Deep Neural Network

AIML Module 1

BITS Pilani

BITS Pilani

Pilani | Dubai | Goa | Hyderabad

The Lecture Materials gratefully acknowledge the authors who made their course materials freely available online.

Course Logistics

What we Learn…. (Module Structure)

1. Fundamentals of Neural Network
2. Multilayer Perceptron
3. Deep Feedforward Neural Network
4. Improve the DNN performance by Optimization and Regularization
5. Convolutional Neural Networks
6. Sequence Models
7. Attention Mechanism
8. Neural Network Search
9. Time series Modelling and Forecasting
10. Other Learning Techniques

Books

Textbook

- Dive into Deep Learning by Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. https://d2l.ai/chapter_introduction/index.html

Reference book

- Deep Learning by Ian Goodfellow, Yoshua Bengio, Aaron Courville https://www.deeplearningbook.org/

Evaluation Components and Schedule

	Evaluation Component	Mode	Duration	Weightage	Date and Time
EC1	Quiz I*	Online	30 Mins	5%	Refer to LMS
	Quiz II*		30 Mins	5%	
	Assignment 01		4 Weeks	10%	
	Assignment 02		4 Weeks	15%	
EC2	Mid-sem exam (Closed book)		2 Hours	30%	
EC3	Comprehensive exam (Open book)		2.5 Hours	40%	

* Best of Two

Lab Sessions

- L1 Introduction to PyTorch
- L2 Deep Neural Network with Back-propagation and optimization
- L3 CNN
- L4 RNN
- L5 LSTM
- L6 Auto-encoders

Course Logistics

- Refer Taxila for the following
    - Handout
    - Schedule for Webinar
    - Schedule of Quiz, and Assignments.
    - Evaluation scheme
    - Session Slide Deck
    - Demo Lab Sheets
    - Quiz-I, Quiz-II
    - Assignment-I, Assignment-II
    - Sample QPs
- Lecture Recordings
    - Available on Microsoft teams

Honour Code

All submissions for graded components must be the result of your original effort. It is strictly prohibited to copy and paste verbatim from any sources, whether online or from your peers. The use of unauthorized sources or materials, as well as collusion or unauthorized collaboration to gain an unfair advantage, is also strictly prohibited. Please note that we will not distinguish between the person sharing their resources and the one receiving them for plagiarism, and the consequences will apply to both parties equally.

In cases where suspicious circumstances arise, such as identical verbatim answers or a significant overlap of unreasonable similarities in a set of submissions, will be investigated, and severe punishments will be imposed on all those found guilty of plagiarism.

In case of Queries regarding the course….

Step 1: Post in the discussion forum.

- Read through the existing post and if you find any topic similar to your concern, add on to the existing discussion.
- Avoid duplication of queries or issues.

Step 2: Email the IC at bharatesh.c@wilp.bits-pilani.ac.in if the query or issue is not resolved within 1 week. Turn around for a response to the email is 48 hours.

- In the subject pl mention the phrase ”DNN” clearly.
- Use BITS email id for correspondence. Emails from personal emails will be ignored without any reply.

PATIENCE is highly APPRECIATED

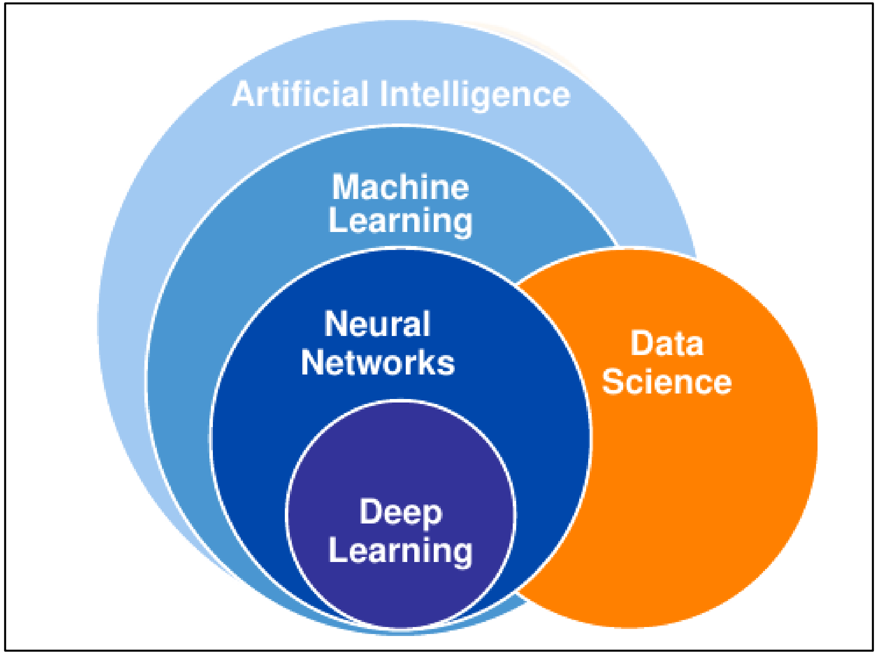
What is Deep Learning?

Definitions of Deep Learning

- Deep Learning is a type of **machine learning** based on **artificial neural networks** in which multiple layers of processing are used to extract progressively higher-level features from data.
- Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is **inspired by the human brain**.
- Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: **learn by example**.
- Deep learning is a subset of machine learning, which is essentially a neural network with **three or more layers**.
- Deep Learning gets its name from the fact that we add more "**Layers**" to learn from the data.

Where in AI sits DL?

AI is a general field that encompasses machine learning and deep learning, but that also includes many more approaches that don't involve any learning.



AI - ML - DL

AI : Artificial intelligence is the science of making things smart. The aim is make machines perform human tasks. Eg: Robot cleaning a room.

ML : Machine learning is an approach to AI. The machine learns or perform tasks through learning by experience.

DL : Deep Learning is a technique for implementing machine learning to recognise patterns.

Deep (Machine) Learning

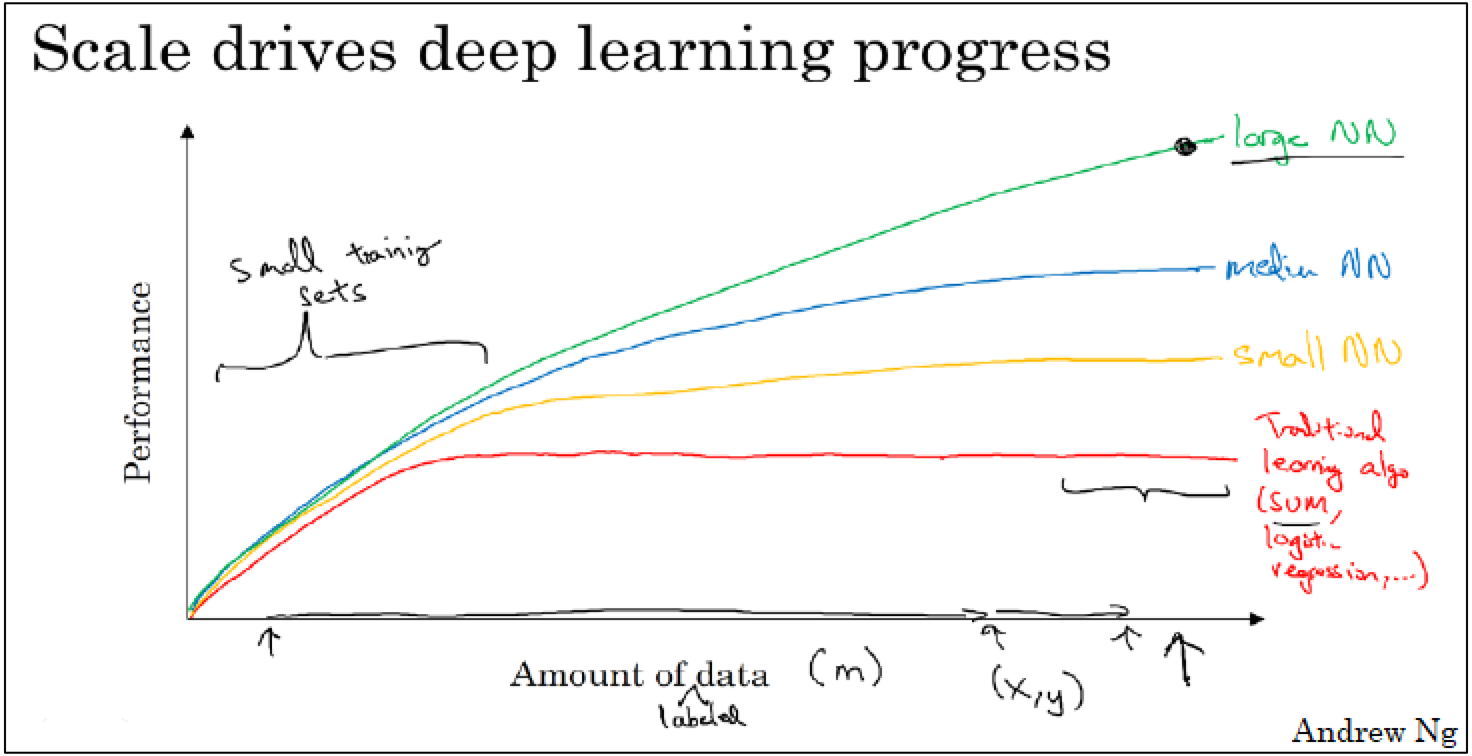
- Deep learning is a specific subfield of machine learning.
- Learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations.
- The **deep** in deep learning stands for this idea of successive layers of representations.
- The number of layers that contribute to model the data is called the **depth** of the model.
- In deep learning, the layered representations are learned via models called **neural networks**, structured in literal layers stacked on top of each other.

