SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with ‘A’ Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of AI &DS

Course Name – 19AD602 DEEP LEARNING

III Year / VI Semester

Unit 2-DEEP NETWORKS
Topic: HISTORY OF DEEP LEARNING

Case Study

A healthcare system uses Bayesian Neural Networks to predict diseases based on patient data. These models provide not only predictions but also uncertainty estimates, enabling doctors to review ambiguous cases and improve diagnostic safety. This approach reduced diagnostic errors by 15% and improved trust in AI systems.

History of Deep Learning

Early Beginnings (1940s - 1960s)

- **1943**: The journey began with Warren McCulloch and Walter Pitts' model of artificial neurons, the McCulloch-Pitts neuron, which laid the foundation for neural network theory.
- **1957**: Frank Rosenblatt introduced the Perceptron, an early neural network model capable of learning and recognizing patterns.

The Winter of AI (1970s - 1980s)

- Despite early enthusiasm, neural networks faced challenges, including computational limitations and the inability to train multi-layer networks, leading to reduced interest in the field, known as the "AI winter."
- **1974**: Paul Werbos developed backpropagation, a key algorithm for training neural networks, but it remained largely unnoticed until the mid-1980s.

Revival and Growth (1980s - 1990s)

- **1986**: Geoffrey Hinton, David Rumelhart, and Ronald Williams popularized backpropagation, reviving interest in neural networks.
- **1989**: Yann LeCun applied backpropagation to handwritten digit recognition, leading to the development of Convolutional Neural Networks (CNNs).

HISTORY OF DEEP LEARNING

The Emergence of Deep Learning (2000s)

- **2006**: Hinton and his colleagues introduced the concept of deep belief networks (DBNs), marking the formal beginning of deep learning.
- **2009**: Fei-Fei Li's ImageNet project provided a large-scale dataset for training deep learning models, fueling advancements in computer vision.

Breakthroughs and Dominance (2010s)

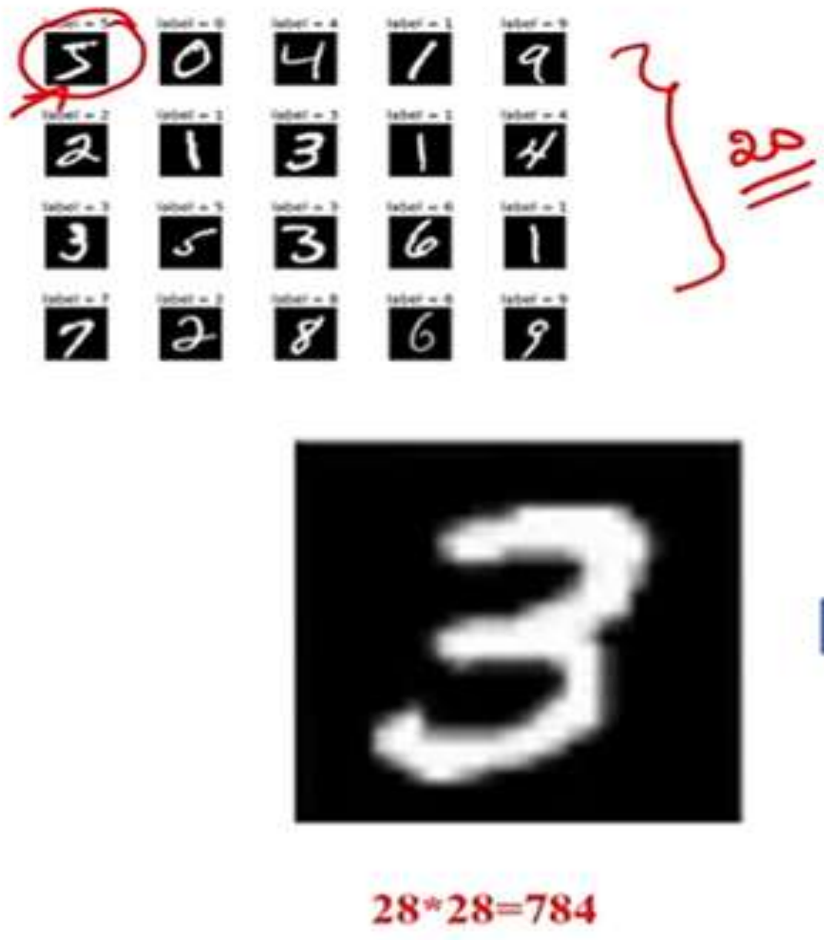
- **2012**: Alex Krizhevsky, Ilya Sutskever, and Hinton won the ImageNet competition with AlexNet, a deep CNN, demonstrating the power of deep learning in image recognition.
- **2014**: The introduction of Generative Adversarial Networks (GANs) by Ian Goodfellow opened new possibilities in generative modeling.
- **2015**: Google's DeepMind developed AlphaGo, which defeated the world champion Go player, showcasing deep learning's potential in complex strategy games.
- **2016**: The emergence of frameworks like TensorFlow and PyTorch made deep learning more accessible to researchers and practitioners.

