and produce a response
- In this model:
  - All inputs are binary i.e. [0,1]
  - Inputs are “inhibitory” or “excitatory”.
  - Inhibitory have maximum influence on the model
  - It has an aggregator ‘g’ and a function ‘f’
  - There is a threshold
  - If g is more than threshold, ŷ = 1 else 0



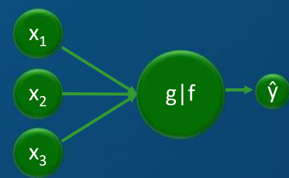
- ŷ = 0 if any x_i is inhibitory, else g(x) = $\sum x_i$
- ŷ = 1 if g(x) ≥ threshold else ŷ = 0

11/8/2025

pra-sâmi

49

MP Neuron



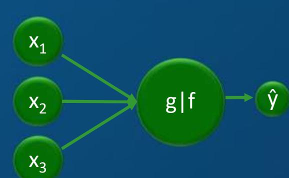
id	Dry Weather	Low Temp	Homework Done	Team Members	Equipment	Ground	Sum	Played
1	1	1	1	1	0	1	5	1
2	1	1	1	1	1	1	6	1
3	1	1	1	1	1	1	6	1
4	0	1	0	1	1	1	4	0
5	0	0	1	1	1	0	3	0
6	0	0	0	0	0	1	1	0

11/8/2025

pra-sâmi

50

MP Neuron



id	Dry Weather	Low Temp	Homework Done	Team Members	Equipment	Ground	Sum	Played
1	1	1	1	1	0	1	5	1
2	1	1	1	1	1	1	6	1
3	1	1	1	1	1	1	6	1
4	0	1	0	1	1	1	4	0
5	0	0	1	1	1	0	3	0
6	0	0	0	0	0	1	1	0

The logic is straight forward. Let's implement this model on a dataset.

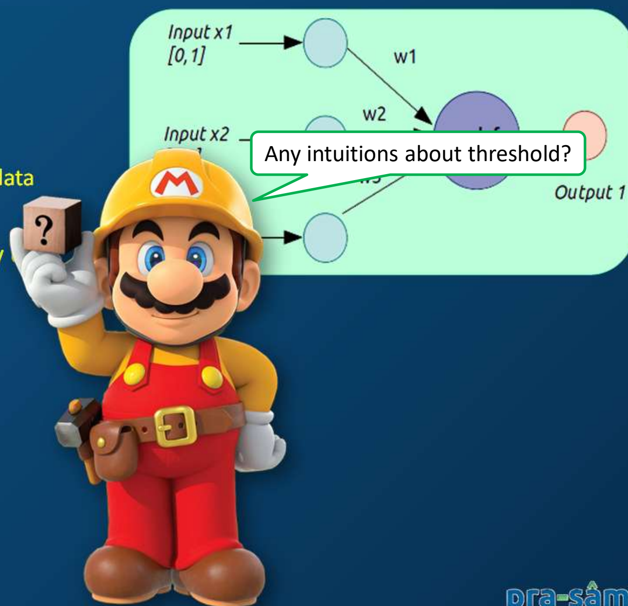
11/8/2025

pra-sâmi

51

Code Example1 – MP Neurons

- Need a dataset with plenty of features and binary output
- Load modified Breast Cancer dataset
  - This dataset is based on scikit-learn breast cancer data
- Its features are a continuous and we need binary
  - For b in range [0, num_features+1]
    - Sum it by row and compare with b
- Converted file is in the shared folder



11/8/2025

52

Next Session

- Perceptron
- Single Layer Neural Network
- Overview of back propagation of errors

11/8/2025

53

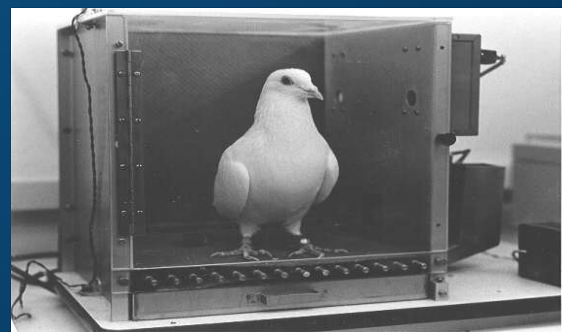
11/8/2025

ADDITIONAL MATERIAL

55

Biological Neural Nets

- Pigeons as art experts (Watanabe et al. 1995)