Why Deep Learning?

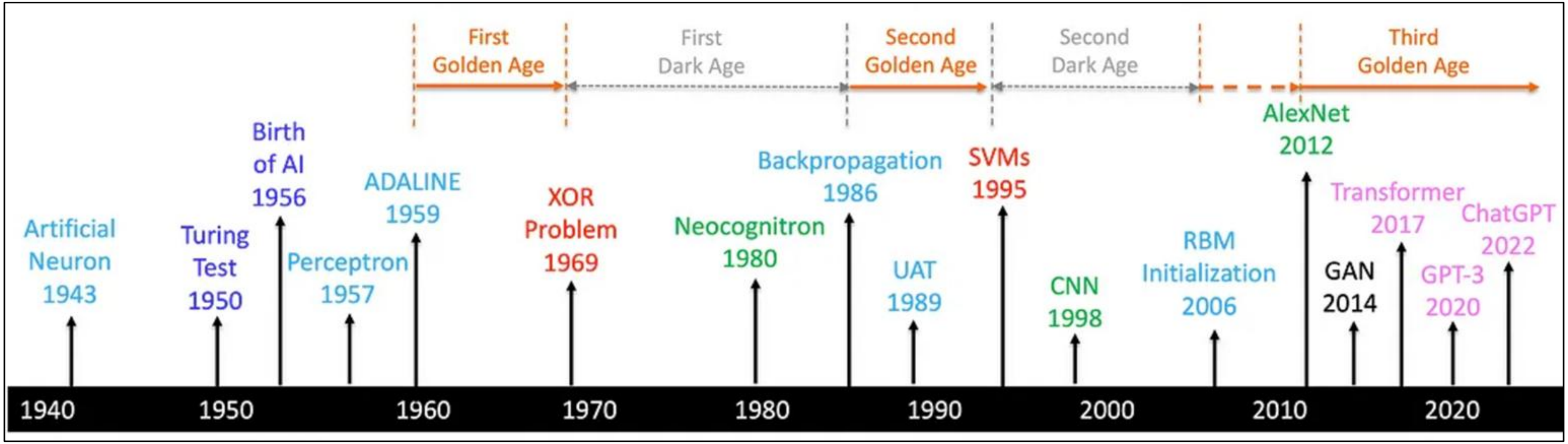
Why Deep Learning?

- Large amounts of data
- Lots and lots of unstructured data like images, text, audio, video
- Cheap, high-quality sensors
- Cheap computation - CPU, GPU, Distributed clusters
- Cheap data storage
- Learn by examples
- Automated feature generation
- Better learning capabilities
- Scalability
- Advance analytics can be applied

Why deep learning?



Deep **Learning** Timeline



Applications of Deep Learning

TECHNEWSWORLD EMERGING TECH SEARCH

Computing Internet IT Mobile Tech Reviews Security Technology Tech Blog Reader Service

Microsoft AI Beats Humans at Speech Recognition

By Richard Adhikari
Oct 20, 2016 11:40 AM PT

Print
Email

Image: Adobe Stock

Microsoft's Artificial Intelligence and Research Unit earlier this week reported that its speech recognition technology had surpassed the performance of human transcriptionists.

Most Popular Newsletters News Aler

How do you feel about Black Friday and Cyber Monday?

- They're great -- I get a lot of bargains!
- The deals are too spread out -- I'd prefer just one day.
- They're a fun way to kick off the holiday season.
- I don't like the commercialization of Thanksgiving Day.
- They're crucial for the retail industry and the economy.
- The deals typically aren't that good.