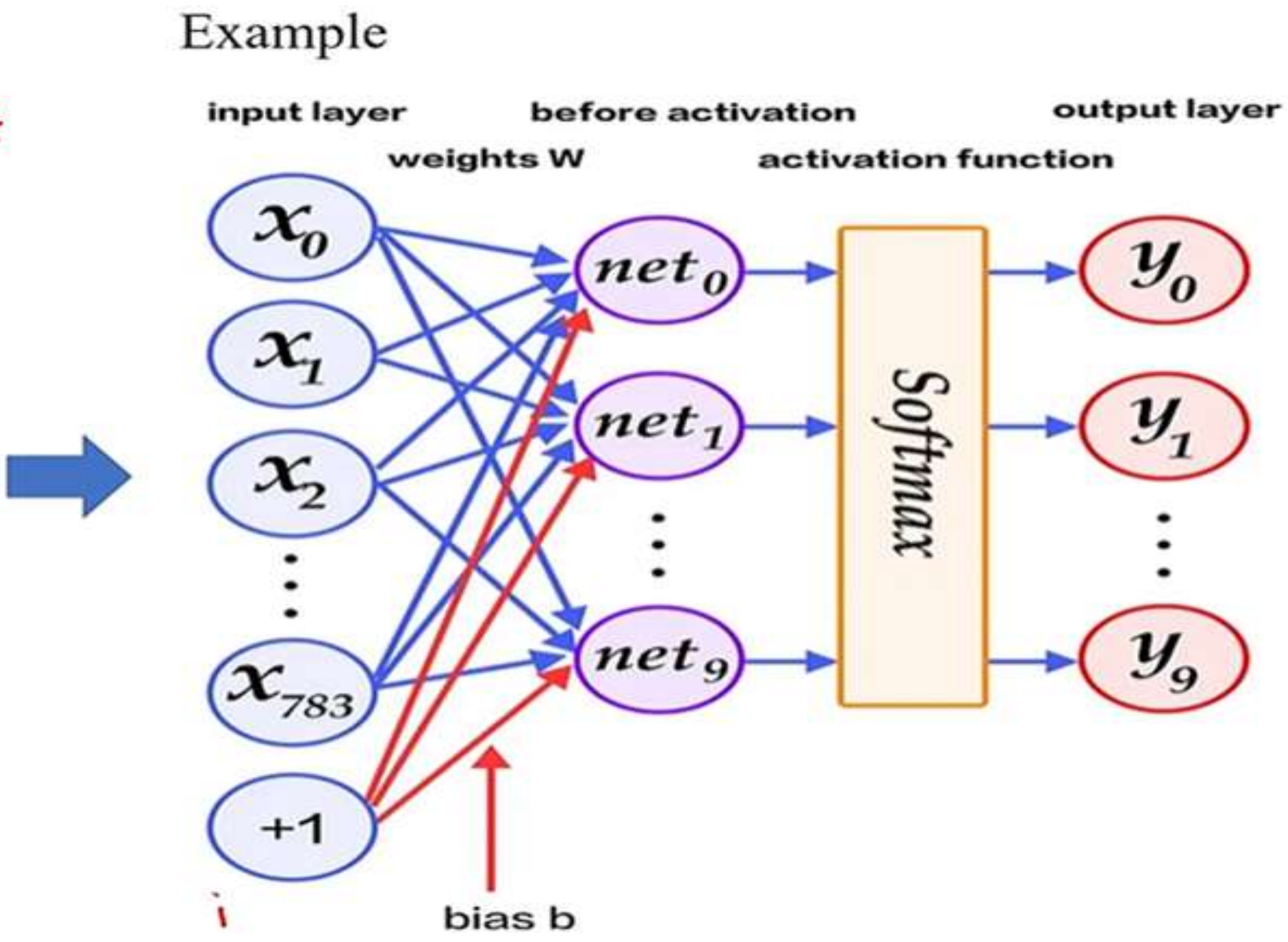
Recent Advances and Future Directions (2020s)