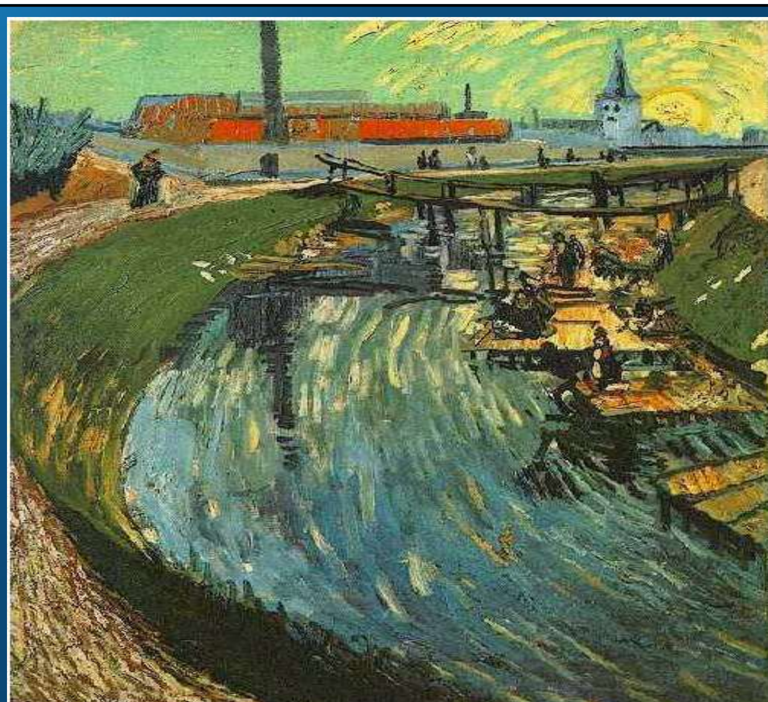


- Experiment:
  - Pigeon in Skinner box
  - Present paintings of two different artists (e.g. Chagall / Van Gogh)
  - Reward for pecking when presented a particular artist (e.g. Van Gogh)

11/8/2025

pra-sâmi

56



11/8/2025

pra-sâmi

58

Biological Neural Nets

- Pigeons were able to discriminate between Van Gogh and Chagall
  - With 95% accuracy on train set (when presented with pictures they had been trained on)
  - Discrimination, still 85% successful for previously unseen paintings of the artists

- Pigeons do not simply memorise the pictures
- They can extract and recognise patterns (the 'style')
- They generalise from the already seen to make predictions

- This is what neural networks (biological and artificial) are good at (unlike conventional computer)

59

Brain and Machine

- The Brain
  - Pattern Recognition
  - Association
  - Complexity
  - Noise Tolerance

- The Machine
  - Calculation
  - Precision
  - Logic

11/8/2025

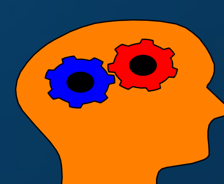
pra-sâmi

60

The contrast in architecture

- The Von Neumann architecture uses a single processing unit;
  - Tens of millions of operations per second
  - Absolute arithmetic precision

- The brain uses many slow unreliable processors acting in parallel



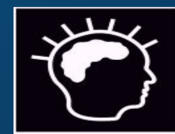
11/8/2025

pra-sâmi

61

The biological inspiration

- Features of the Brain
  - Ten billion (10^{10}) neurons
  - On average, several thousand connections
  - Hundreds of operations per second
  - Die off frequently (never replaced)
  - Compensates for problems by massive parallelism



- The brain has been extensively studied by scientists
- Vast complexity prevents all but rudimentary understanding
- Even the behavior of an individual neuron is extremely complex
- Single “percepts” distributed among many neurons
- Localized parts of the brain are responsible for certain well-defined functions (e.g. vision, motion).

62

Journey So far

Search → Search and Inference → Uncertainty → Bayesian Network

Listen → Reflect → Implement → Share

63

Uncertainty Everywhere

Goal:
Delivering a passenger to the airport on time

- The agent forms a plan, lets say… A90,
  - Leave home 90 minutes before the flight departs
  - Driving at a reasonable speed
- Are you certain *"Plan A90 will get us to the airport in time."*?
  - Not in absolute sense but with some riders

11/8/2025 pra-sâmi

64

Uncertainty Everywhere

Goal:
Delivering a passenger to the airport on time

- How about other plans, such as A180,
  - Might increase the agent's belief that it will get to the airport on time,
  - But also increase the likelihood of a long wait
- Probability is an agent's measure of belief in some proposition — subjective probability.
- An agent's belief depends on its prior belief and what it observes.

11/8/2025 pra-sâmi

65

Sample Formats

Edward L. Thorndike, 1898

Responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a discomforting effect become less likely to occur again in that situation.

Sapphire Blue: hex code #0F52BA,
Emerald : hex code #50C878,
Sapphire Yellow: hex code is #CDC7B4

11/8/2025

pra-sâmi