Vote to See Results

E-Commerce Times

Black Friday Shoppers Hungry for New Experiences, New Tech

Pay TV's Newest Innovation: Giving Users Control

Apple Celebrates Itself in $300 Coffee Table Tome

AWS Enjoys Top Perch in IaaS, PaaS Markets

US Comptroller Gears Up for Blockchain and

The Keyword Latest Stories Product News Topics

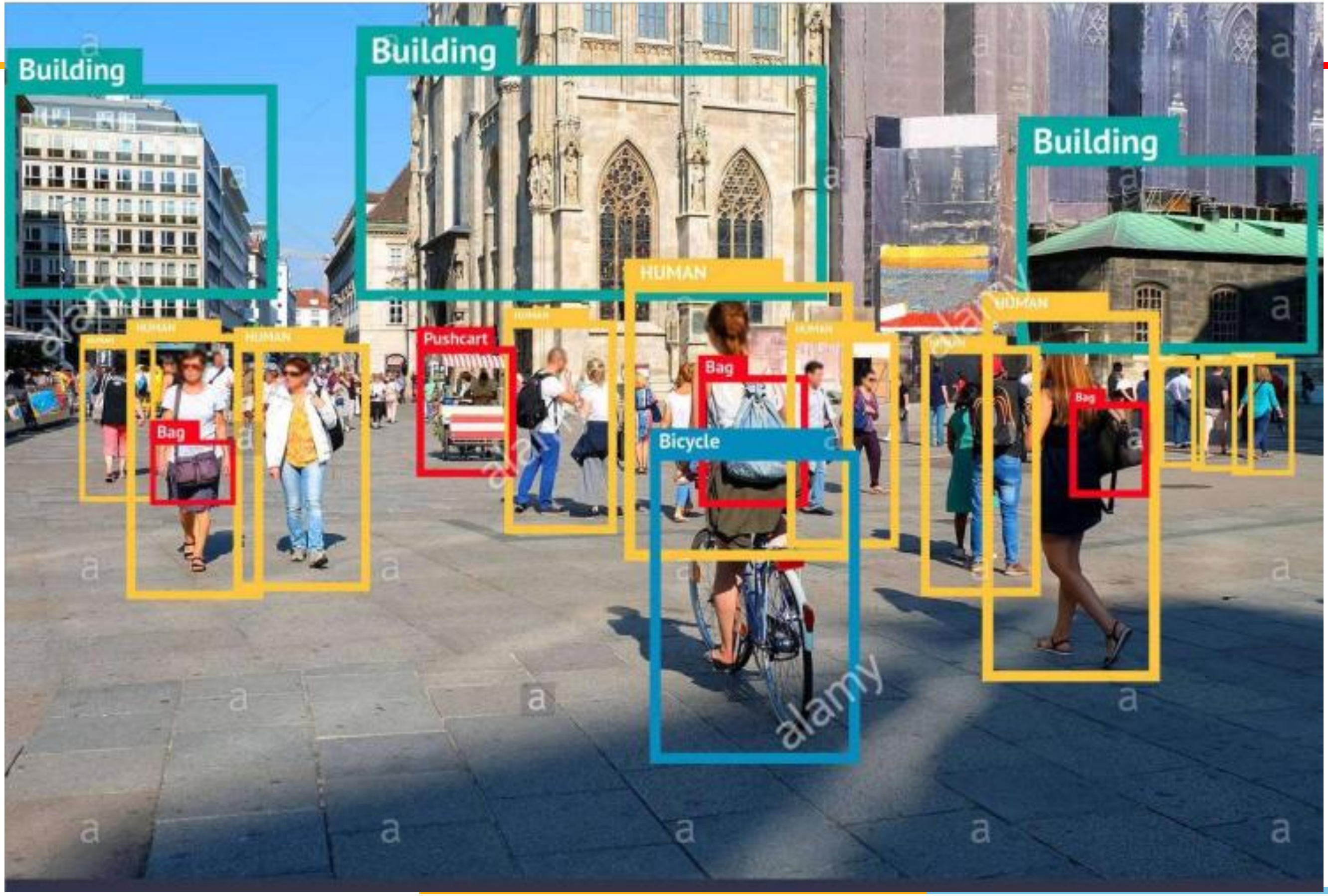
TRANSLATE NOV 15, 2016

Found in translation: More accurate, fluent sentences in Google Translate

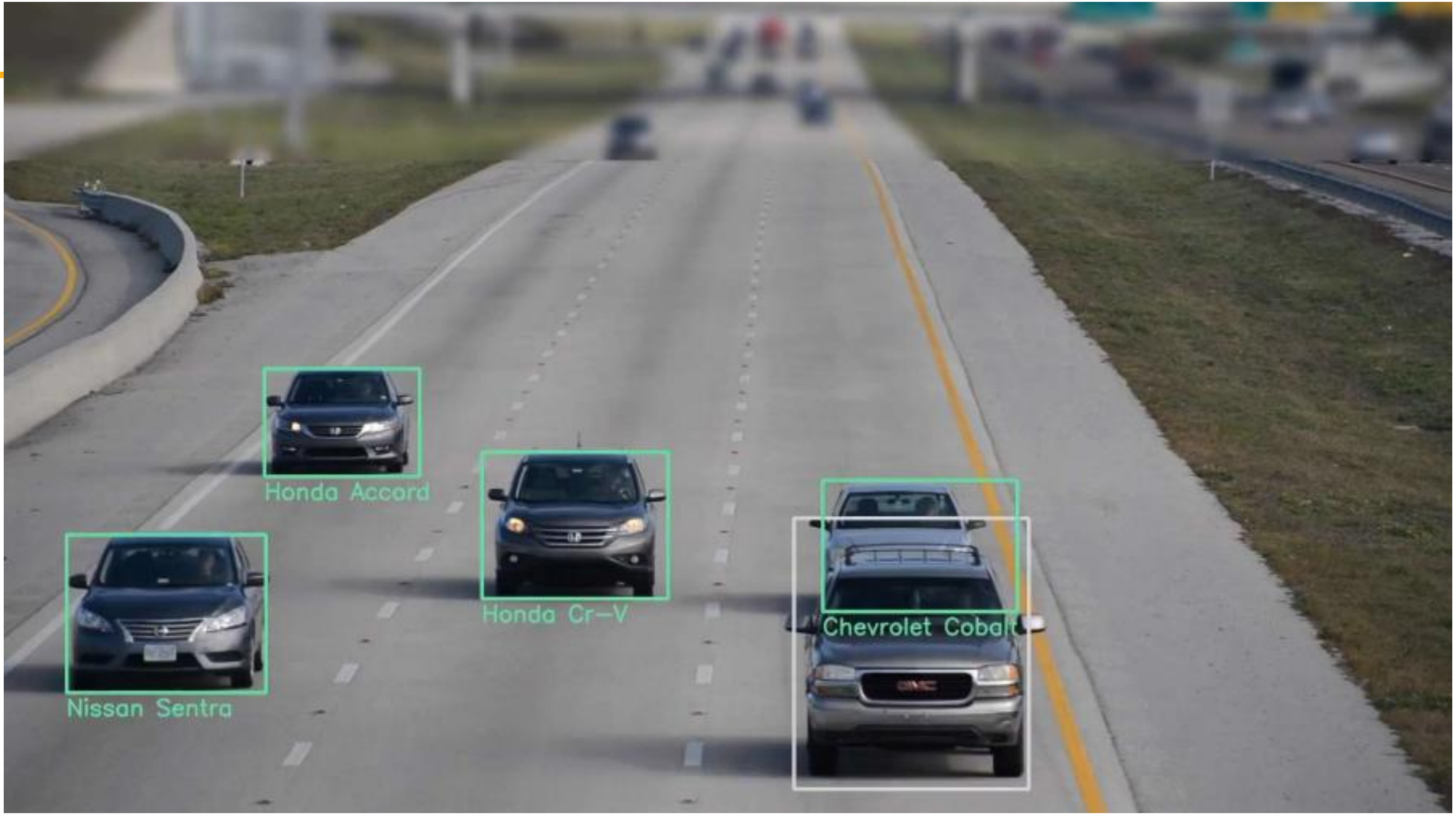
Barak Turovsky
PRODUCT LEAD, GOOGLE TRANSLATE

In 10 years, Google Translate has gone from supporting just a few languages to 103, connecting strangers, reaching across language barriers and even helping

Breakthroughs with Neural Networks



Breakthroughs with Neural Networks



Breakthroughs with Neural Networks

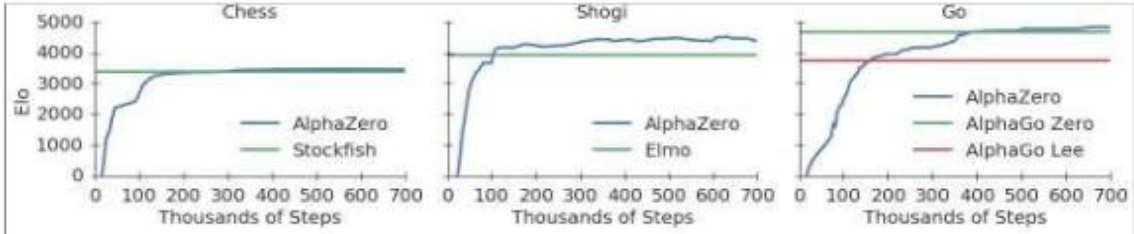


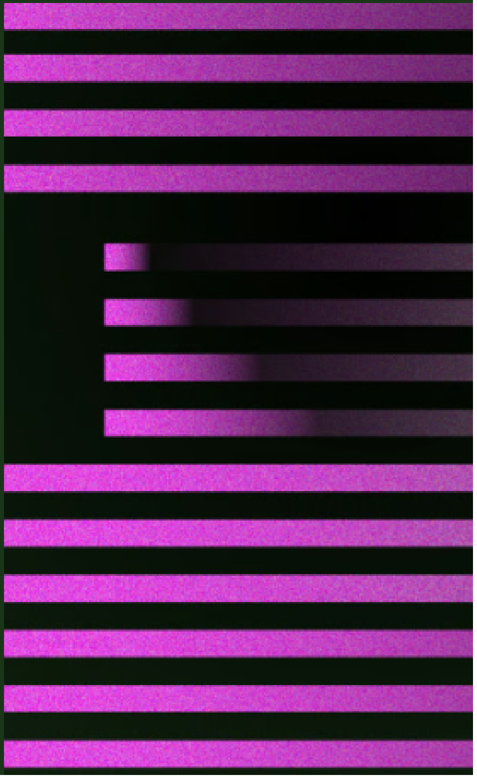
Figure 1: Training *AlphaZero* for 700,000 steps. Elo ratings were computed from evaluation games between different players when given one second per move. **a** Performance of *AlphaZero* in chess, compared to 2016 TCEC world-champion program *Stockfish*. **b** Performance of *AlphaZero* in shogi, compared to 2017 CSA world-champion program *Elmo*. **c** Performance of *AlphaZero* in Go, compared to *AlphaGo Lee* and *AlphaGo Zero* (20 block / 3 day) (29).

Breakthroughs with Neural Networks

Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

Try ChatGPT ↗   Read about ChatGPT Plus



Applications of Deep Learning

Application	Input	Output	Neural Network
Real Estate	House features	House Price	Std NN
Photo Tagging	Image	Text	CNN
Object detection	Image	Bounding box	CNN
Speech Recognition	Audio	Text transcript	RNN
Translation	English Text	French Text	RNN
Autonomous driving	Image, Sensors Radars	Position of other cars, objects, signals	Hybrid NN

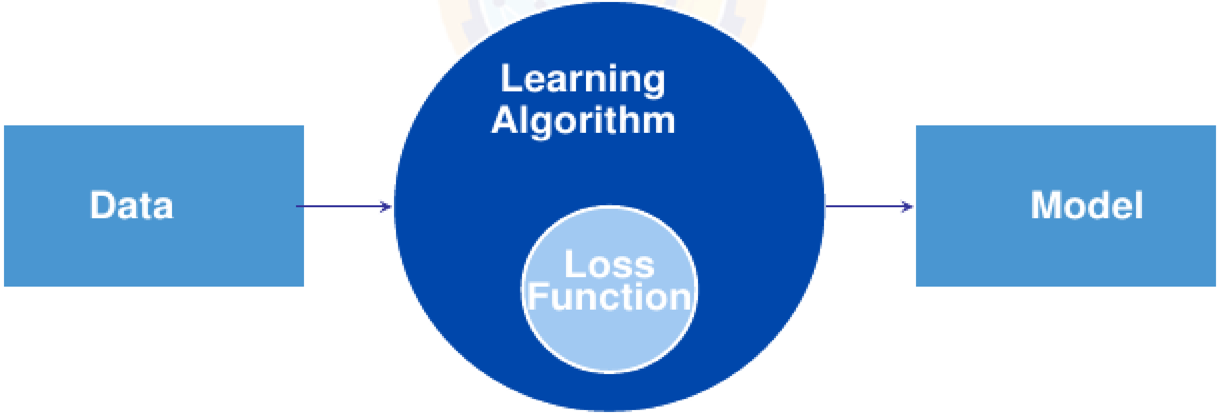
Many more applications….