- **2020**: OpenAI's GPT-3, a language model with 175 billion parameters, demonstrated the capabilities of deep learning in natural language processing.
- **Ongoing Research**: Deep learning continues to evolve with advancements in areas like reinforcement learning, unsupervised learning, and multimodal learning.

Probability

- Probability is the science of quantifying uncertain things.
- Most of machine learning and deep learning systems utilize a huge dataset to learn patterns from the data.
- *Whenever data is utilized in a system rather than sole logic, uncertainty grows up.*
- *Whenever uncertainty grows up, probability becomes relevant.*


20
28*28=784


Example
input layer
before activation
output layer
weights W
activation function
x_0
x_1
x_2
x_{783}
+1
i
net_0
net_1
net_9
Softmax
y_0
y_1
y_9
bias b

Procedure

- The **input layer** is a flattened vector of the size of the input image (28*28=784).
- It is passed to a ~~hidden~~ o/p layer, where the input vector is multiplied by the weights, and added with the bias vector. This layer has 10 neurons.
- This is the implication that there are 10 digits. Then they go through a softmax activation function.
- After this step they do not output the exact digit but a vector of length 10 with **each element being a probability value for each digit.**
- Apply argmax to get the index of the probability with the highest value in the output vector i.e., predicted class.

Where is probability used?

Sample space:

- The **set of all possible values** in an **experiment**. From the example, the input can be from a set of images. Thus it is the sample space for the input.
- Similarly, the output prediction can take any value from the digits 0 to 9, thus the digits are the sample space for the output prediction.

Probability distribution:

- The probability distribution is a description of **how likely** the **random variable** is to take on **different values** of the **sample space**.
- vector $\boldsymbol{y} = [y_0, y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9,]$
- The output vector $\boldsymbol{y}$ follows softmax distribution which is also a probability distribution that shows the probability of X taking different digit values.

Where is probability used?

- *E.g., **output vector***

$$y = [0.03, 0.07, 0.04, \mathbf{0.5}, 0.06, 0.05, 0.05, 0.06, 0.04, 0.1]$$

- Total probability is 1.0.
- The argmax shows that index **3** has maximum value of 0.5 indicating the value should be 3.
- This property of adding upto 1.0 is called **normalization**.
- Also the values must be between 0 and 1. An impossible event is denoted by 0 and a possible event is denoted by 1.
- **Joint Probability**: *What is the probability of two events occurring simultaneously denoted by P(y=y, x=x) or p(y and x)?*

Example:

- Probability of seeing sun and moon **at the same time** is very low.

Probability distribution

- No of observations / total gives the probability distribution

	Male	Female	Total
Football	120	75	195
Rugby	100	25	125
Other	50	130	180
	270	230	500

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Joint Probability

- The Joint probability is a statistical measure that is used to calculate the probability of two events occurring together at the same time i.e., P(A and B).

$$P(A \cap B) = P(A \mid B) * P(B)$$

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Marginal Probability

- Marginal distribution of a **subset** of a collection of random variables is the probability distribution of the variables contained in the subset.

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Conditional Probability

- It defines the probability of one event occurring given that another event has occurred (by assumption, presumption, assertion or evidence).

$$P(A|B) = P(A, B) / P(B)$$

- P(Rugby | Female) = P(Rugby ,Female) / P(Female)

  = 0.05 / ------ = -----

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Activity

Train a Bayesian Neural Network on noisy image data to classify categories and observe how it provides uncertainty estimates for ambiguous inputs compared to a traditional neural network.

THANK YOU