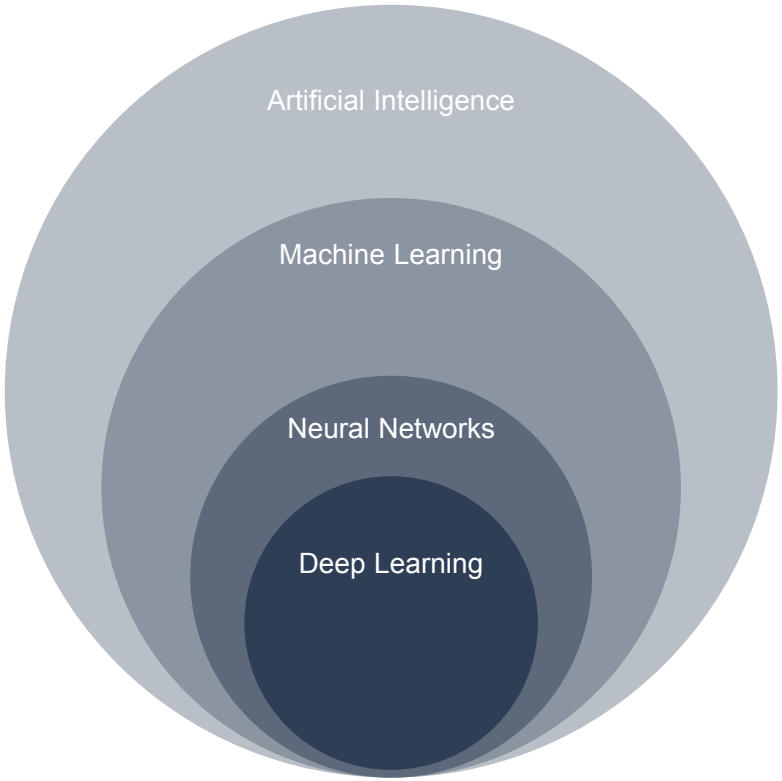
University of Illinois at Urbana-Champaign
Dept. of Electrical and Computer Engineering

ECE 101: Exploring Digital Information Technologies for Non-Engineers

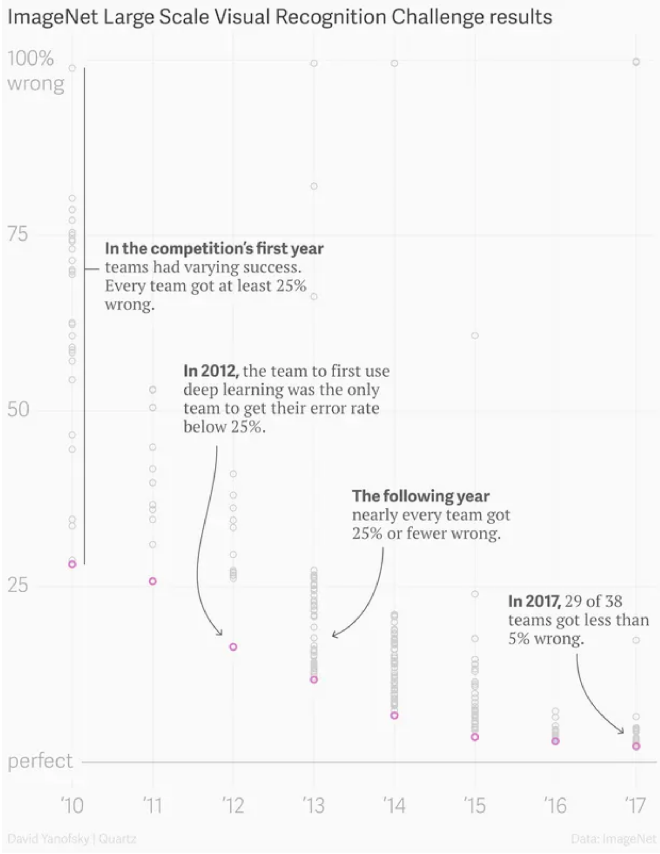
Machine Learning - Neural Networks

Neural Nets for AI



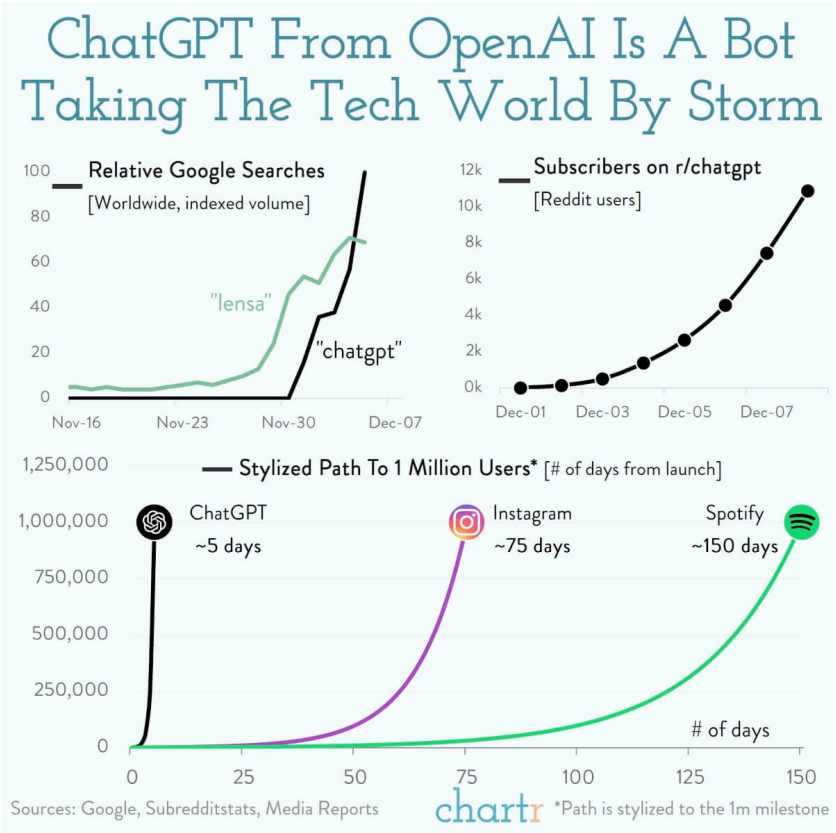
Neural Nets through the Years

- **1942**—First computational model for neural networks
- **1965**—First functional networks of many layers
- **1975**—Backpropagation algorithms for training multilayer networks
- **1990s**—Datasets quite small; computers not that fast; other methods doing better
- **2005-2007**—Unsupervised learning with deep nets; use of GPUs
- **2009**—ImageNet: Image database of 14 million images for more than 21000 concepts
- **2012**—AlexNet: Winner of ImageNet Large Scale Visual Recognition Challenge 2012



