Big Data Analytics, Deep learning, Explainable AI (XAI)

Big data

- Big data refers to data sets that are too large or complex to be dealt with by traditional data-processing application software.
- Data with many fields offer greater statistical power, while data with higher complexity may lead to a higher false discovery rate.

Big data

Type of Data

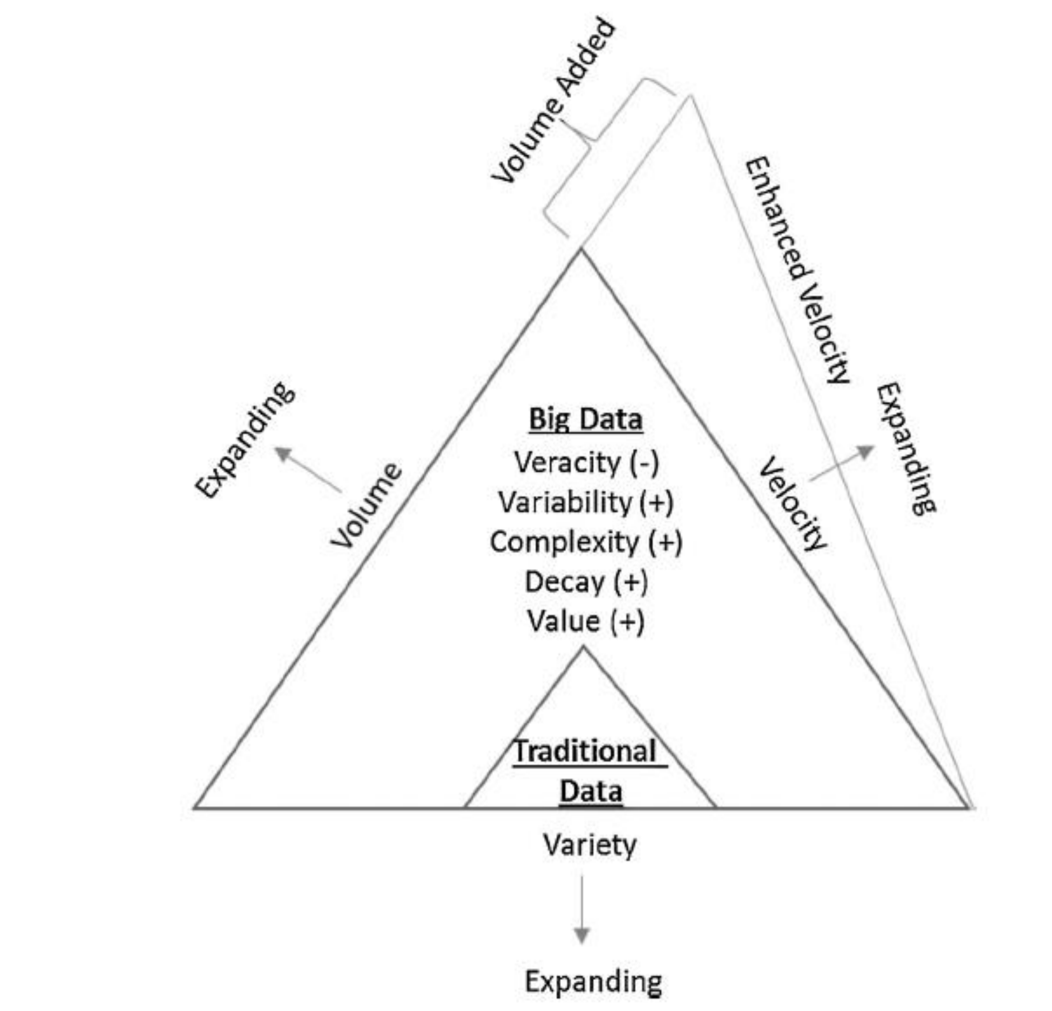
- **Structured**
  - Tables
- **Semi-Structured**
  - XML, JSON
  - Clickstream data, Ecommerce, APIs
- **Unstructured**
  - Audio, Video, Images, Text, Sensory Data
  - IoT Devices, Social Media

Data

- Storage
  - SQL, NoSQL
- Processing
  - Hadoop

Dimensions of big data

- Volume
- Velocity
- Variety
- Veracity
- Variability
- Complexity
- Value
- Decay



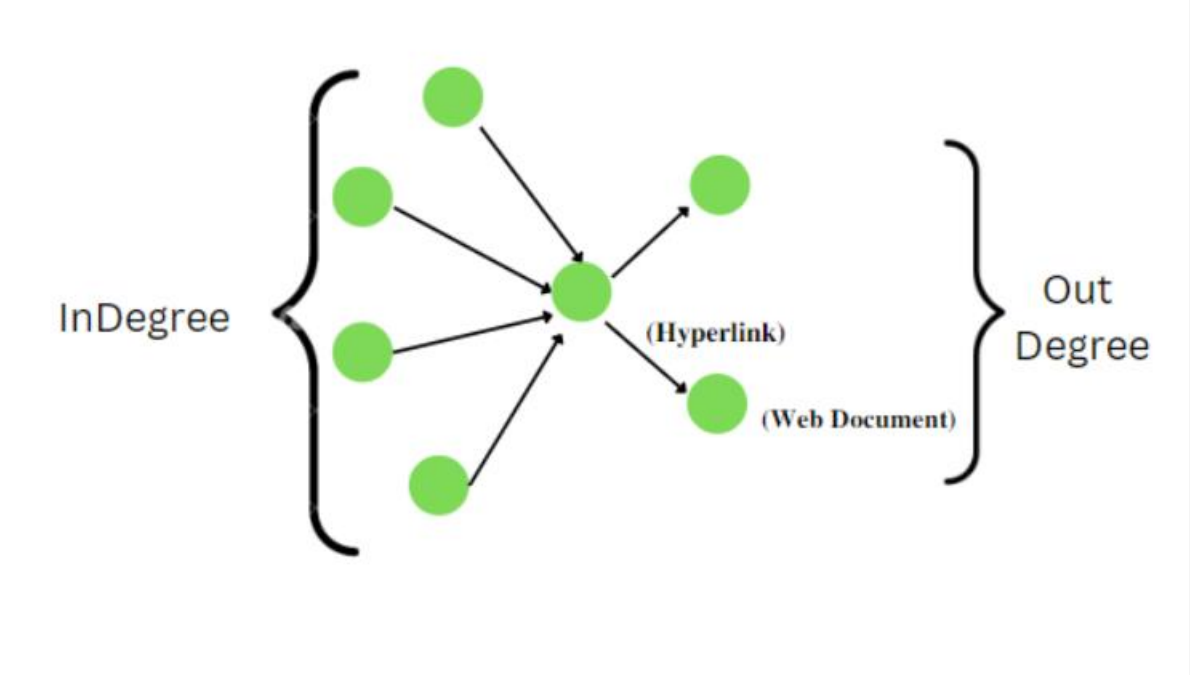
Evolution of big data and data analytics

- 1950s to mid-1990s
- Big Data 1.0
- Big Data 2.0
- Big Data 3.0

Big Data 1.0

- Web usage mining
- Web structure mining
- Web content mining

Web content mining

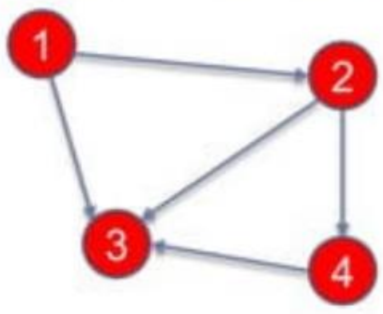


Big Data 2.0

- Social Media
  - Sentiment Analysis
    - Entity, sentence, document level
    - Lexicon-based, Machine Learning
  - Social Network Analysis
    - network structure, connections, nodes,
    - network density, network centrality, network flows

Networks – How to represent Networks

Graph (directed)



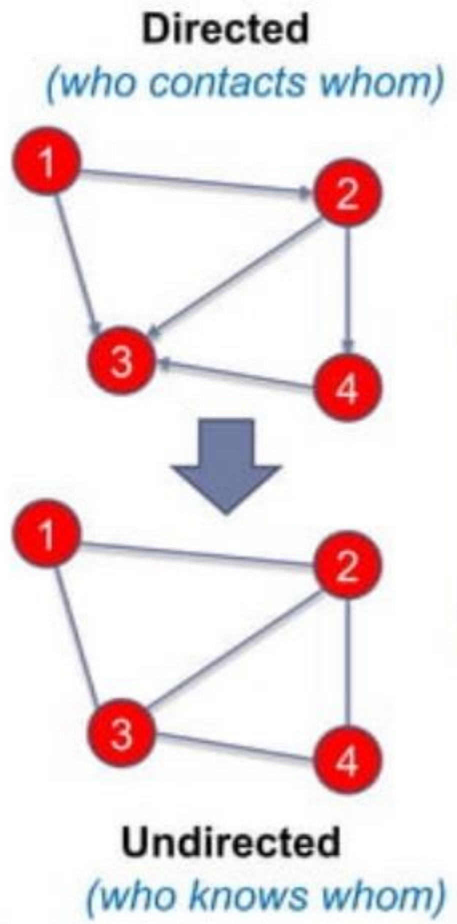
Edge list

Vertex	Vertex
1	2
1	3
2	3
2	4
3	4

Adjacency matrix

Vertex	1	2	3	4
1	-	1	1	0
2	0	-	1	1
3	0	0	-	0
4	0	0	1	-

Networks – How to represent Networks



Edge list remains the same

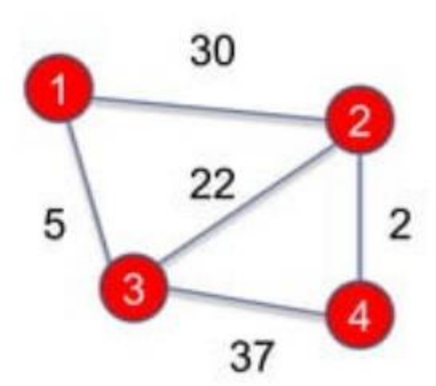
Vertex	Vertex
1	2
1	3
2	3
2	4
3	4

But interpretation is different now

Adjacency matrix becomes symmetric

Vertex	1	2	3	4
1	-	1	1	0
2	1	-	1	1
3	1	1	-	1
4	0	1	1	-

Ties – Adding weights



Edge list: add column of weights

Vertex	Vertex	Weight
1	2	30
1	3	5
2	3	22
2	4	2
3	4	37

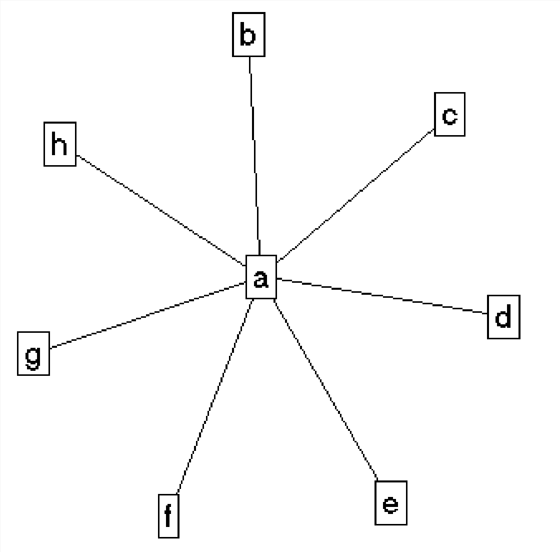
Adjacency matrix: add weights instead of 1