- a program that predicts tomorrow's weather given geographic information, satellite images, and a trailing window of past weather.
- a program that takes in a question, expressed in free-form text, and answers it correctly.
- a program that given an image can identify all the people it contains, drawing outlines around each.
- a program that presents users with products that they are likely to enjoy but unlikely, in the natural course of browsing, to encounter.

Key components of DL problem

Core components of DL problem

1. The data that we can learn from.
2. A model of how to transform the data.
3. An objective function that quantifies how well (or badly) the model is doing.
4. An algorithm to adjust the model's parameters to optimize the obje



1. Data

- Collection of examples.
- Data has to be converted to an useful and a suitable numerical **representation**.
- Each example (or data point, data instance, sample) typically consists of a set of attributes called **features** (or covariates), from which the model must make its predictions.
- In the supervised learning problems, the attribute to be predicted is designated as the **label** (or target).
- Mathematically, a set of *m* examples,
- We need **right** data.

$$Data = \mathcal{D} = \{X, t\}$$

1. Data

- **Dimensionality** of data
  - Each example has the same number of numerical values. This data consist of fixed-length vectors. Eg: Image
  - The constant length of the vectors as the dimensionality of the data.
  - Text data has varying-length data.

2. Model

- Model denotes the **computational machinery** for ingesting data of one type, and spitting out predictions of a possibly different type.
- Deep learning models consist of many successive transformations of the data that are chained together top to bottom, thus the name deep learning.

Fig. 1.1.1: Identify a wake word.

3. Objective Function

- **Learning means improving at some task over time.**
- A formal mathematical system of learning machines is defined using formal measures of how good (or bad) the models are. These formal measures are called as **objective functions**.
- By convention, objective functions are defined so that lower is better.
- Because lower is better, these functions are sometimes called **loss functions**.

3. Loss Functions

- To predict numerical values (regression), the most common loss function is squared error.
- For classification, the most common objective is to minimize error rate, i.e., the fraction of examples on which our predictions disagree with the ground truth.

3. Loss Functions

- Loss function is defined **with respect to the model parameters** and **depends upon the dataset**.
- We learn the best values of our model parameters by **minimizing the loss** incurred on a set consisting of some number of examples collected for training. However, doing well on the training data does not guarantee that we will do well on unseen data. I.e **Model has to generalize better.**
- When a model performs well on the training set but fails to generalize to unseen data, we say that it is **overfitting**.

4. Optimization Algorithms

- Optimization Algorithm is an algorithm capable of **searching for the best possible parameters** for **minimizing the loss function**.
- Popular optimization algorithms for deep learning are based on an approach called **gradient descent**.

Example of the Framework

- We have to tell a computer explicitly how to map from inputs to outputs.
- We have to define the problem precisely, pinning down the exact nature of the inputs and outputs, and choosing an appropriate model family.
- Collect a huge dataset containing examples of audio and label those that do and that do not contain the wake word.

Fig. 1.1.1: Identify a wake word.

Example of the Framework

- Create a Model
  - Define a flexible program whose behavior is determined by a number of parameters.
  - To determine the best possible set of parameters, use the data. The parameters should improve the performance of the program with respect to some measure of performance on the task of interest.
  - After fixing the parameters, we call the program a model.
    - Eg: The model receives a snippet of audio as input, and the model generates a selection among yes, no as output.
  - The set of all distinct programs (input-output mappings) that we can produce just by manipulating the parameters is called a family of models.
    - Eg: We expect that the same model family should be suitable for ”Alexa” recognition and ”Hey Siri” recognition because they seem, intuitively, to be similar tasks.

Example of the Framework

- The meta-program that uses our dataset to choose the parameters is called a learning algorithm.
- In machine learning, the learning is the process by which we discover the right setting of the parameter coercing the desired behavior from our model.
- Train the model with data.

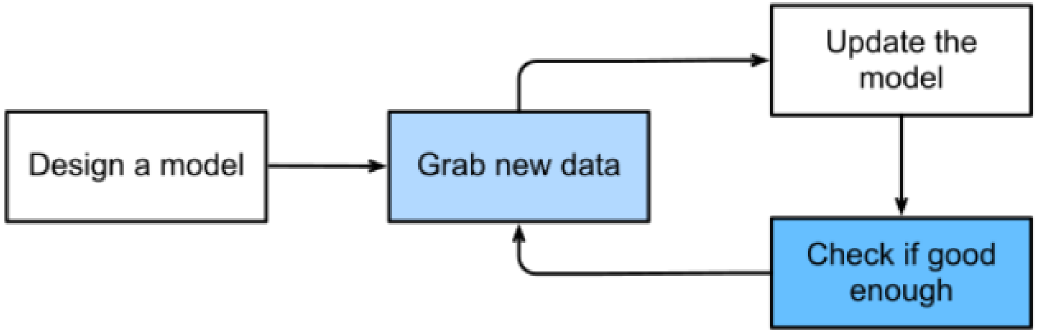


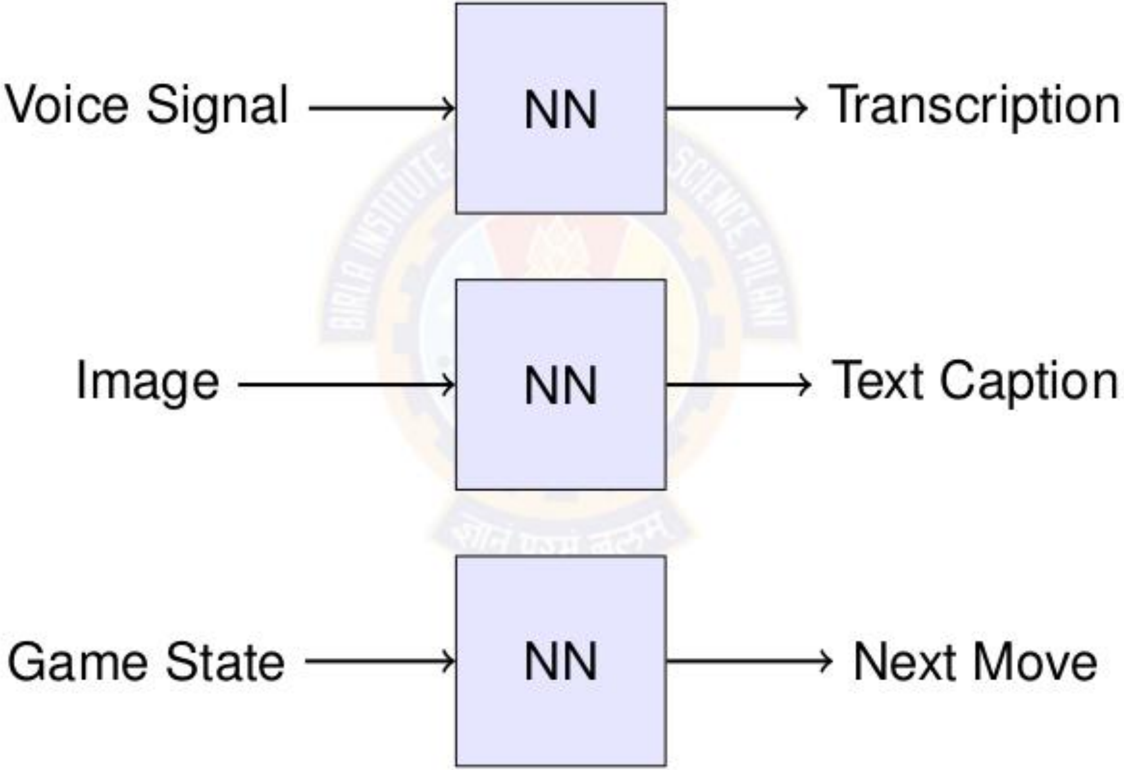
Fig. 1.1.2: A typical training process.