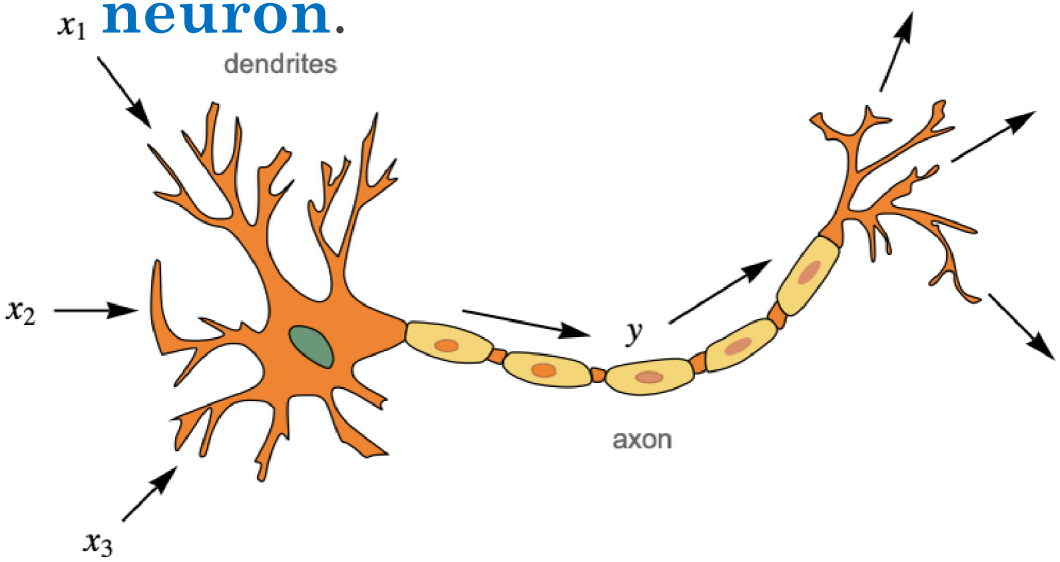
Neural Nets through the Years

- **2015**—Wolfram Image Identification Project announced
- **2017**—A team of Google researchers proposes a new simple network architecture, the Transformer. Transformers enabled advancements in generative models compared to older long short-term memory models
- **2018**—OpenAI releases GPT (Generative Pre-trained Transformer), a language model that achieves state-of-the-art performance on various natural language processing tasks
- **2018**—Wolfram Neural Net Repository launched
- **2021**—DALL.E, a transformer-based neural network-based system developed by OpenAI, generates images from textual descriptions
- **2022**—ChatGPT releases GPT-3.5, an AI tool that reached one million users within five days. The tool can access data from the web from up to 2021.



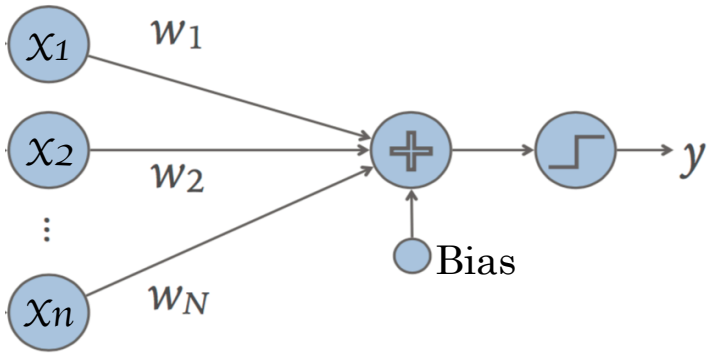
Perceptron

A **perceptron,** the simplest form of an artificial neural network **mimics a human neuron**.