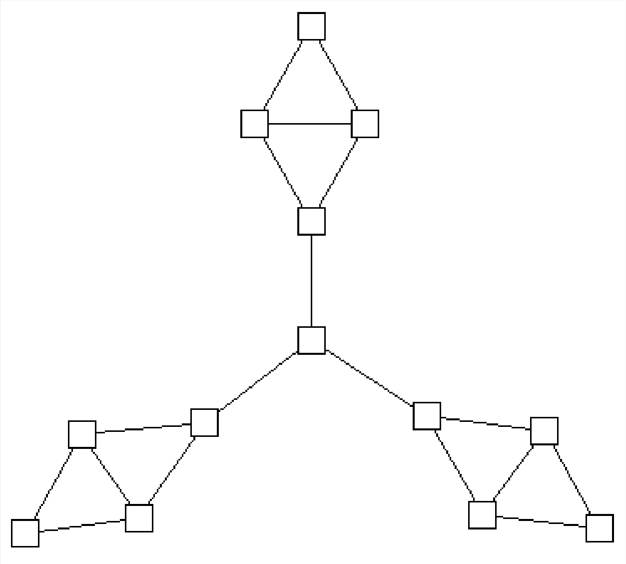
Vertex	1	2	3	4
1	-	30	5	0
2	30	-	22	2
3	5	22	-	37
4	0	2	37	-

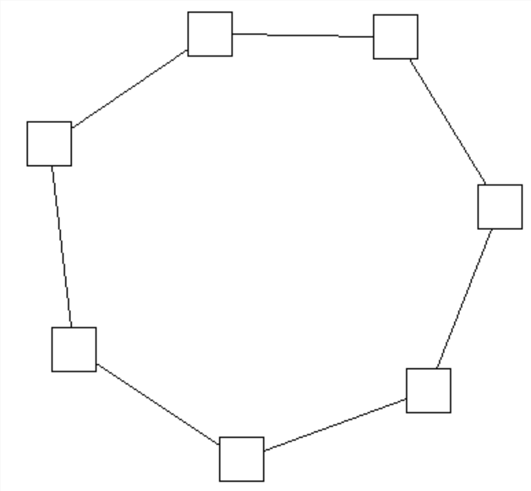
Centrality in Social Networks

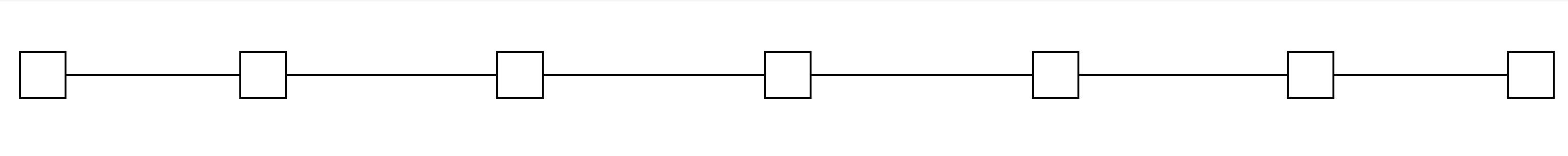
Intuitively, we want a method that allows us to distinguish “**important**” actors.

Consider the following graphs:









Centrality

In social networks, some people, like **celebrities** and **politicians** have a lot of followers and can propagate information easier than ordinary subjects.

Hence, these nodes can be considered as central.

However, *this definition of centrality is not unique*, since we can define it in terms of the load that each node receives.

Centrality Measures

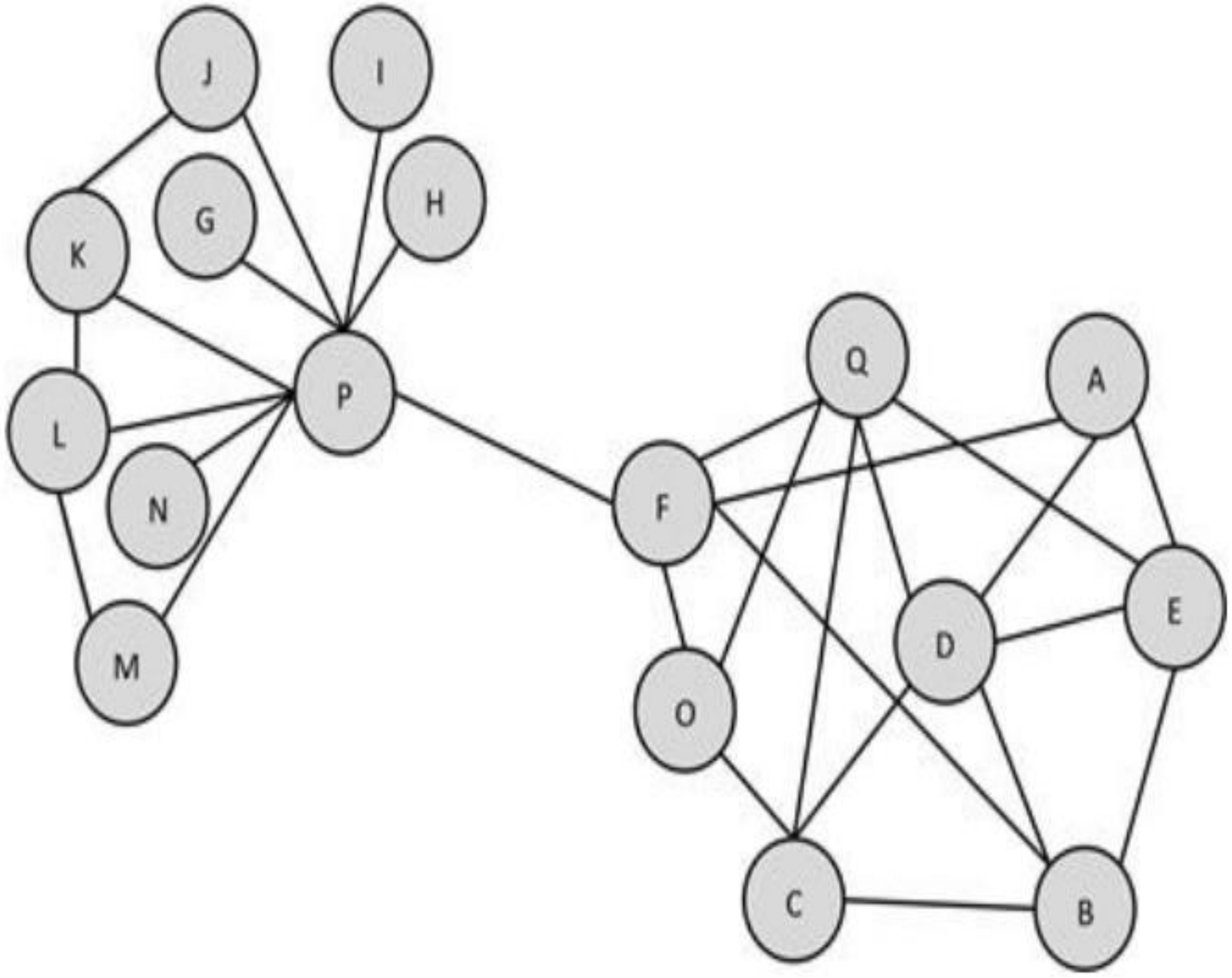
- ❑ Degree
- ❑ Closeness
- ❑ Harmonic
- ❑ Between

Degree Centrality

In Figure, node P has the highest degree centrality of 9.

Meanwhile, node F has a relatively low degree centrality of 5.

Many other nodes have that same centrality value or higher (e.g., node D has a degree centrality of 5).

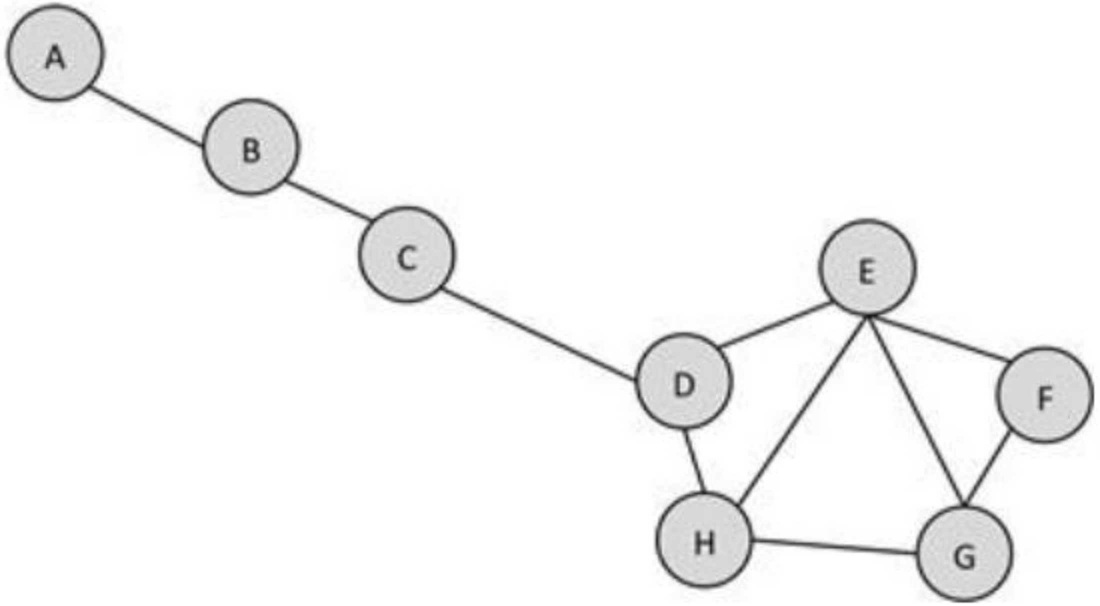


Closeness centrality

Let's start by computing the average shortest path length of node D.
Table shows each node and the length of the shortest path from D.