Reading from TB Dive into Deep Learning

- Chapter 1
- Chapter 2 for Python Prelims, Linear Algebra, Calculus, Probability

Artificial Neural Network

What are Neural Networks?



What are Neural Networks?

- It begins with human brain.

- Humans learn, solve problems, recognize patterns, create, think deeply about something, meditate and many many more.....
- Humans learn through **association**. [Refer to Associationism for more details.]

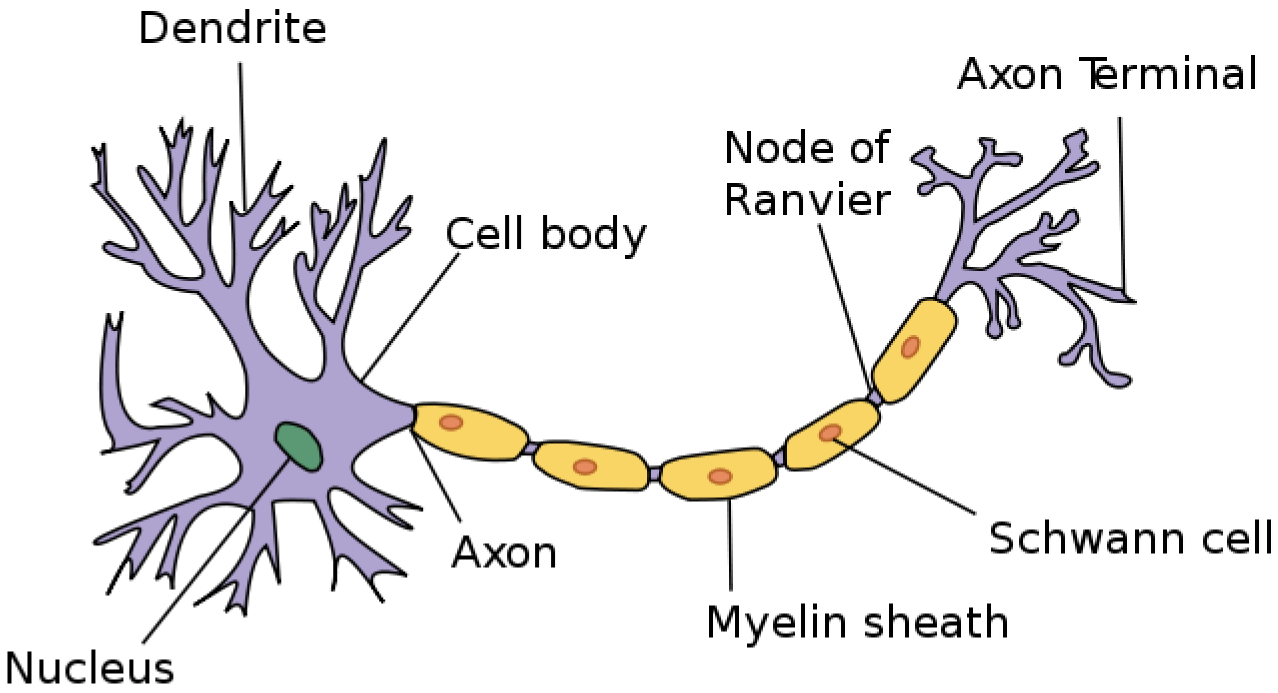
Observation: The Brain

- The brain is a mass of interconnected neurons.
- Number of neurons is approximately 10^(10).
- Connections per neuron is approximately 10^(4 to 5).
- Neuron switching time is approximately 0.001 second.
- Scene recognition time is 1 second.
- 100 inference steps doesn't seem like enough. Lot of parallel computation.

Brain: Interconnected Neurons

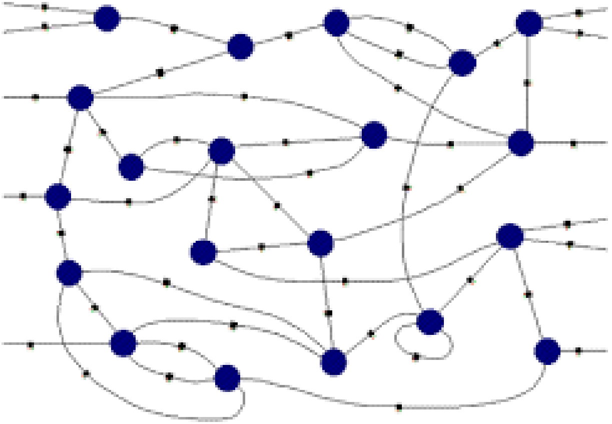
- Many neurons connect in to each neuron.
- Each neuron connects out to many neurons.

Biological Neuron



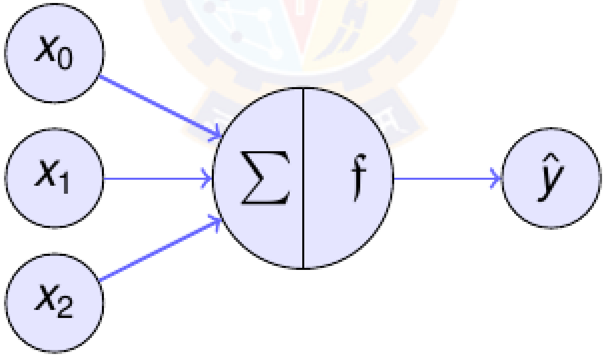
Connectionist Machines

- Network of processing elements, called artificial neural unit.
- The neurons are interconnected to form a network.
- All world knowledge is stored in the connections between these elements.
- Neural networks are connectionist machines.



What are Artificial Neurons?

- Neuron is a processing element inspired by how the brain works.
- Similar to biological neuron, each artificial neuron will be do some computation. Each neuron is interconnected to other neurons.
- Similar to brain, the interconnections between neurons store the knowledge it learns. The knowledge is stored as parameters.



Properties of Artificial Neural Nets (ANNs)

- Many neuron-like threshold switching units.
- Many weighted interconnections among units.
- Highly parallel, distributed process.
- Emphasis on tuning parameters or weights automatically.

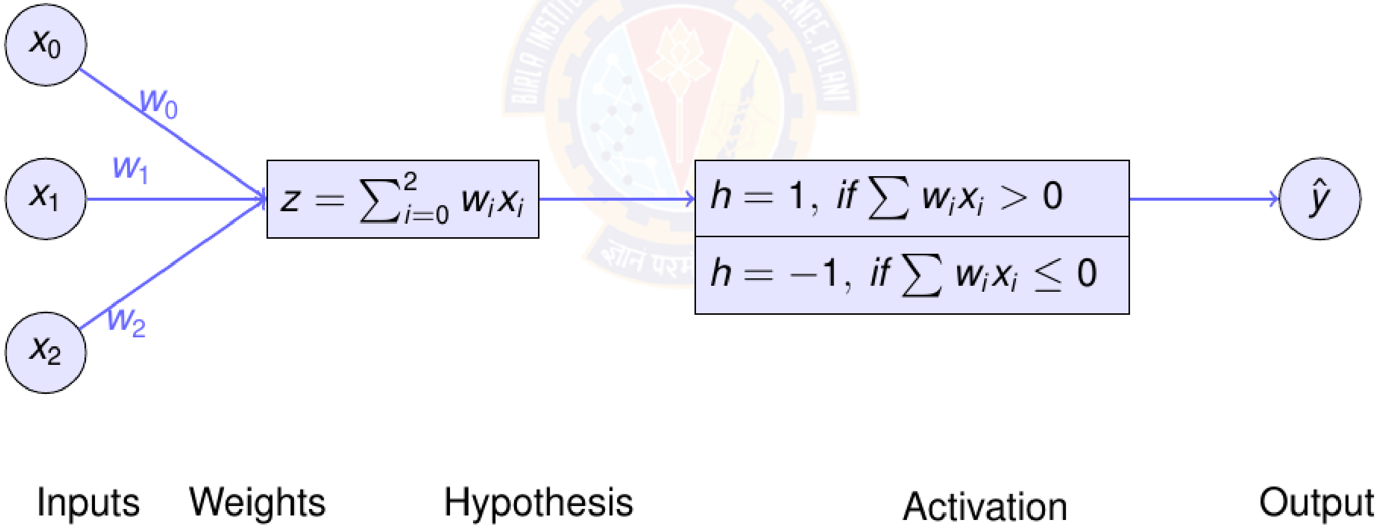
When to consider Neural Networks?