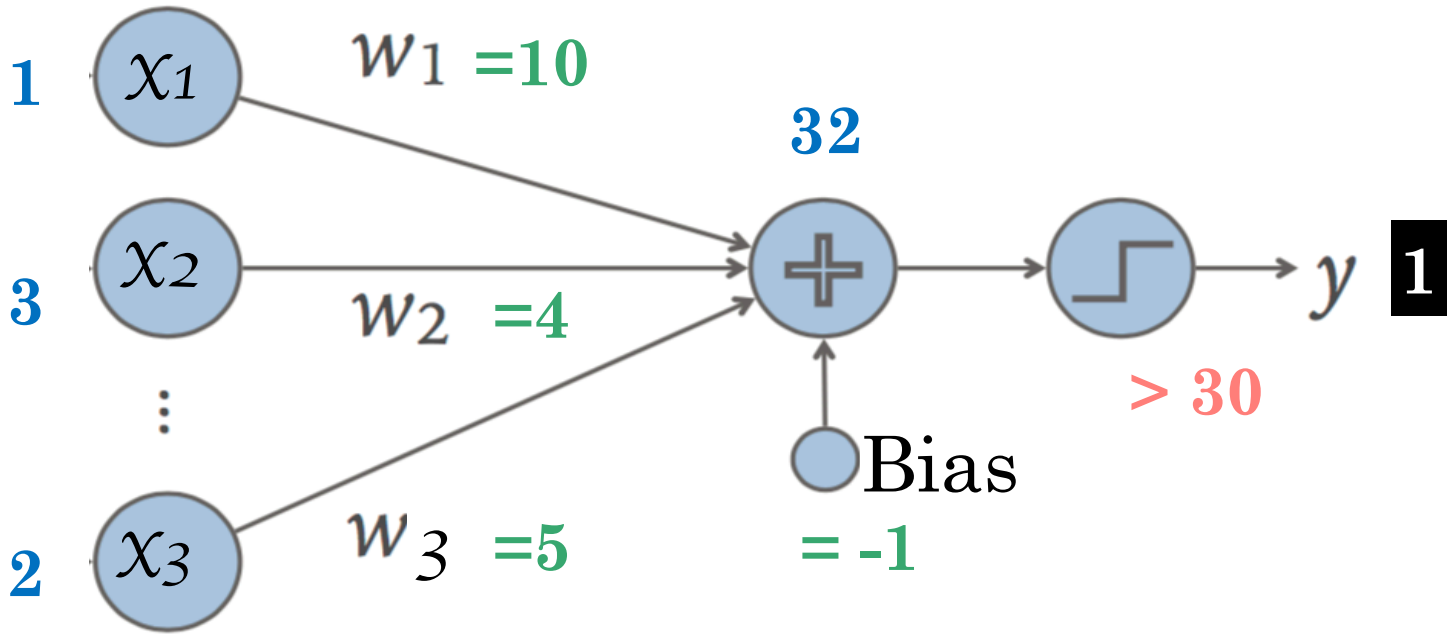


Inputs (on the left) are multiplied by weights, then summed together with a bias.

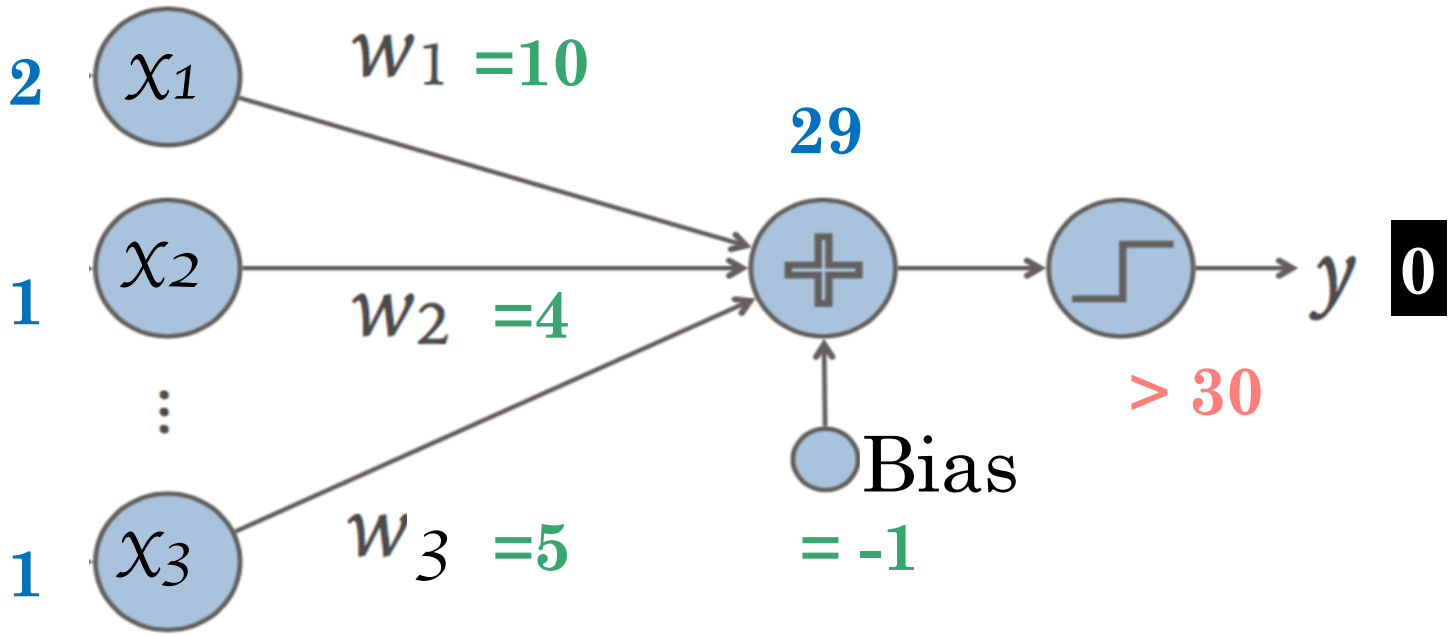
The sum is then converted to +1 or -1.