The Shortest Path Lengths from D to each Other Node in the Network

Node	Shortest Path from D
A	3 (D–C–B–A)
B	2
C	1
E	1
F	2
G	2
H	1



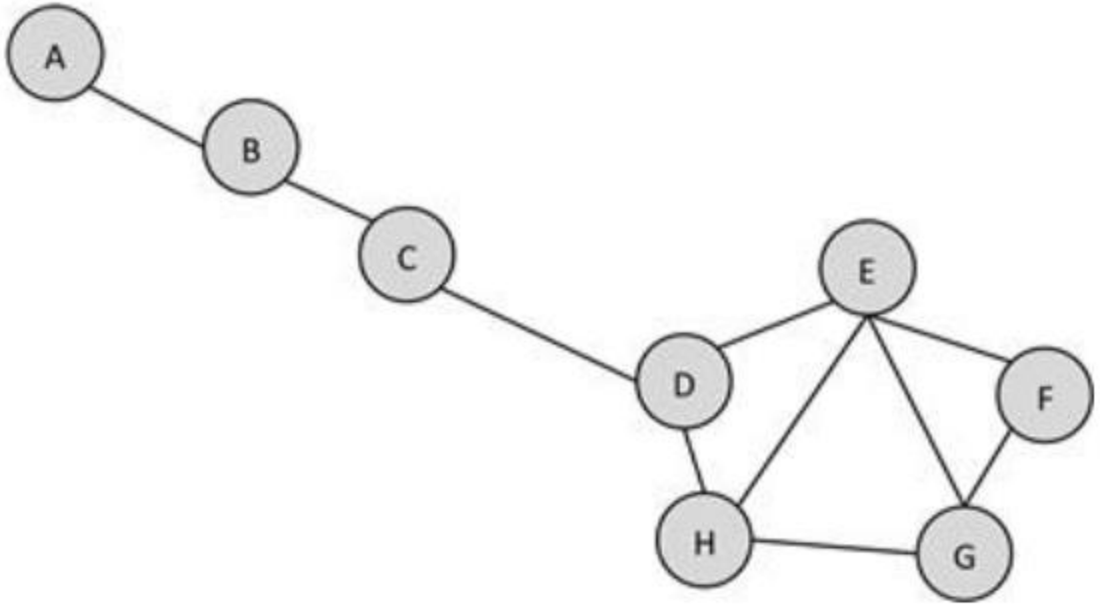
Closeness centrality

Closeness is based on the inverse of the distance of the actor (in question) to every other actor in the network.

The Shortest Path Lengths from D to each Other Node in the Network

Node	Shortest Path from D
A	3 (D–C–B–A)
B	2
C	1
E	1
F	2
G	2
H	1

Closeness (D)= 1/(3+2+1+1+2+2+1)= 1/12= 0.083



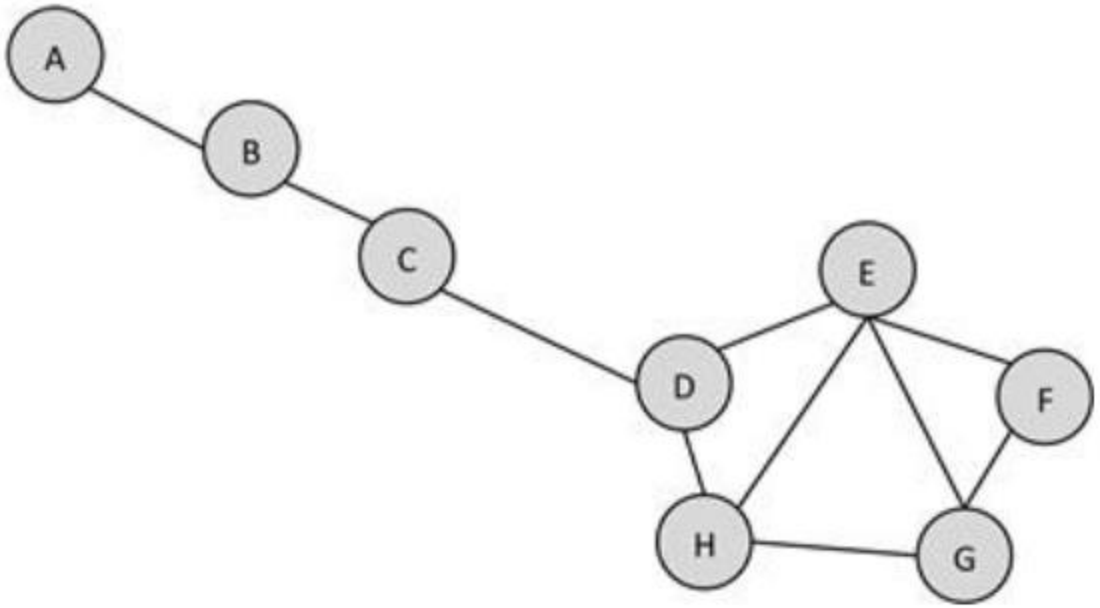
Closeness centrality

Closeness centrality is based on the inverse of the distance of the actor (in question) to every other actor in the network.

The Shortest Path Length from node A to Every Other Node in the Network

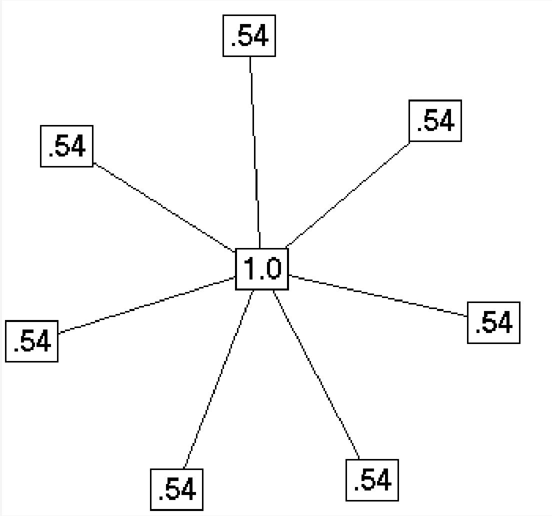
Node	Shortest Path from A
B	1
C	2
D	3
E	4
F	5
G	5
H	4

Closeness (A)= 1/(1+2+3+4+5+5+4)= 1/24= 0.042

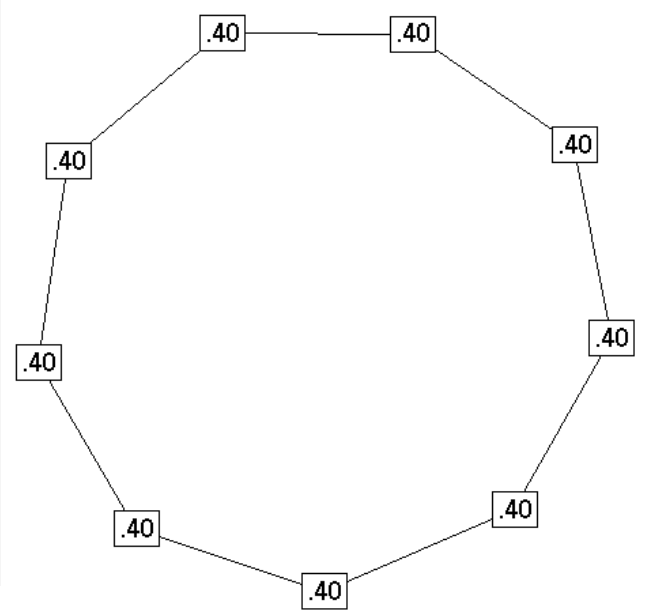


Closeness centrality

Closeness Centrality in the examples

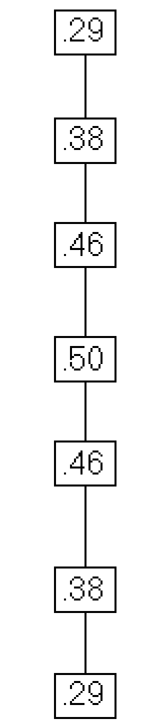


Distance	Closeness	normalized
0 1 1 1 1 1 1 1	.143	1.00
1 0 2 2 2 2 2 2	.077	.538
1 2 0 2 2 2 2 2	.077	.538
1 2 2 0 2 2 2 2	.077	.538
1 2 2 2 0 2 2 2	.077	.538
1 2 2 2 2 0 2 2	.077	.538
1 2 2 2 2 2 0 2	.077	.538
1 2 2 2 2 2 2 0	.077	.538



Distance	Closeness	normalized
2 3 4 4 3 2 1	.050	.400
1 2 3 4 4 3 2	.050	.400
0 1 2 3 4 4 3	.050	.400
1 0 1 2 3 4 4	.050	.400
2 1 0 1 2 3 4	.050	.400
3 2 1 0 1 2 3	.050	.400
4 3 2 1 0 1 2	.050	.400
4 4 3 2 1 0 1	.050	.400
3 4 4 3 2 1 0	.050	.400

Closeness centrality



Closeness Centrality in the examples

Distance	Closeness	normalized
0 1 2 3 4 5 6	.048	.286
1 0 1 2 3 4 5	.063	.375
2 1 0 1 2 3 4	.077	.462
3 2 1 0 1 2 3	.083	.500
4 3 2 1 0 1 2	.077	.462
5 4 3 2 1 0 1	.063	.375
6 5 4 3 2 1 0	.048	.286

Betweenness Centrality

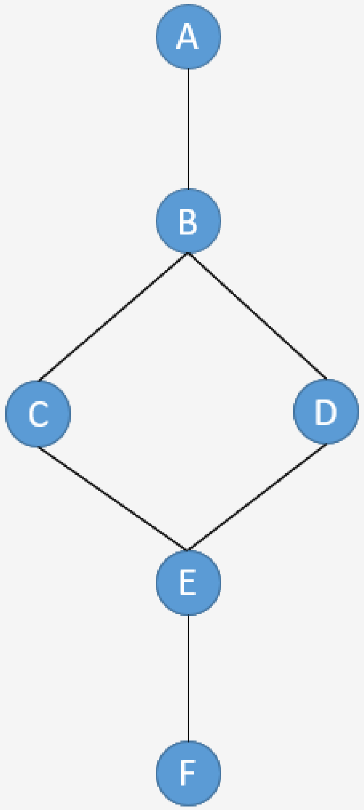
Let's compute betweenness centrality for node B.

There are 10 pairs of nodes to consider: AC, AD, AE, AF, CD, CE, CF, DE, DF, and EF.

Without counting, we know that 100% of the shortest paths from A to every other node in the network go through B, since A can't reach the rest of the network without B. Thus, the fractions for AC, AD, AE, and AF are all 1.