- Input is high-dimensional discrete or real-valued (e.g. raw sensor input).
- Possibly noisy data. Data has lots of errors.
- Output is discrete or real valued or a vector of values.
- Form of target function is unknown.
- Human readability, in other words, explainability, of result is unimportant.
- Examples:
  - Speech phoneme recognition
  - Image classification
  - Financial prediction

Perceptron

Perceptron

- One type of ANN system is based on a unit called a perceptron.
- **A perceptron takes a vector of real-valued inputs, calculates a linear combination of these inputs, and then outputs a 1 if the result is greater than some threshold and -1 otherwise.**



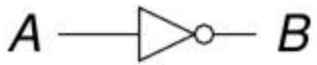
Representing Logic Gates using Perceptron

- **Perceptron** can represent all of the primitive Boolean functions AND, OR, NAND, and NOR – Linearly separable data
- some Boolean functions cannot be represented by a single perceptron, such as the XOR function – Linearly nonseparable data
- Perceptron learning rule finds a successful weight vector when the training examples are linearly separable, **it can fail to converge if the examples are not linearly separable.**

NOT Logic Gate

Question:

- How to represent NOT gate using a perceptron?
- What are the parameters for the NOT perceptron?
- Data is given below.



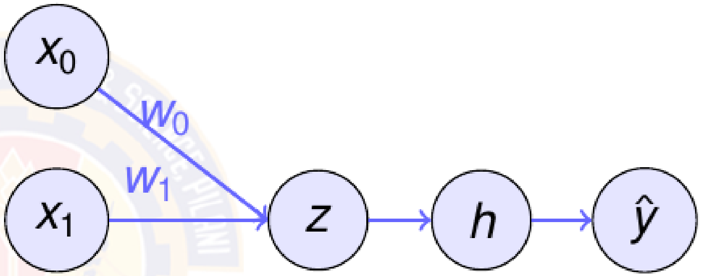
A	B
0	1
1	0

Rewrite as

x_1	t
-1	1
1	-1

Perceptron for NOT gate

- Perceptron equation is $\hat{y} = w_0 x_0 + w_1 x_1$.
- $x_0 = 1$ always.
- $h > 0$ for output to be 1.
- For each row of truth table, the equations are given.
- One solution is $w_0 = 1$ and $w_1 = -1$. (Intuitive solution)
- This give a beautiful linear decision boundary.



x_1	t_1	h	h
-1	1	$w_0 x_0 + w_1(-1) > 0$	$w_0 - w_1 > 0$
1	-1	$w_0 x_0 + w_1(1) < 0$	$w_0 + w_1 < 0$

AND Logic Gate

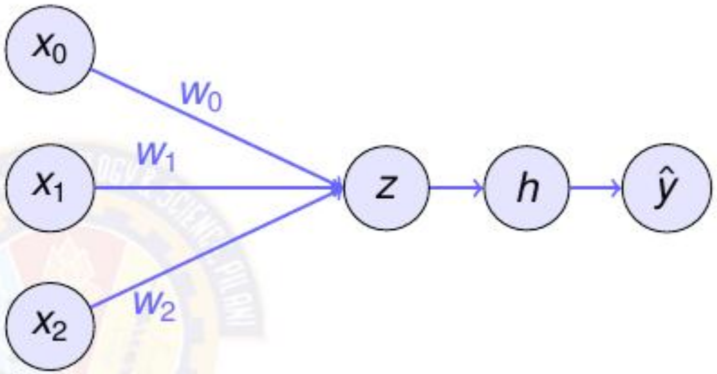
Question:

- How to represent AND gate using a perceptron?
- What are the parameters for the AND perceptron?
- Data is given below.

A	B	Out
0	0	0
0	1	0
1	0	0
1	1	1

Perceptron for AND gate

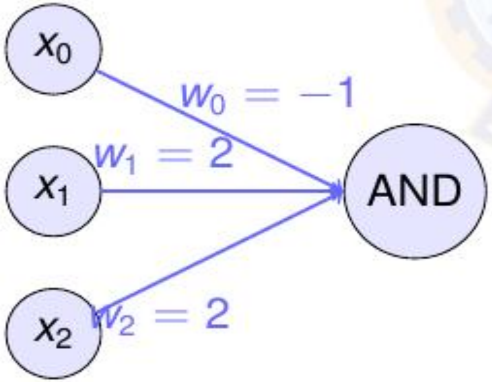
- Perceptron equation is $\hat{y} = w_0 x_0 + w_1 x_1 + w_2 x_2$.
- $h > 0$ for output to be 1.
- For each row of the truth table, the equations are given.
- Solve for the inequalities.
- One solution is $w_1 = w_2 = 2$, $w_0 = (-1)$.
- This give a beautiful linear decision boundary.



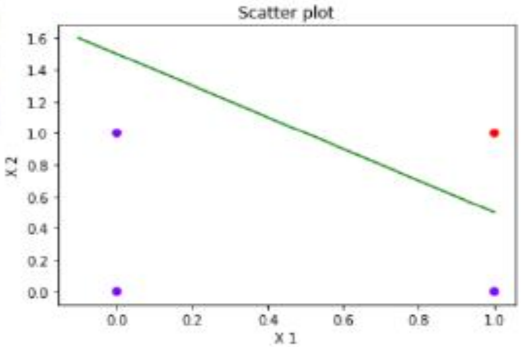
x_1	x_2	t	h
-1	-1	-1	$w_0 + w_1(-1) + w_2(-1) < 0$
-1	1	-1	$w_0 + w_1(-1) + w_2(1) < 0$
1	-1	-1	$w_0 + w_1(1) + w_2(-1) < 0$
1	1	1	$w_0 + w_1(1) + w_2(1) > 0$

Perceptron for AND gate

A	B	AB
0	0	0
0	1	0
1	0	0
1	1	1



DECISION BOUNDARY

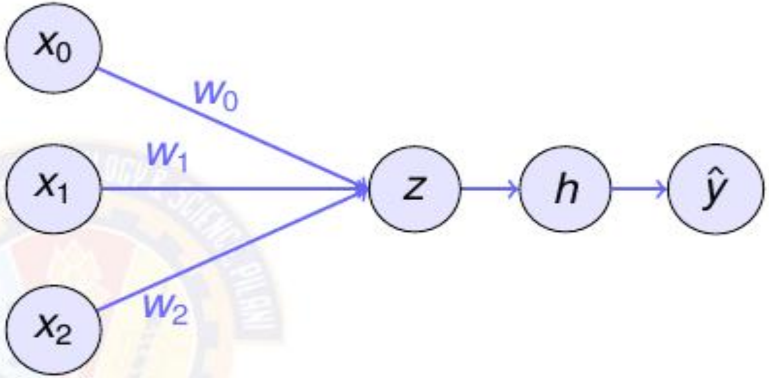


Exercise

1. Represent OR gate using Perceptron. Compute the parameters of the perceptron.
2. Represent NOR gate using Perceptron. Compute the parameters of the perceptron.
3. Represent NAND gate using Perceptron. Compute the parameters of the perceptron.

Perceptron for OR gate

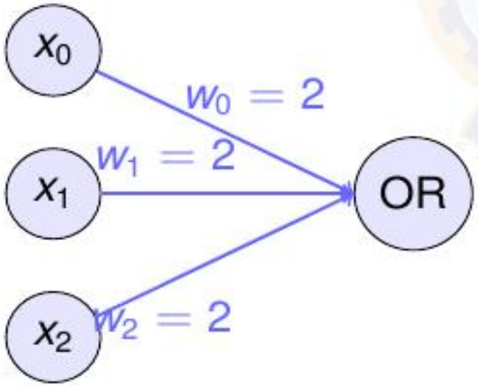
- Perceptron equation is $\hat{y} = w_0 x_0 + w_1 x_1 + w_2 x_2$.
- $h > 0$ for output to be 1.
- For each row of the truth table, the equations are given.
- Solve for the inequalities.
- One solution is $w_0 = w_1 = w_2 = 2$.
- This give a beautiful linear decision boundary.