Computation in a Perceptron

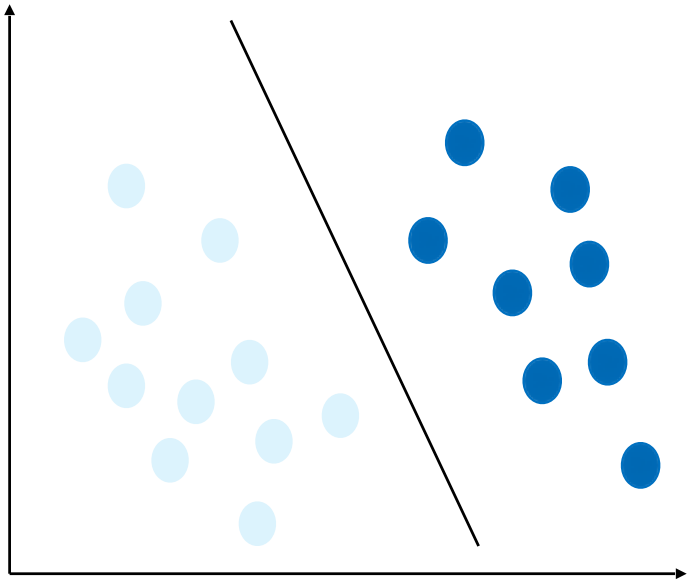


Computation in a Perceptron



The Perceptron can Classify: +1 or -1, Yes or No, A or B

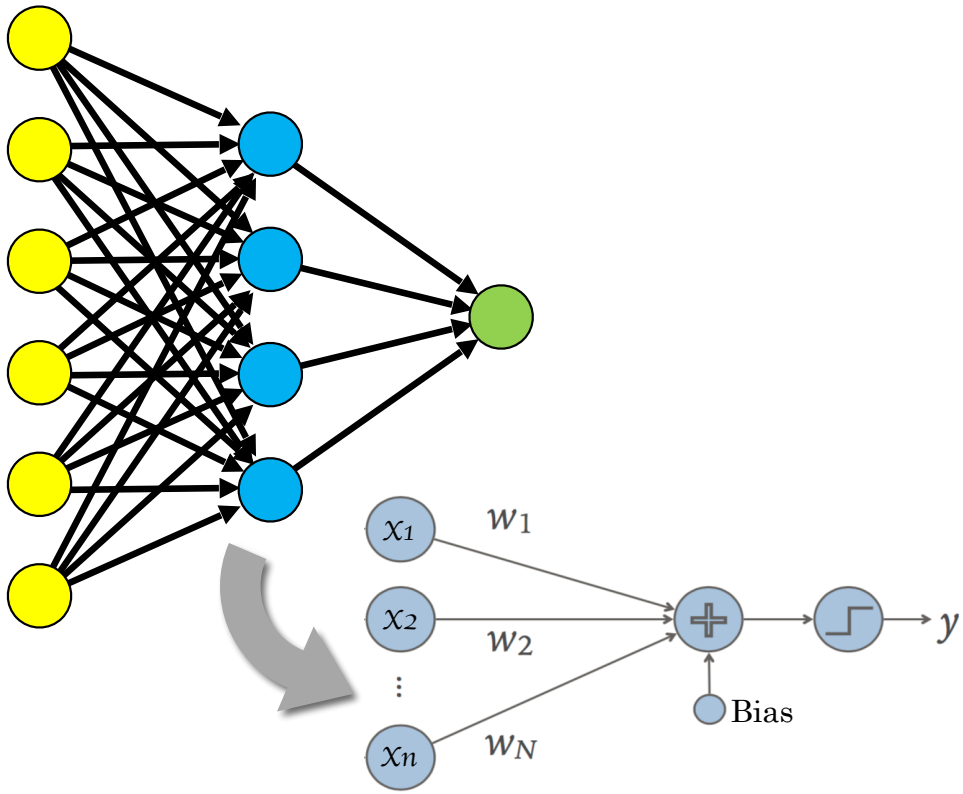
It's a linear classifier!



Neural Networks Consist of Many Artificial Neurons

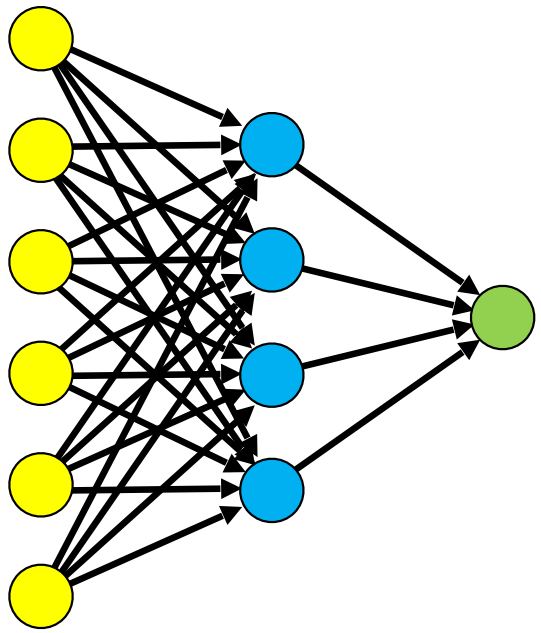
To **perform more complex** (non-linear) **tasks**,

- perceptrons can be **connected in a network**
- by using the output of one perceptron
- as the input to a second,
- then a third, and so forth.



Input, Output and Hidden Layers

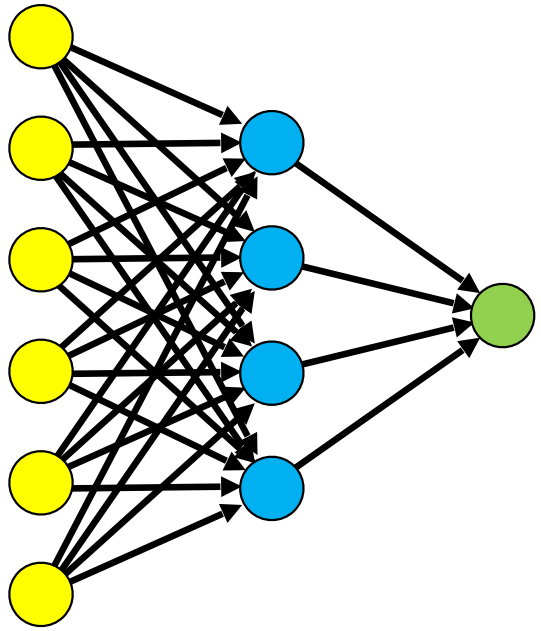
- **Input layer** (in yellow)
- Every input is connected to every node in the **hidden layer** (in blue)
- **Output layer** (in green) consists of a single output perceptron
- In a fully connected network, every node is connected every node in the next layer



The Network is a Complex Function made up of Simple Parts

Think of it like a brain
made of tiny decision-makers.

Each one looks at a piece of information
and passes it along.



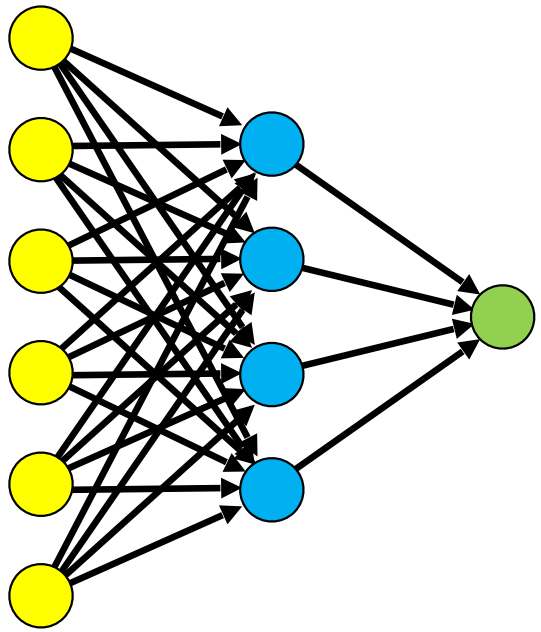
Advancement: Deeper Networks that Leverage Input Relationships

In theory, a small number of fully connected hidden layers can learn anything.