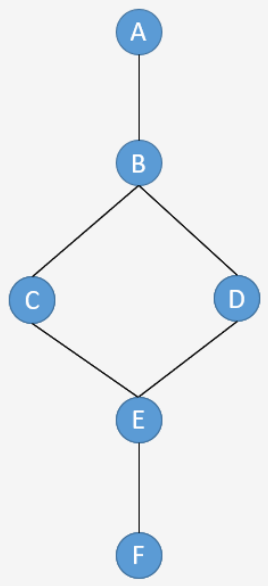
1. We repeat this process for every pair of nodes in the network.

2. We then add up the fractions we computed, and this is the betweenness centrality for node N.



Betweenness Centrality

From C to D, there are two shortest paths: one through B and one through E. Thus, 1÷2=0.5 go through B.



For the remaining pairs—CE, CF, DE, DF, and EF—no shortest paths go through B.

Thus, the fraction for all of these is zero.

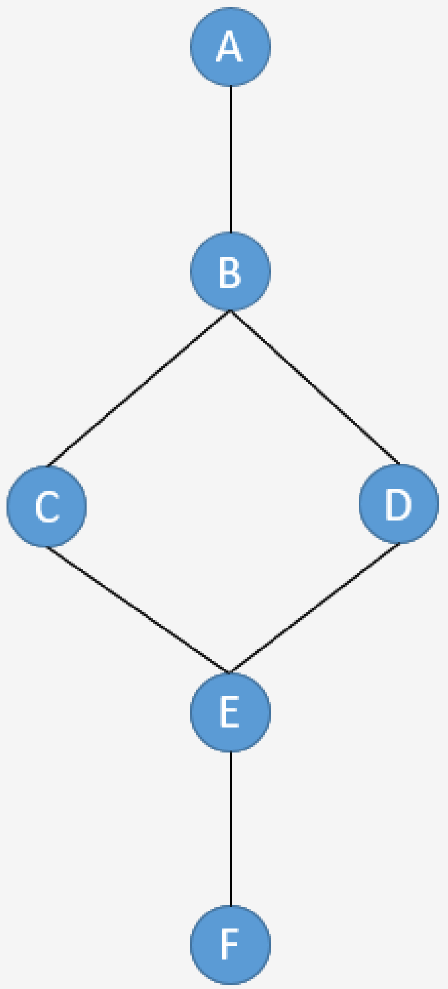
Now we can calculate the betweenness for B:

4×1 (A to all others)$+0.5$ (CD)$+5 \times 0$ (all remaining pairs)$=4+0.5+0=4.5$

Betweenness Centrality

In contrast, the betweenness centrality of A is zero, since no shortest paths between D, C, D, E, and F go through A.

The heights the score, highest Betweenness Centrality.



Which is best centrality measure?

Degree	How many people can this person reach directly?
Betweenness	How likely is this person to be the most direct route between two people in the network?
Closeness	How fast can this person reach everyone in the network?

Big Data 3.0

- IoT Applications
  - Streaming Analytics
  - Edge Computing

Monitoring

1. Sensors and external data sources enable the comprehensive monitoring of:
   - the product's condition
   - the external environment
   - the product's operation and usage

   Monitoring also enables alerts and notifications of changes

Control

2. Software embedded in the product or in the product cloud enables:
   - Control of product functions
   - Personalization of the user experience

Optimization

3. Monitoring and control capabilities enable algorithms that optimize product operation and use in order to:
   - Enhance product performance
   - Allow predictive diagnostics, service, and repair

Autonomy

4. Combining monitoring, control, and optimization allows:
   - Autonomous product operation
   - Self-coordination of operation with other products and systems
   - Autonomous product enhancement and personalization
   - Self-diagnosis and service

Digital Twins

Impact of Big Data

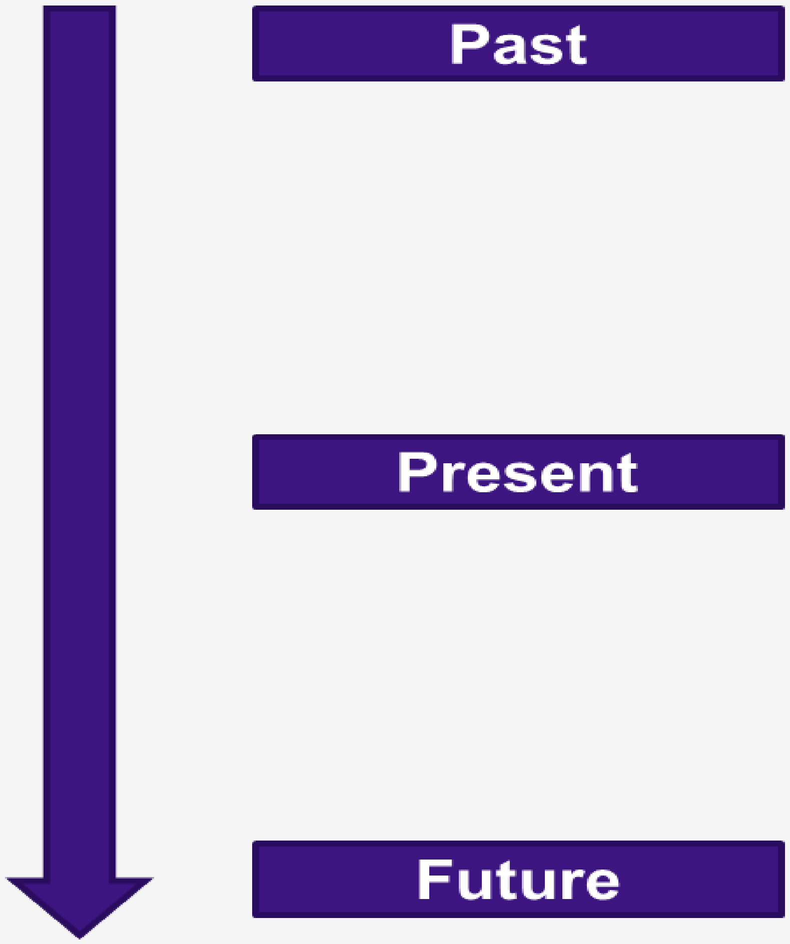
- Personalization marketing
- Better Pricing
- Cost Reduction
- Improved Customer Service

Challenges in big data

- Data Quality
- Data Security
- Privacy
- Investment Justification
- Data Management
- Shortage of Experts

Type of Data Analytics

- **What happened?**
  - Descriptive Analytics (Data aggregation, Summary)
  - What were our total sales this month?
- **Why did it happen?**
  - Diagnostic Analytics (Data discovery, Drill-down)
  - Why have sales gone down?
- **What will happen?**
  - Predictive Analytics (Regression, Neural Network)
  - Forecasting & What If Analysis
- **What do I want to happen?**
  - Prescriptive Analytics (Optimization, Recommendation)
  - Planning & Targets



Thought Exercise

- **What are the skills required to become a successful data scientist?**

What Is Artificial Intelligence? (1 of 3)

- Grand vision
  - Computer hardware and software systems that are as "smart" as humans
- Realistic vision
  - Systems that take data inputs, process them, and produce outputs (like all software programs) and that can perform many complex tasks that would be difficult or impossible for humans to perform.

What Is Artificial Intelligence? (2 of 3)

- Examples:
  - Recognize millions of faces in seconds
  - Interpret millions of CT scans in minutes
  - Analyze millions of financial records
  - Detect patterns in very large Big Data databases
  - Improve their performance over time (“learn”)
  - Navigate a car in certain limited conditions
  - Respond to questions from humans (natural language); speech activated assistants like Siri, Alexa, and Cortana

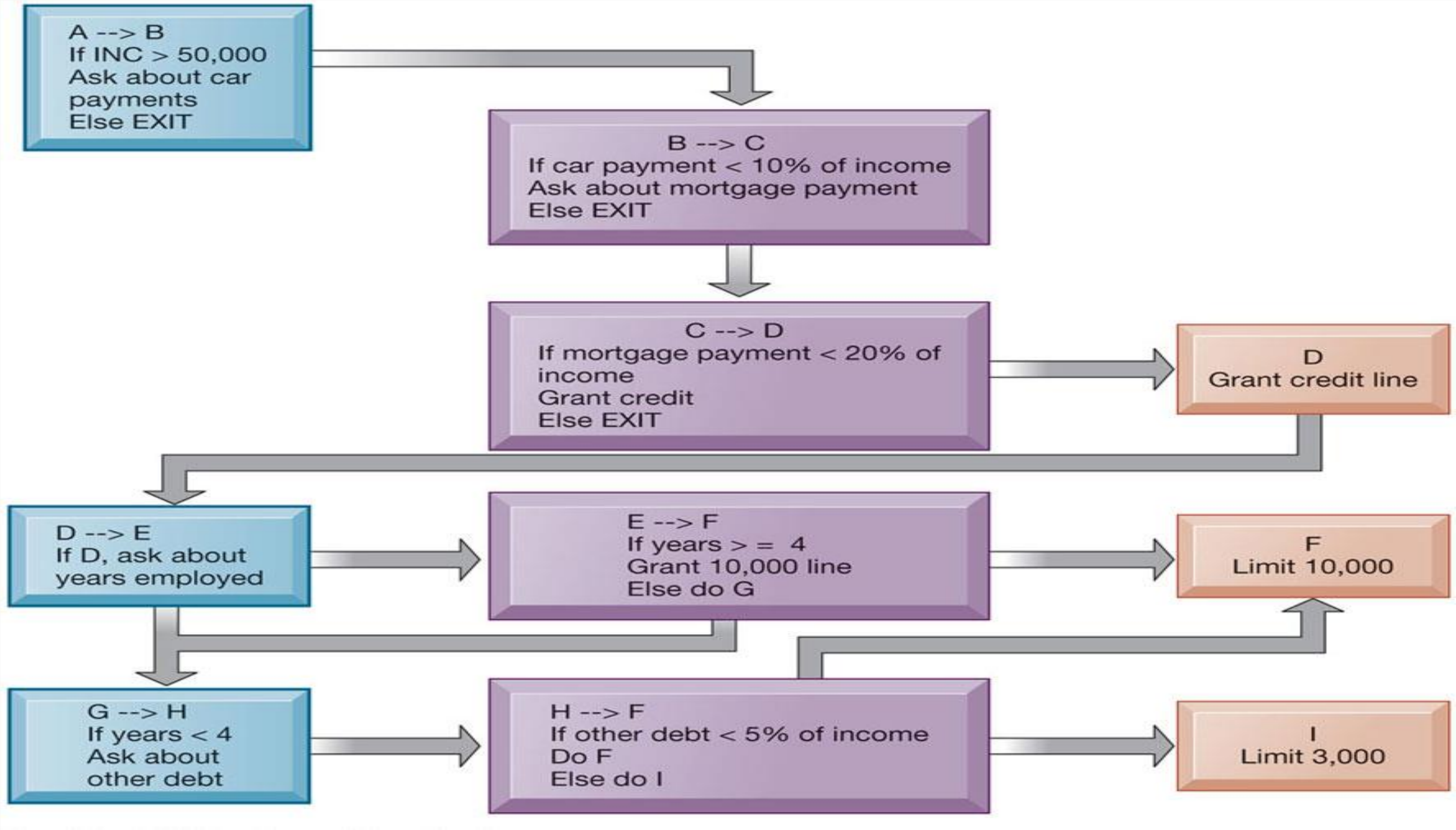
What Is Artificial Intelligence? (3 of 3)