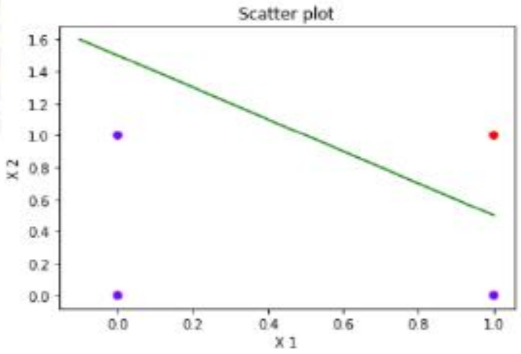
x_1	x_2	t	h
-1	-1	-1	$w_0 + w_1(-1) + w_2(-1) < 0$
-1	1	-1	$w_0 + w_1(-1) + w_2(1) > 0$
1	-1	-1	$w_0 + w_1(1) + w_2(-1) > 0$
1	1	1	$w_0 + w_1(1) + w_2(1) > 0$

Perceptron for OR gate

A	B	AB
0	0	0
0	1	1
1	0	1
1	1	1



DECISION BOUNDARY



Perceptron Learning Algorithm

Perceptron Learning Algorithm

$$w_i \leftarrow w_i + \Delta w_i$$

where

$$\Delta w_i = \eta(t - o)x_i$$

Where:

- $t = c(\vec{x})$ is target value
- o is perceptron output
- η is small constant (e.g., .1) called *learning rate*

Convergence of Perceptron Learning Algorithm

It can be proved that the algorithm will converge

- If training data is linearly separable.
- Learning rate is sufficiently small
  - The role of the learning rate is to moderate the degree to which weights are changed at each step.
  - It is usually set to some small value (e.g., 0.1) and is sometimes made to decay as the number of weight-tuning iterations increases.

Perceptron Learning Algorithm for NOT gate

- Assume $w_0 = w_1 = 0$. Let the learning rate$= \eta = 1$.
- Equations are

$$\begin{aligned}
\text{Hypothesis} \quad & z = w_0 + w_1 x_1 \\
\text{Activation} \quad & h = sign(z) \\
\text{Compute output} \quad & \hat{y} = h = sign(w_0 + w_1 x_1) \\
\text{Compute } \Delta w & = \eta(t - \hat{y})x \\
\text{Update } w_{new} & \leftarrow w_{old} + \Delta w
\end{aligned}$$

NOT
(Inverter)

A	B
0	1
1	0

Perceptron Learning Algorithm for NOT gate

w1=w0=0, η =1

X1	t	W1	W0	z	h	Isequal(t,h)	Δw	New w
-1	1	0	0	0+0=0	-1	No	Δw1=1(1+1)*(-1)=-2 Δw0=1(1+1) =2	w1⍰ 0-2=-2 w0⍰ 0+2=2
1	-1	-2	2	2-2=0	-1	Yes	Δw1=1(-1+1)1=0 Δw0=1(-1+1) =0	w1⍰ -2+0=-2 w0⍰ 2+0=2

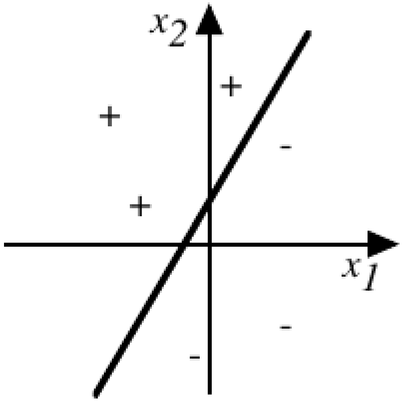
X1	t	W1	W0	z	h	Isequal(t,h)	Δw	New w
-1	1	-2	2	2+2=4	1	Yes	Δw1=1(1-1)*(-1)=0 Δw0=1(1-1) =0	w1⍰ -2+0=-2 w0⍰ 2+0=2

Exercise

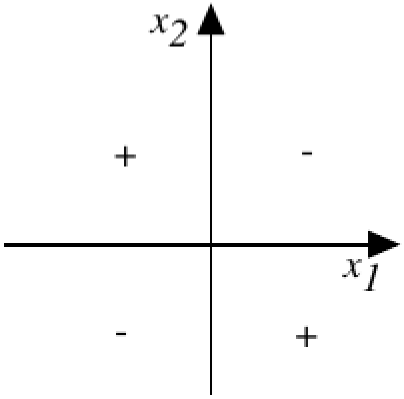
1. Represent OR gate using Perceptron. Compute the parameters of the perceptron using perceptron learning algorithm.
2. Represent AND gate using Perceptron. Compute the parameters of the perceptron using perceptron learning algorithm.

Representational Power of Perceptrons

- A perceptron represents a hyperplane decision surface in the n-dimensional space of examples.
- The perceptron outputs a 1 for examples lying on one side of the hyperplane and outputs a -1 for examples lying on the other side.



(a)

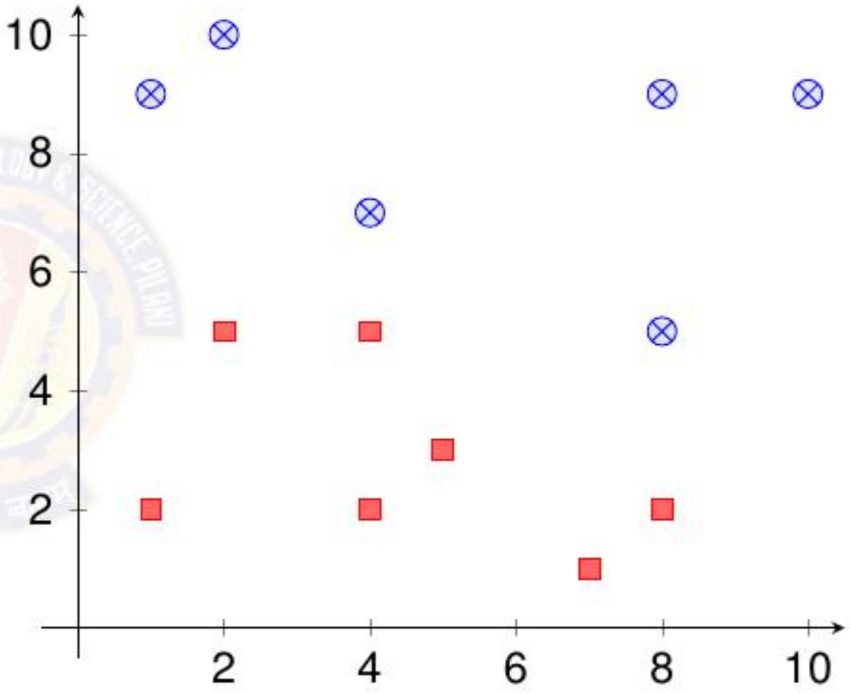


(b)

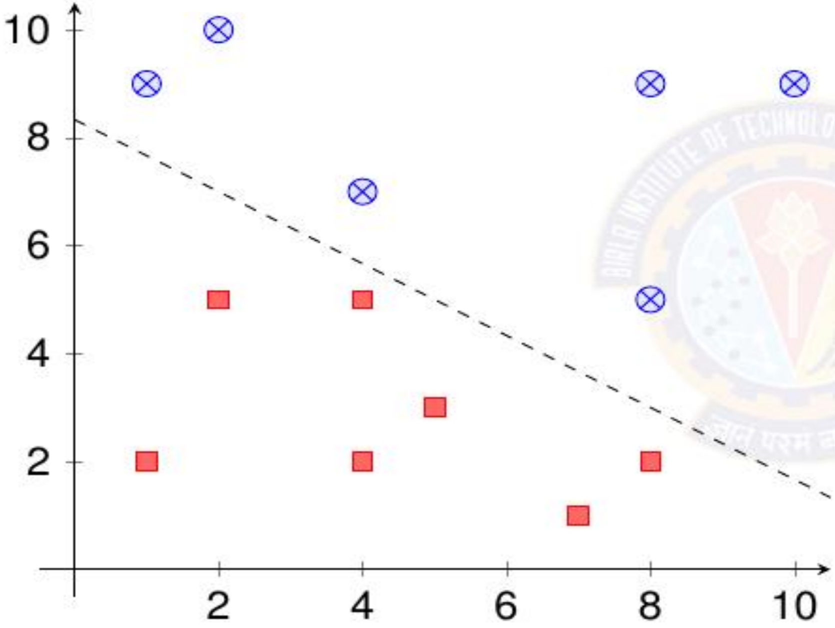
Single Layer Perceptron

Example of Linearly Separable data

x_1	x_2	y
1	9	white
10	9	white
4	7	white
4	5	pink
5	3	pink
8	9	white
4	2	pink
2	5	pink
7	1	pink
2	10	white
8	5	white
1	2	pink
8	2	pink



Example of Linearly Separable data



- Linearly separable data.
- Line equation say

$$2x_1 + 3x_2 - 25 = 0$$

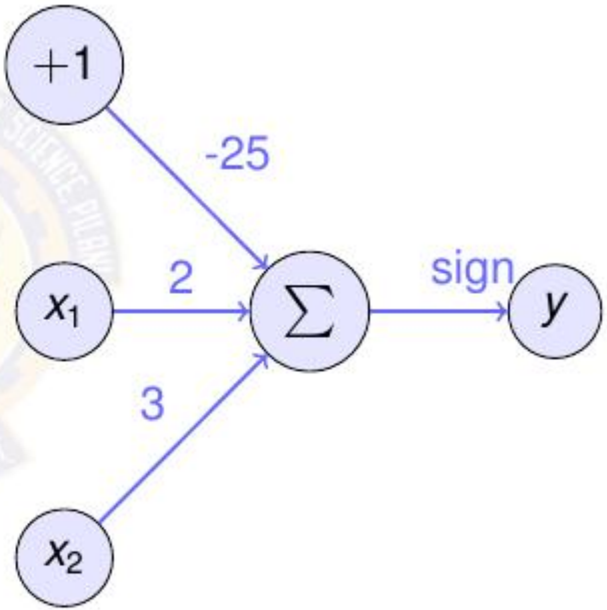
- The line becomes the decision boundary.
- New point $(0, 0) < 25$ so pink interior.
- New point $(10, 10) > 25$ so white interior.