In practice, **three things happened** before neural networks enjoyed major success.

By the late 1990s, researchers had built

- **One: new architectures** that **leveraged relationships between the inputs**, and
- **Two**: **deeper networks** to **capture more complex functions more quickly**.

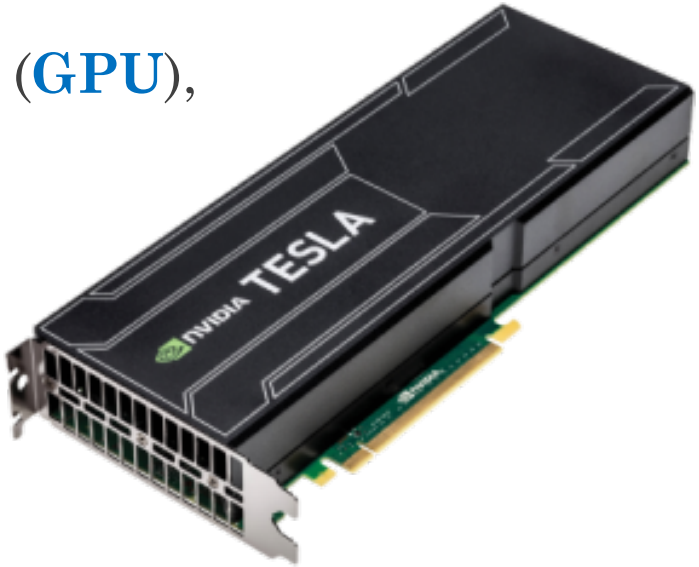


Advancement: Ubiquitous Availability of Fast Computation

Three: The last change came in early 2007
- with the release of the first easily programmable **graphics processing unit** (**GPU**),
- NVIDIA's GTX80 (not the one shown).

GPUs had developed
- to meet the entertainment market demand for high-resolution graphics,
- and by 2007 offered much **more raw computational power than processors**.

They are **well-suited to training and applying neural networks**.



UIUC Offered First Class on GPGPU Programming

The **first class on** general-purpose GPU (**GPGPU**) **programming**

- was offered at UIUC in Spring 2007
- by David Kirk, chief architect from NVIDIA,
- and Wen-mei Hwu, ECE professor.

Deep Learning Derives Features from Data

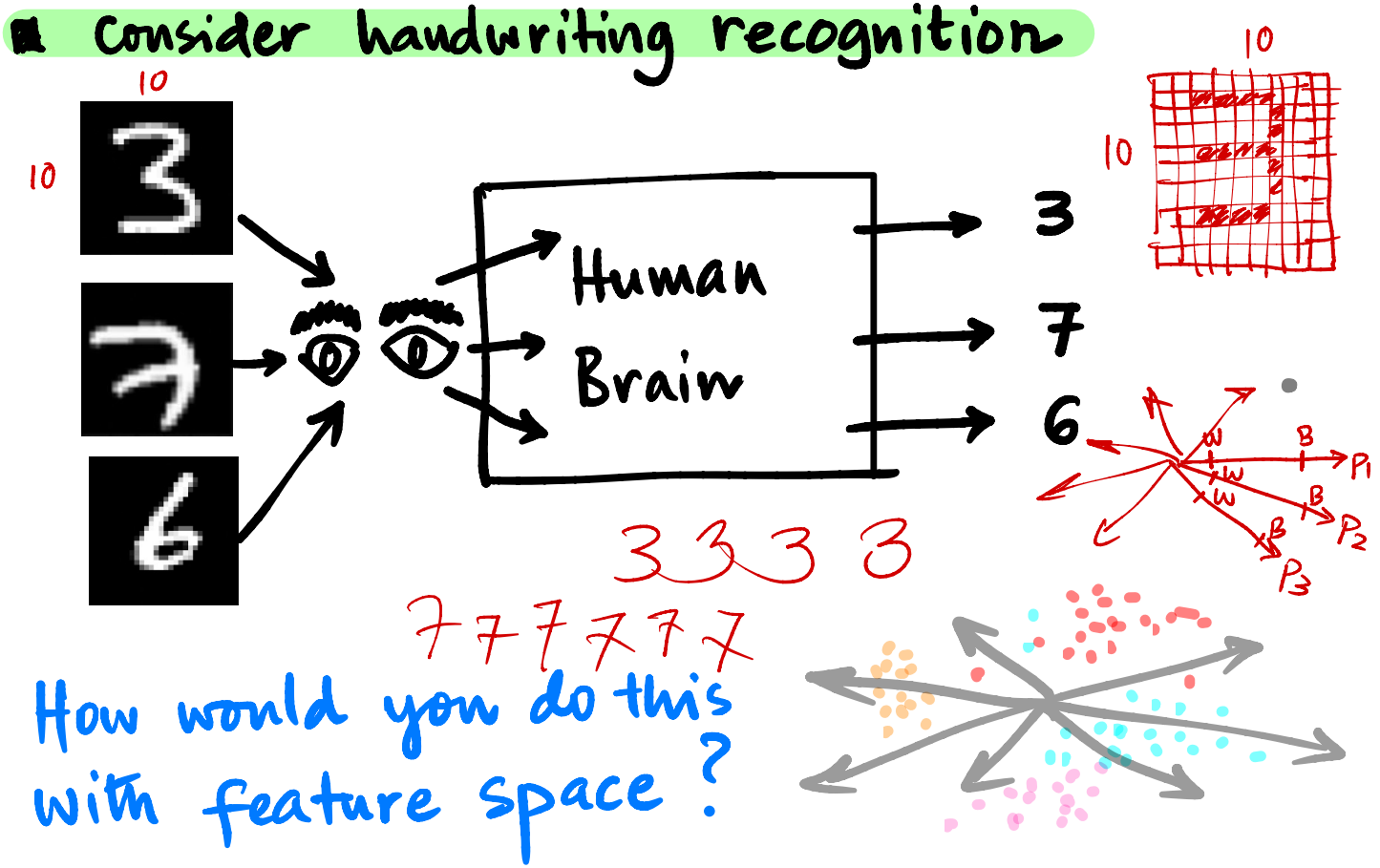
More recent designs
- have stopped using human-modeled features, and
- instead **allow training** of the neural network
- **to derive the features of importance** from the data.