- Major Types of AI
  - Expert systems
  - Machine learning
  - Neural networks and deep learning networks
  - Genetic algorithms
  - Natural language Processing
  - Computer vision
  - Robotics

Capturing Knowledge: Expert Systems

- Capture tacit knowledge in very specific and limited domain of human expertise
- Capture knowledge as set of rules
- Typically perform limited tasks
  - Diagnosing malfunctioning machine
  - Determining whether to grant credit for loan
- Used for discrete, highly structured decision making
- Knowledge base: Set of hundreds or thousands of rules
- Inference engine: Strategy used to search knowledge base
  - Forward chaining
  - Backward chaining

Rules in an Expert System



Machine Learning

- Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming.
  - Recognizing patterns
  - Supervised vs. unsupervised learning
- Contemporary examples
  - Google searches
  - Recommender systems on Amazon, Netflix

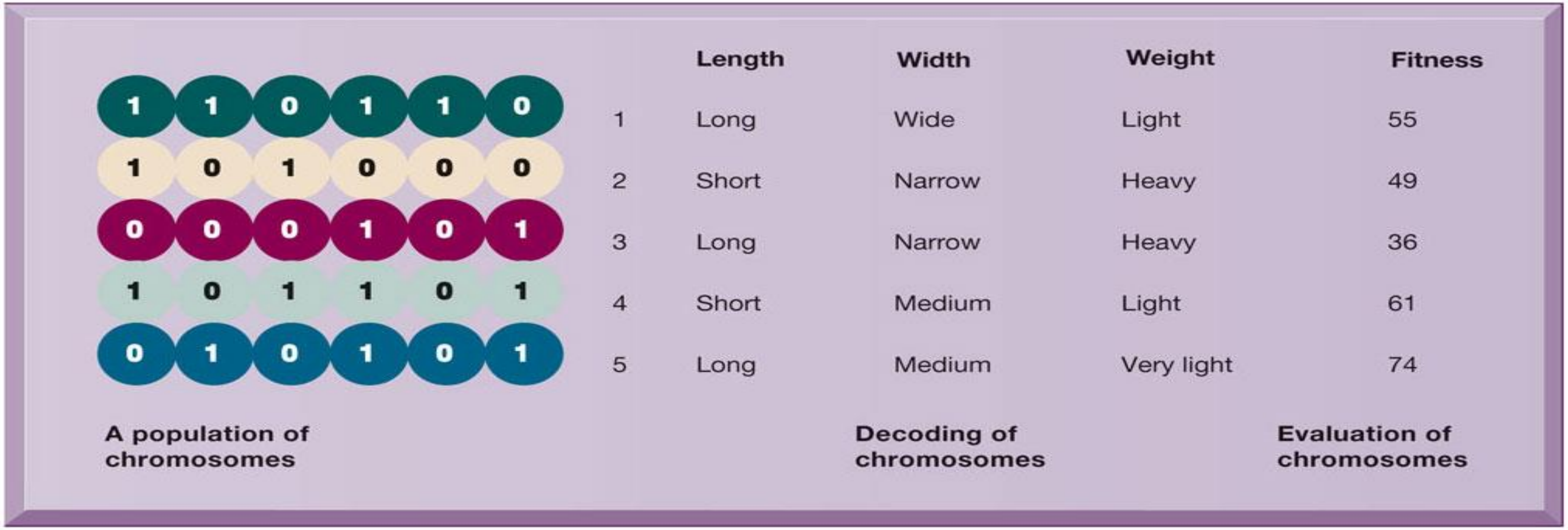
Supervised vs. Unsupervised learning

Supervised Learning	**Unsupervised Learning**
Input data is labelled	Input data is unlabeled
Uses training dataset	Uses just input dataset
Used for prediction	Used for analysis
E.g., classification, regression	Clustering, density estimation and dimensionality reduction

Genetic Algorithms

- Useful for finding optimal solution for specific problem by examining very large number of possible solutions for that problem
- Conceptually based on process of evolution
  - Search among solution variables by changing and reorganizing component parts using processes such as inheritance, mutation, and selection
- Used in optimization problems (minimization of costs, efficient scheduling, optimal jet engine design) in which hundreds or thousands of variables exist
- Able to evaluate many solution alternatives quickly

The Components of a Genetic Algorithm



Natural Language Processing

- Understand, and speak in natural language. Read natural language and translate
- Typically today based on machine learning, aided by very large databases of common phrases and sentences in a given language
- Example: Google Translate
- Spam filtering systems
- Customer call center interactions: What is the customer's problem? What solutions worked in the past?
- Digital assistances: Sire, Alexa, Cortana, Google Assistant
- Not useful for an ordinary common sense human conversation but can be very useful in limited domains, e.g. interacting with your car's heating system.

Computer Vision Systems

- Digital image systems that create a digital map of an image (like a face, or a street sign), and recognize this image in large data bases of images in near real time
- Every image has a unique pattern of pixels
- Facebook's DeepFace can identify friends in photos across their system, and the entire web
- Autonomous vehicles can recognize signs, road markers, people, animals, and other vehicles with good reliability
- Industrial machine (robot) vision
- Passport control at airports
- Identifying people in crowds

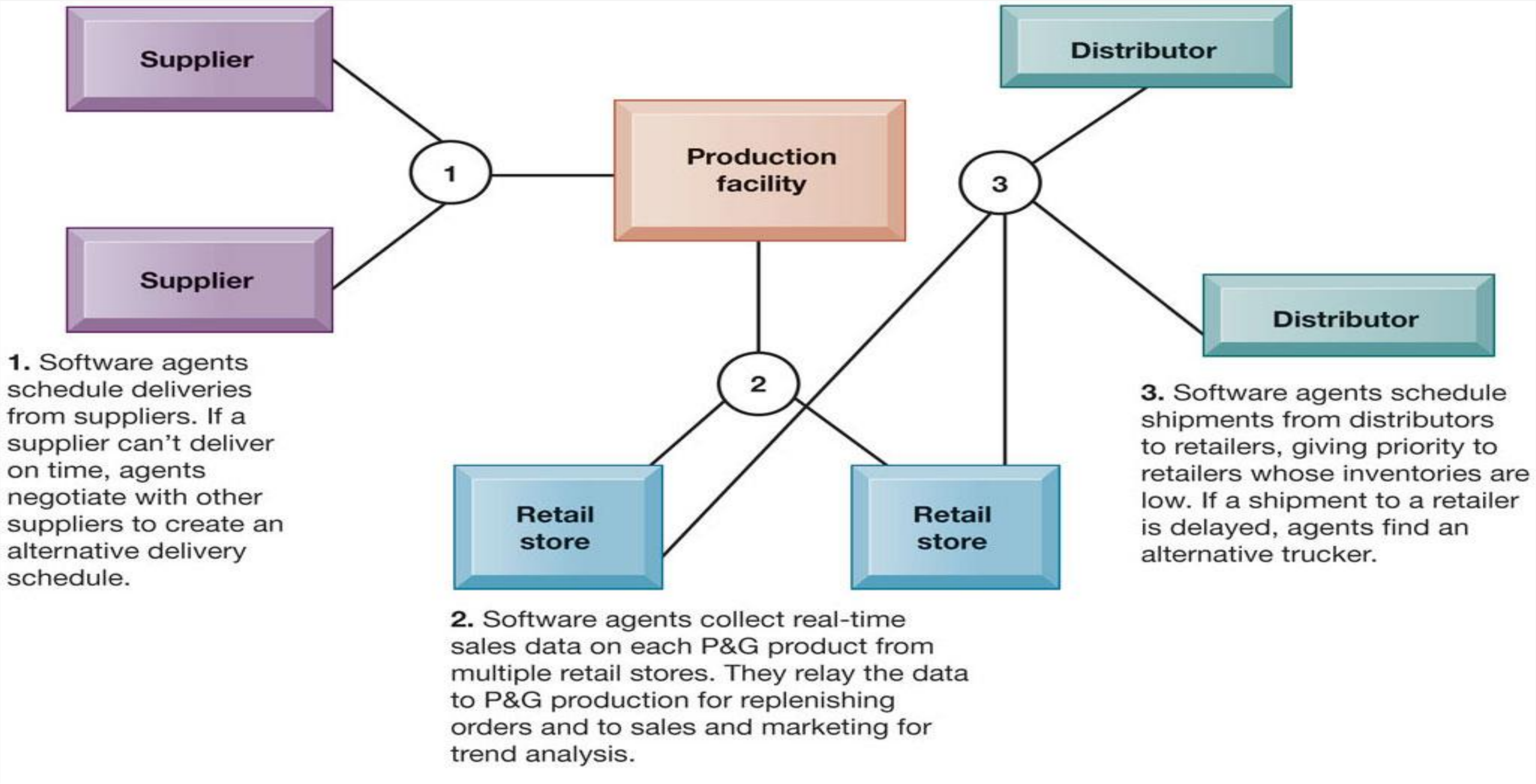
Robotics

- Design, construction, and operation of machines that can substitute for humans in many factory, office, and home applications (home vacuums).
- Generally programmed to perform specific and detailed actions in limited domains, e.g. robots spray paint autos, and assemble certain parts, welding, heavy assembly movement.
- Used in dangerous situations like bomb disposal
- Surgical robots are expanding their capabilities

Intelligent Agents

- Work without direct human intervention to carry out repetitive, predictable tasks
  - Deleting junk e-mail
  - Finding cheapest airfare
- Use limited built-in or learned knowledge base
  - Some are capable of self-adjustment, for example: Siri
- Chatbots
- Agent-based modeling applications:
  - Model behavior of consumers, stock markets, and supply chains; used to predict spread of epidemics