Example of Linearly Separable data

- Linearly separable data.
- Line equation say

$$2x_1 + 3x_2 - 25 = 0$$

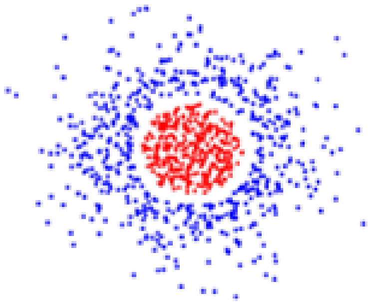
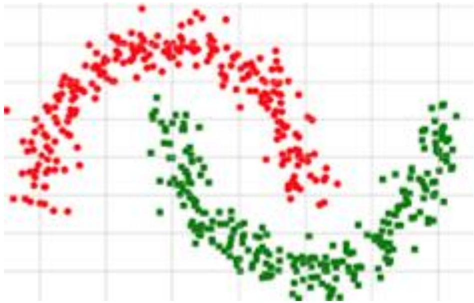
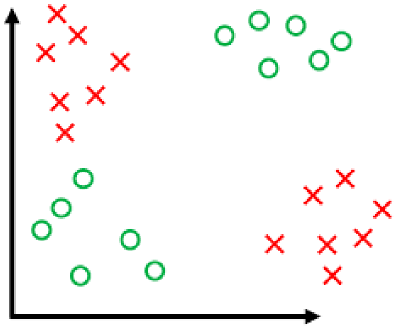
- The line becomes the decision boundary.



Non-Linearly Separable Data

Non-linearly Separable Data

- Two groups of data points are non-linearly separable in a 2-dimensional space if they cannot be easily separated with a linear line.

Perceptron for XOR Gate

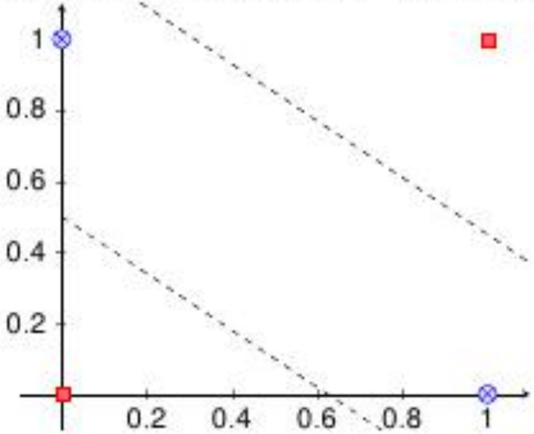
Question:

- How to represent XOR gate using a Perceptron?
- What are the parameters for the XOR Perceptron?
- Data is given below.

A	B	AB
0	0	0
0	1	1
1	0	1
1	1	0

Challenge: Data is non-linearly separable.

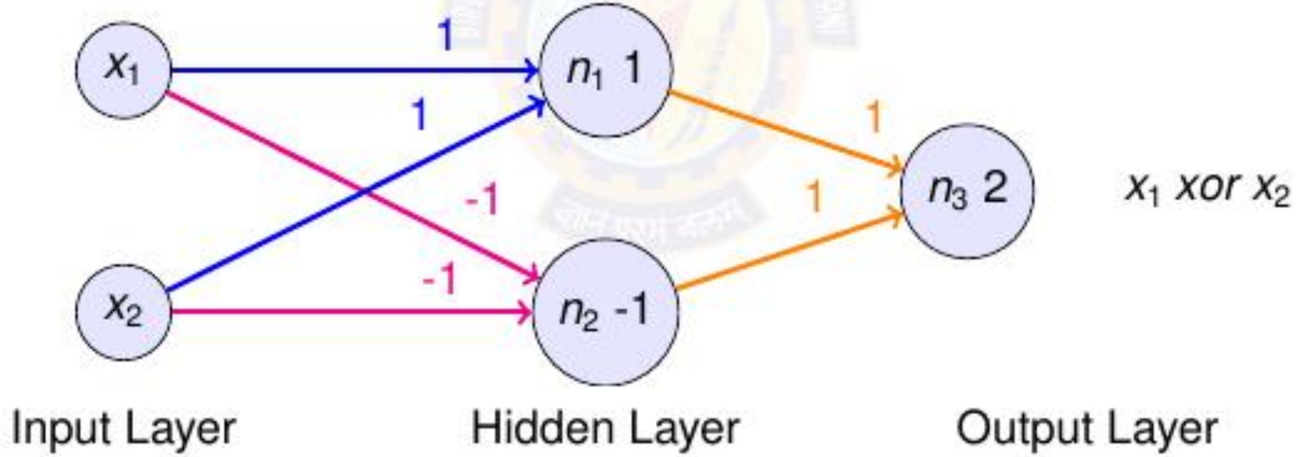
Decision boundary for XOR



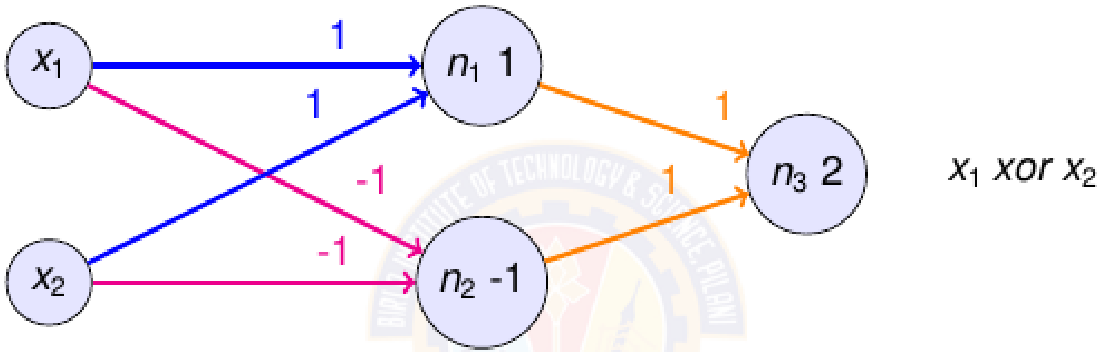
Solution for XOR

Qn: How to represent XOR gate using a Perceptron?

- Use Multilayer Perceptron (MLP)
- Introduce another layer in between the input and output.
- This in-between layer is called hidden layer.



MLP for XOR



x_1	x_2	n_1	n_2	n_3
0	0	$0 \cdot 1 + 0 \cdot 1 = 0 \not\geq th$ $n_1 = 0$	$0 \cdot (-1) + 0 \cdot (-1) = 0 > th$ $n_2 = 1$	$0 \cdot 1 + 1 \cdot 1 = 1 \not\geq th$ $n_3 = 0$
0	1	$0 \cdot 1 + 1 \cdot 1 = 1 \geq th$ $n_1 = 1$	$0 \cdot (-1) + 1 \cdot (-1) = -1 \geq th$ $n_2 = 1$	$1 \cdot 1 + 1 \cdot 1 = 2 \geq th$ $n_3 = 1$
1	0	$1 \cdot 1 + 0 \cdot 1 = 1 \geq th$ $n_1 = 1$	$1 \cdot (-1) + 0 \cdot (-1) = -1 \geq th$ $n_2 = 1$	$1 \cdot 1 + 1 \cdot 1 = 2 \geq th$ $n_3 = 1$
1	1	$1 \cdot 1 + 1 \cdot 1 = 2 \geq th$ $n_1 = 1$	$1 \cdot (-1) + 1 \cdot (-1) = -2 \not\geq th$ $n_2 = 0$	$1 \cdot 1 + 0 \cdot 1 = 1 \not\geq th$ $n_3 = 0$

Ref:
Chapter 4 of Book: Machine Learning by Tom M. Mitchell

innovate achieve lead

Thank You All!