This approach is called **deep learning**.

It's the number of layers, or depth, of neural networks that distinguishes a single neural network from a deep learning algorithm,

Deep learning is **possible due to** the sheer **volume of data** now **available** in many problems.

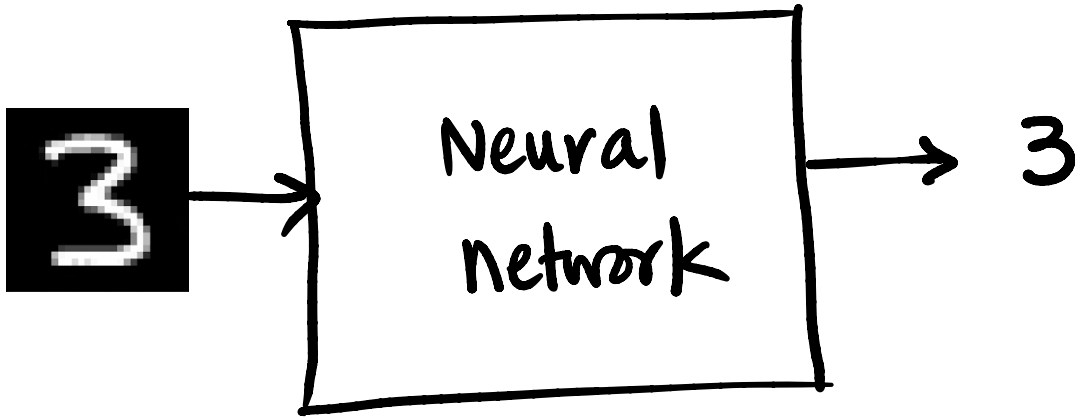
Deep Learning Applications



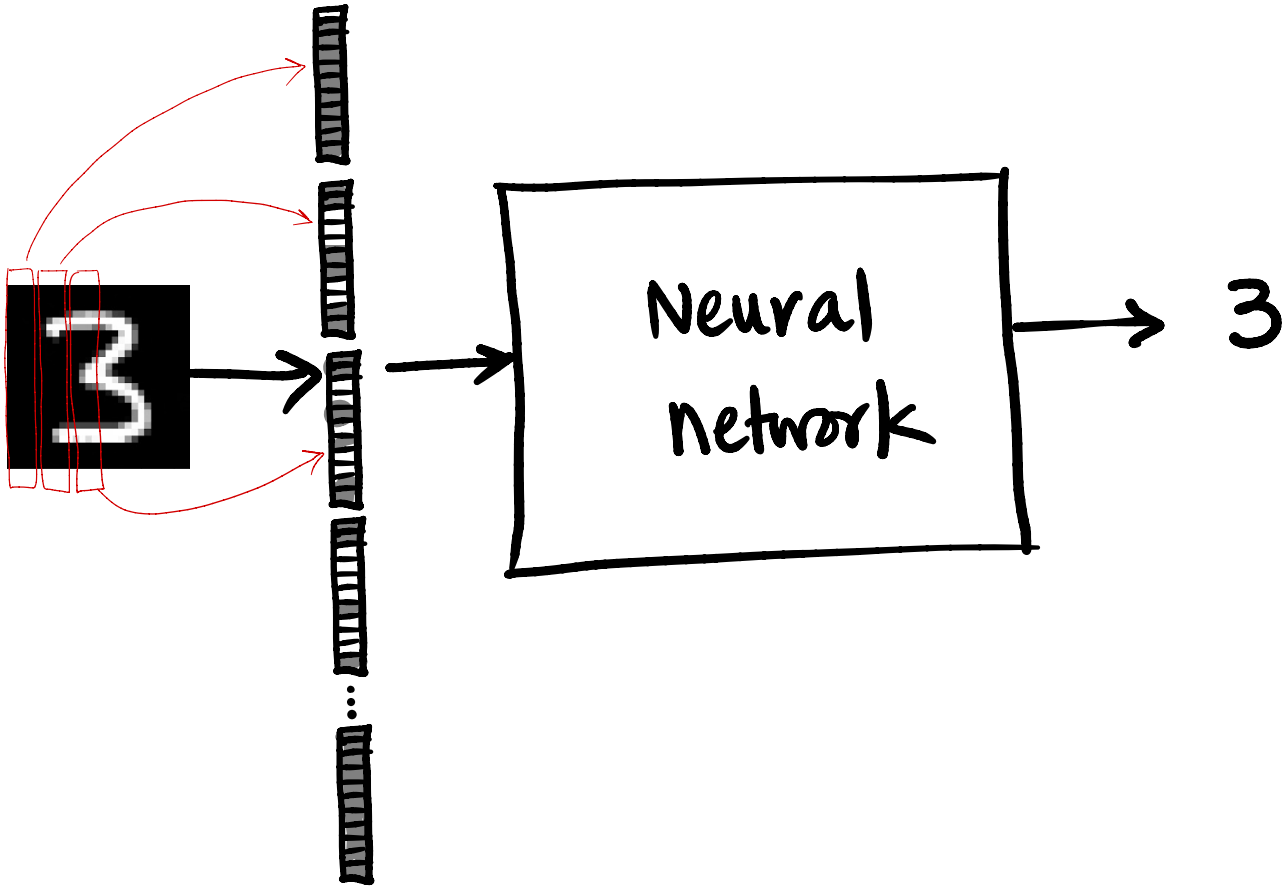
What Features should you Use?