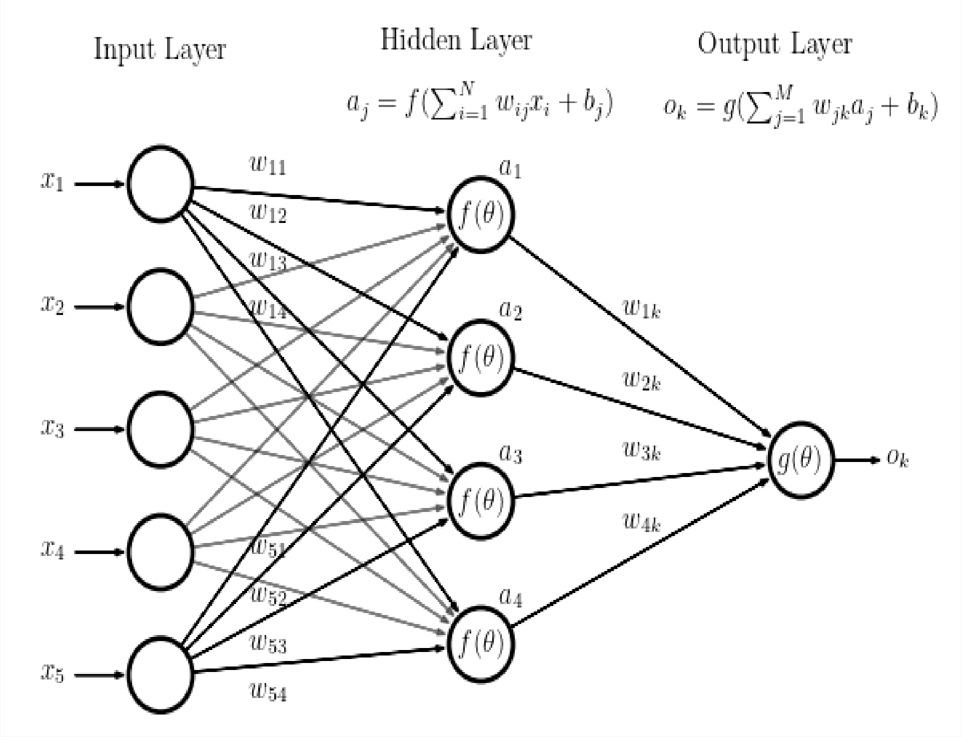
Intelligent Agents in P&G's Supply Chain Network



Neural Networks

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates.

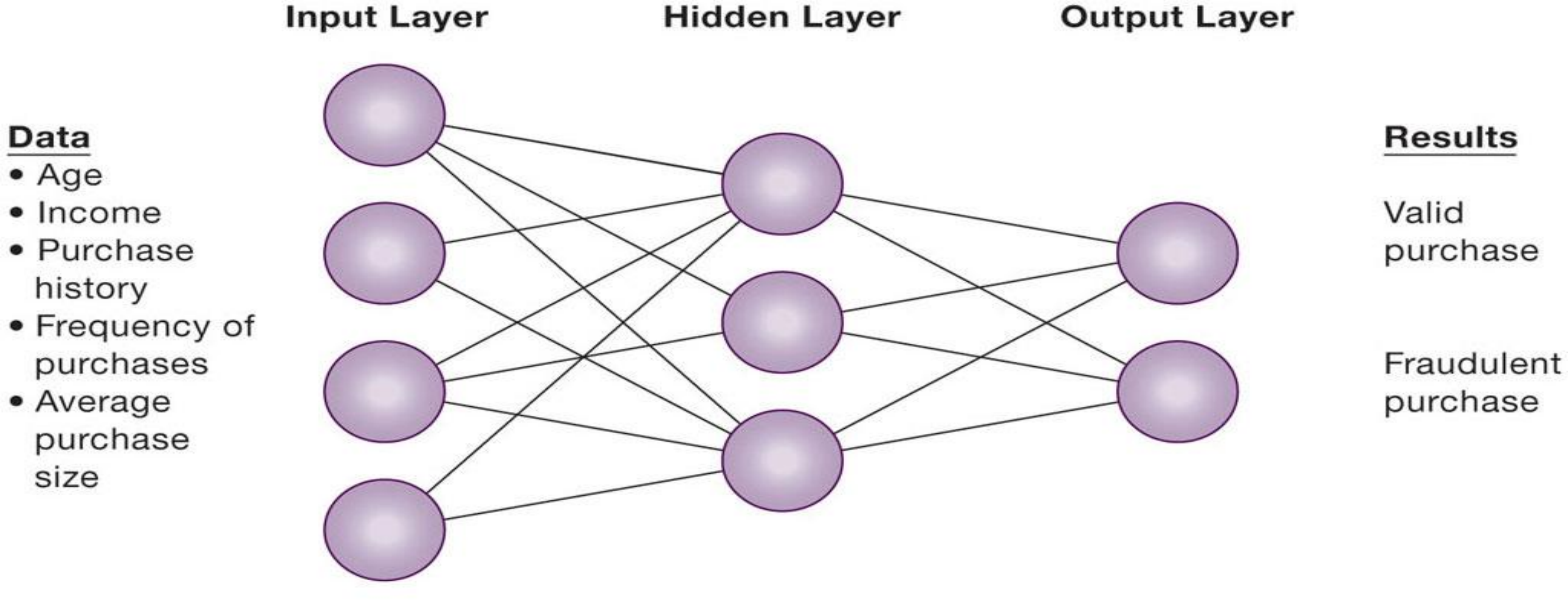
Neural nets are a means of doing machine learning, in which a computer learns to perform some task by analyzing training examples.



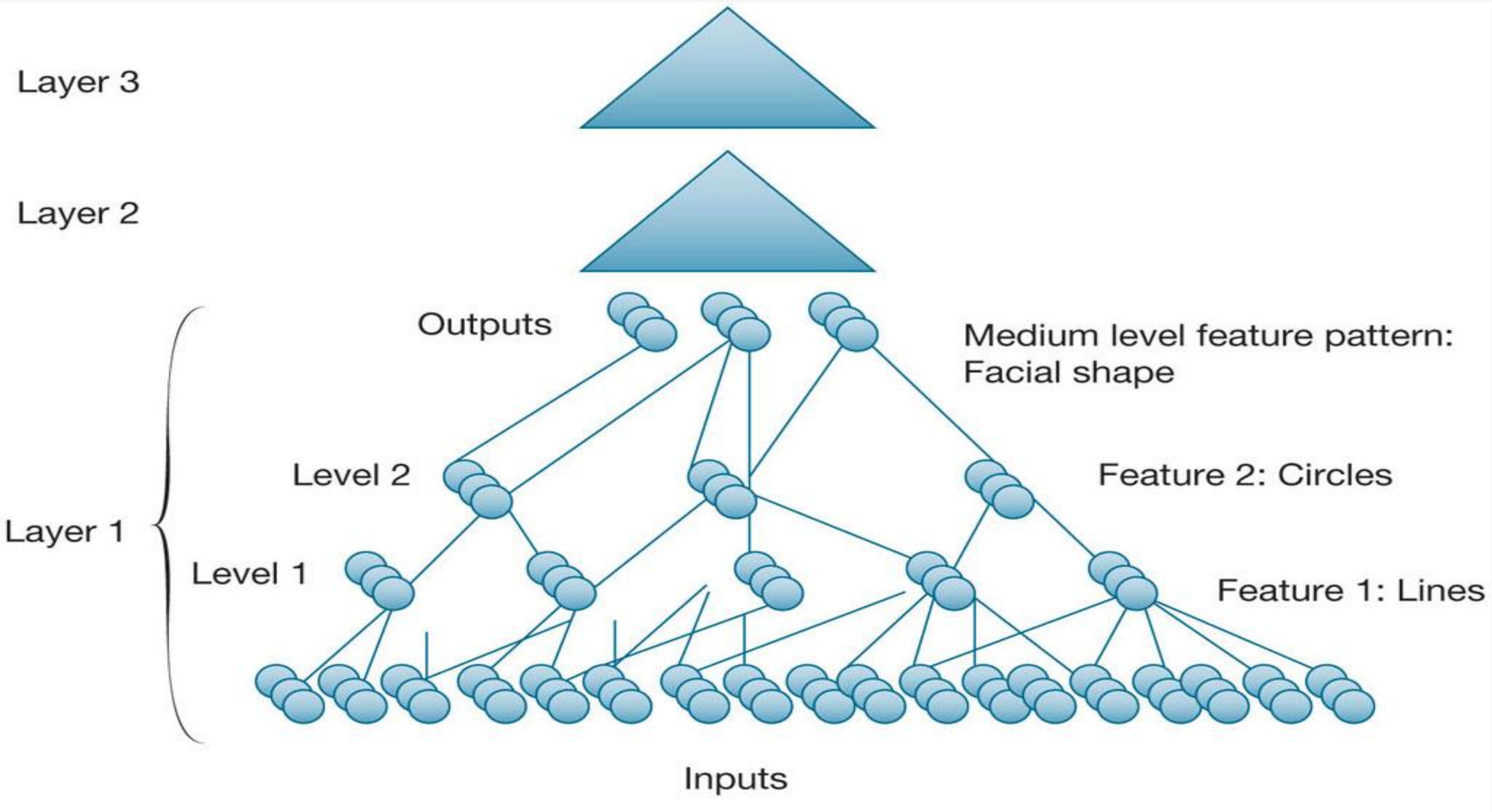
Neural Networks

- **Neural networks** are a series of algorithms that mimic the operations of a human brain to recognize relationships between vast amounts of data.
- Find patterns and relationships in massive amounts of data too complicated for humans to analyze
- "Learn" patterns by searching for relationships, building models, and correcting over and over again
- Humans "train" network by feeding it data inputs for which outputs are known, to help neural network learn solution by example from human experts.
- Used in medicine, science, and business for problems in pattern classification, prediction, financial analysis, and control and optimization

How a Neural Network Works



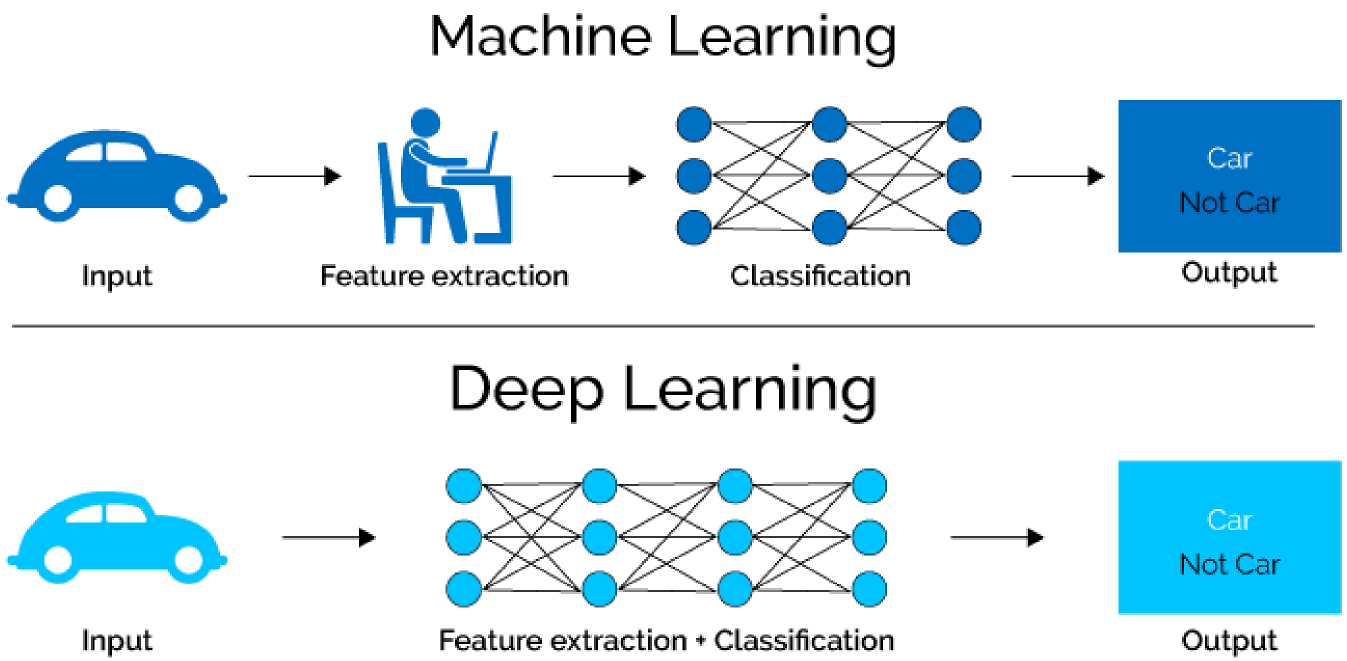
A Deep Learning Network



Deep learning is a specific method of machine learning that incorporates neural networks in successive layers to learn from data in an iterative manner.

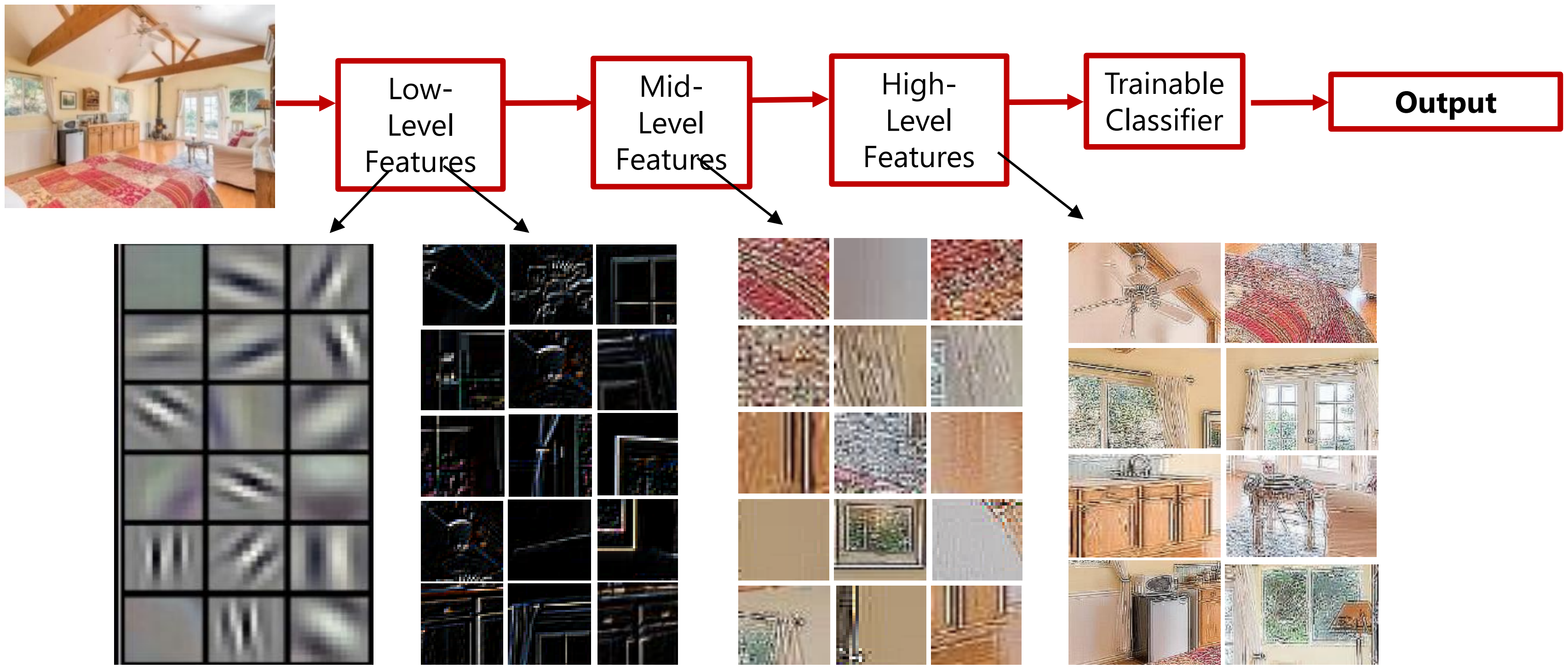
ML vs. Deep Learning

- ***Deep learning*** (DL) is a machine learning subfield that uses multiple layers for learning data representations
  - DL is exceptionally effective at learning patterns



ML vs. Deep Learning

- DL applies a multi-layer process for learning rich hierarchical features (i.e., data representations)
  - Input image pixels → Edges → Textures → Parts → Objects

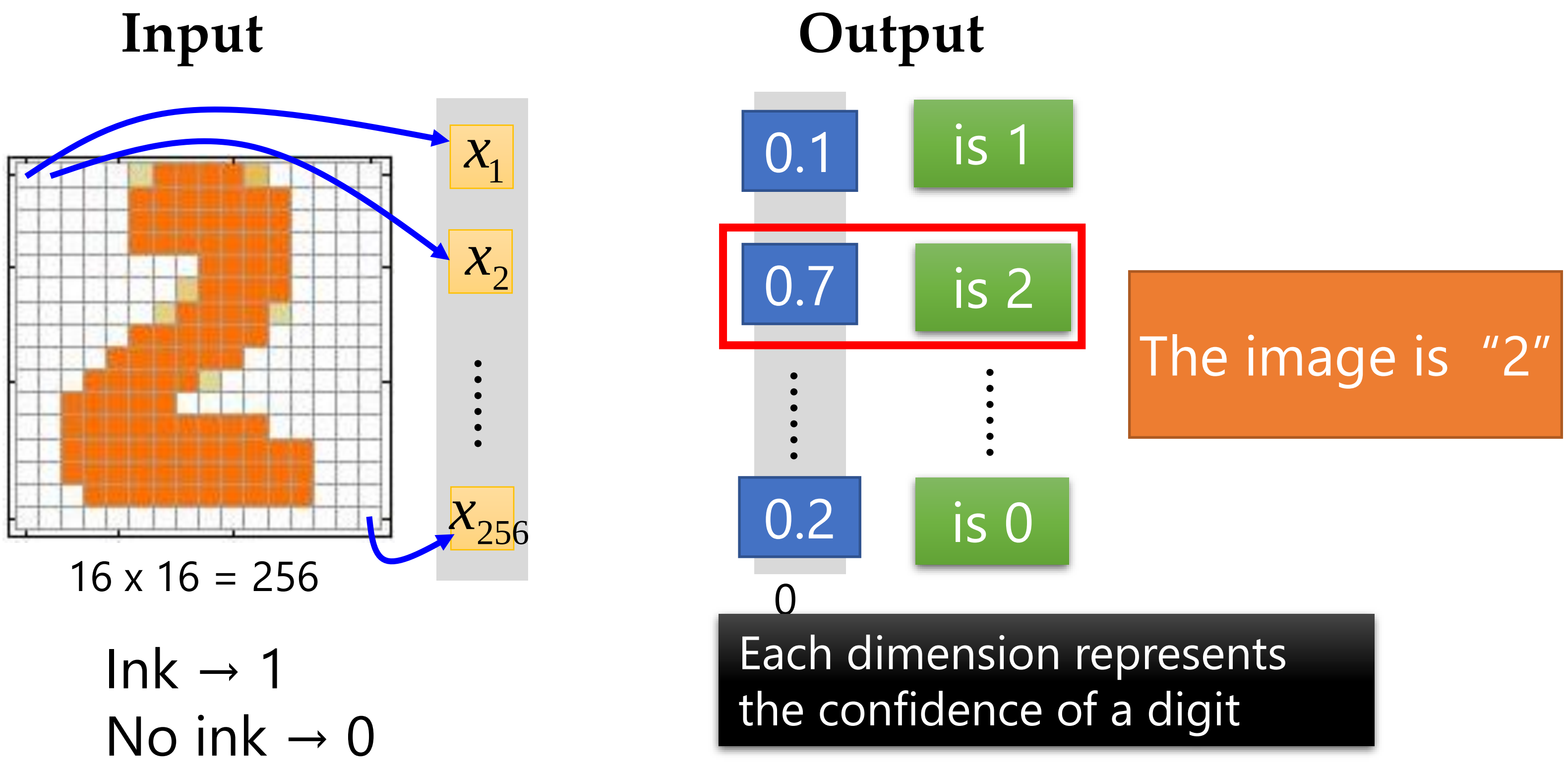


Why is DL Useful?