Feature engineering or feature extraction is often called an “art”.

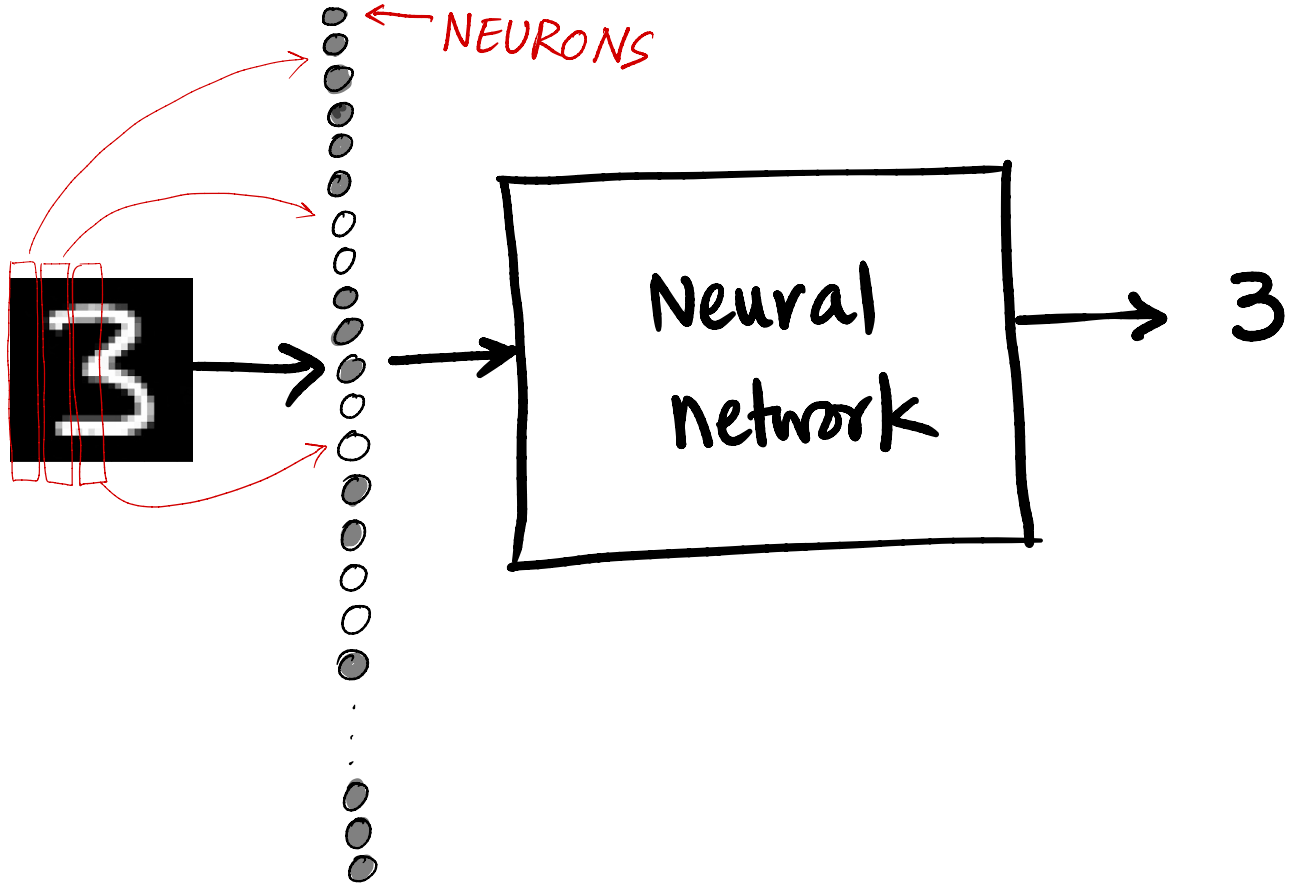
Enter Neural Networks.



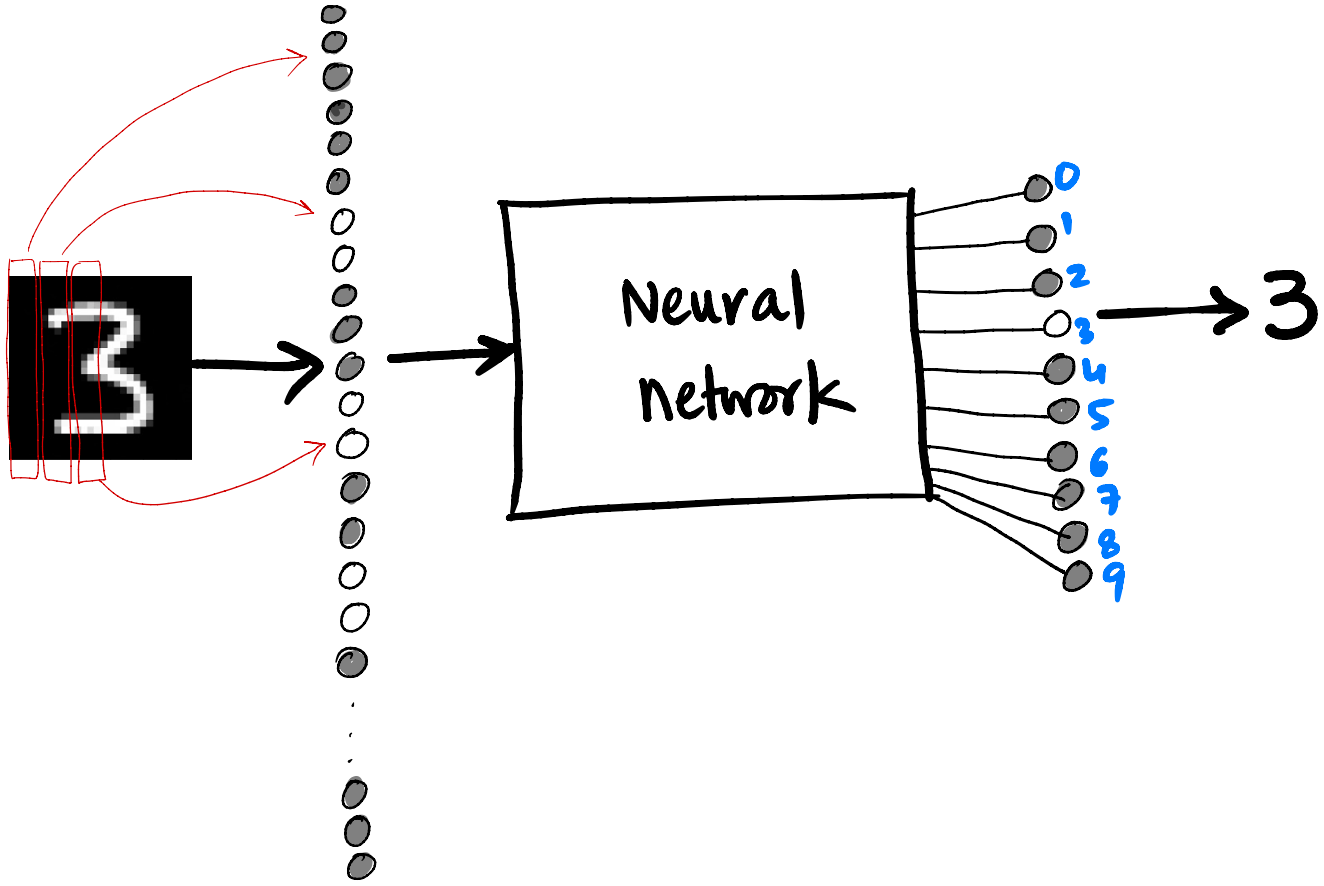
What is the Network Doing?



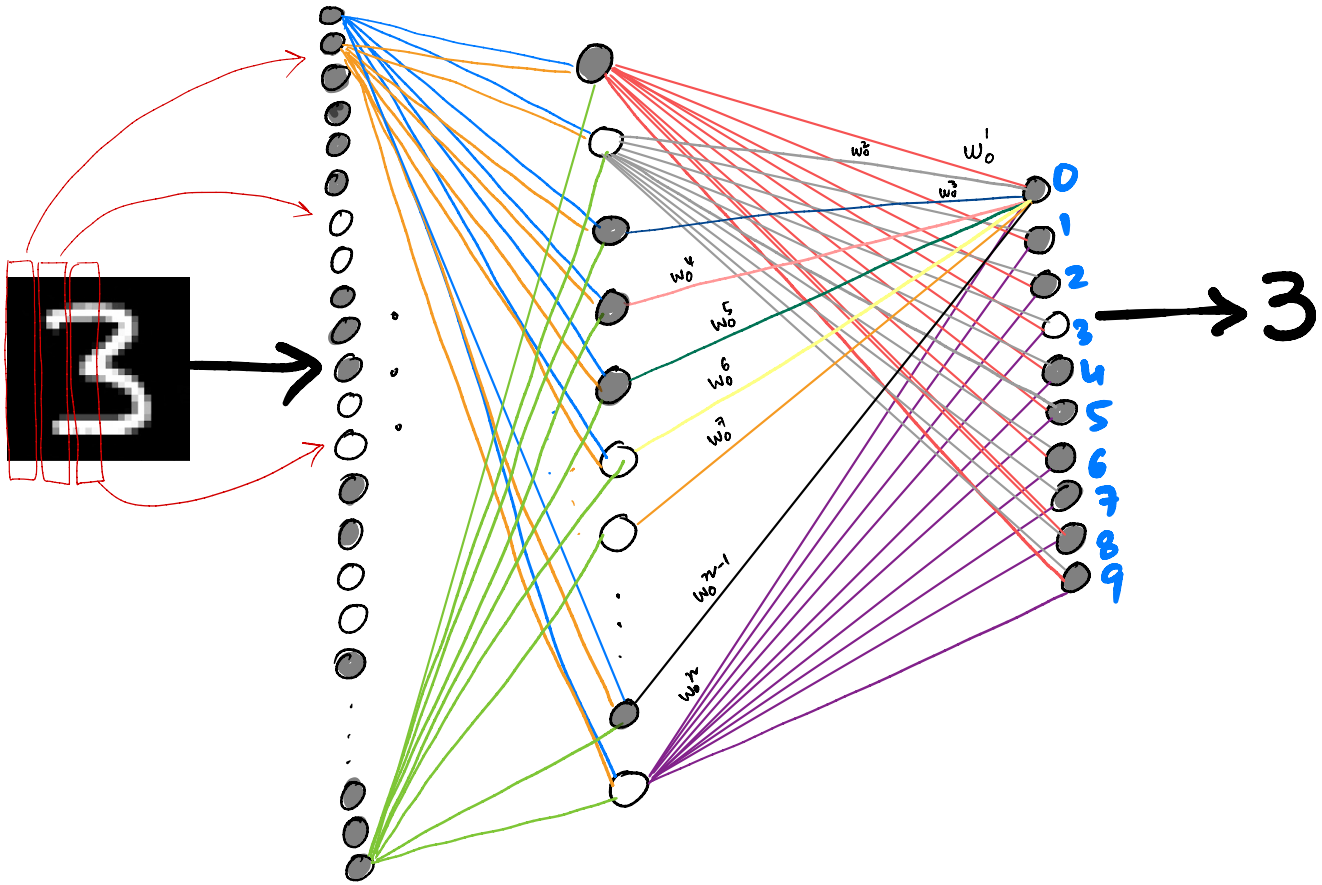
What is the Network Doing?



What is the Network Doing?



What is the Network Doing?



Terminology You Should Know from These Slides

- Deep Neural Networks (DNNs)
- perceptron
- fully connected layer
- Input layer, hidden layer, output layer
- Graphics Processing Unit (GPU)
- Deep learning