- DL provides a flexible, learnable framework for representing visual, text, linguistic information
  - Can learn in supervised and unsupervised manner
- DL represents an effective end-to-end learning system
- Requires large amounts of training data
- DL has outperformed other ML techniques
  - First in vision and speech, then NLP, and other applications

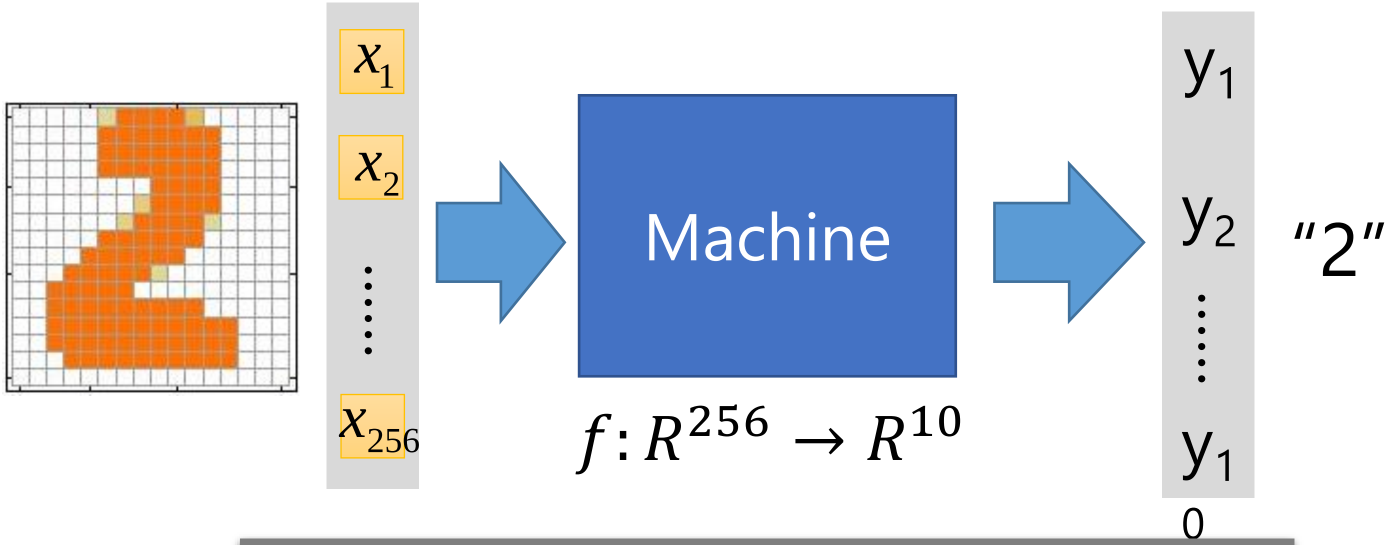
Introduction to Neural Networks

- Handwritten digit recognition (MNIST dataset)
  - The intensity of each pixel is considered an input element
  - Output is the class of the digit



Introduction to Neural Networks

- Handwritten digit recognition



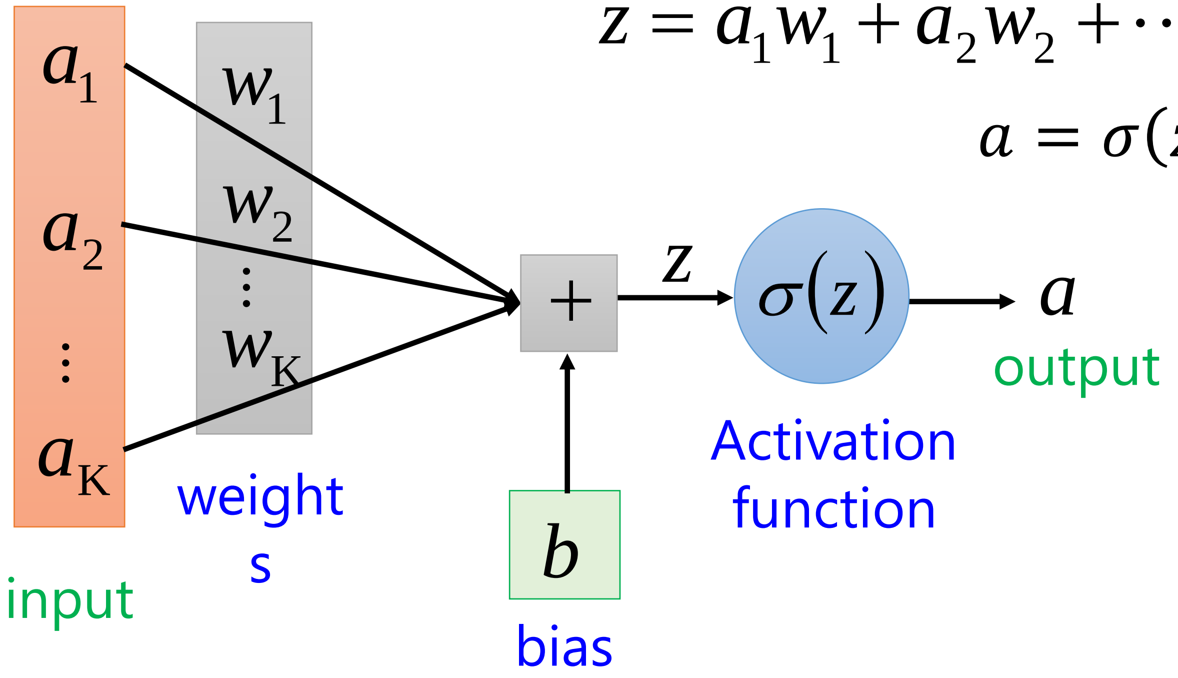
The function f is represented by a neural network

Elements of Neural Networks

- NNs consist of hidden layers with neurons (i.e., computational units)
- A single neuron maps a set of inputs into an output number, or $f: R^K \rightarrow R$

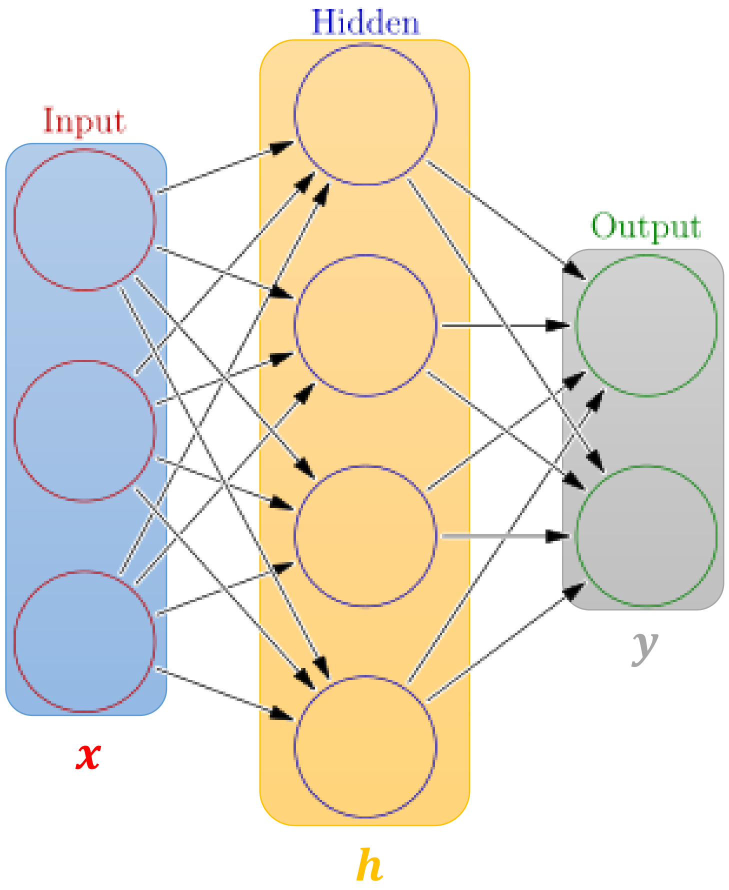
$$z = a_1 w_1 + a_2 w_2 + \cdots + a_K w_K + b$$

$$a = \sigma(z)$$



Elements of Neural Networks

- A NN with one hidden layer and one output layer



Weights Biases

$$\textbf{hidden layer } \boldsymbol{h} = \boldsymbol{\sigma}(\mathbf{W_1}\boldsymbol{x} + \boldsymbol{b_1})$$

$$\textbf{output layer } \boldsymbol{y} = \boldsymbol{\sigma}(\boldsymbol{W_2}\boldsymbol{h} + \boldsymbol{b_2})$$

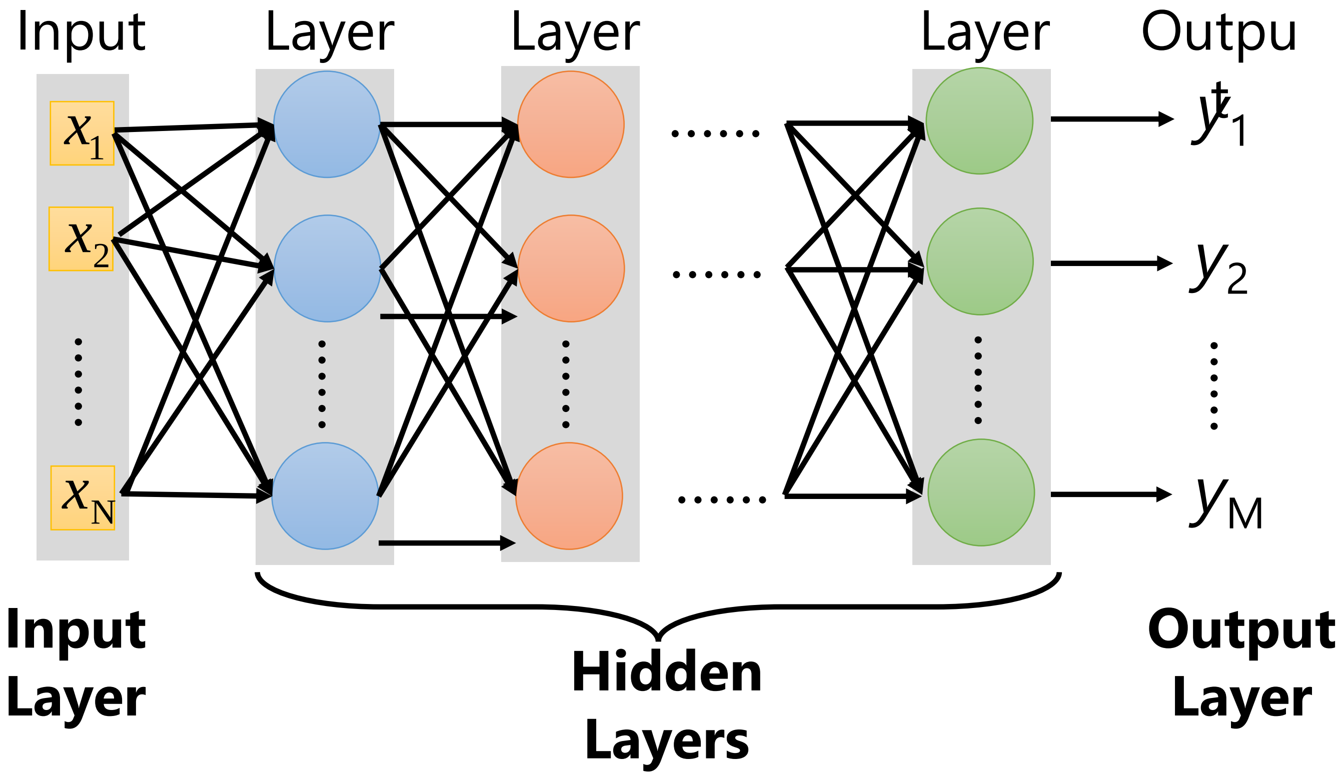
Activation functions

4 + 2 = 6 neurons (not counting inputs)
[3 × 4] + [4 × 2] = 20 weights
4 + 2 = 6 biases

26 learnable parameters

Elements of Neural Networks

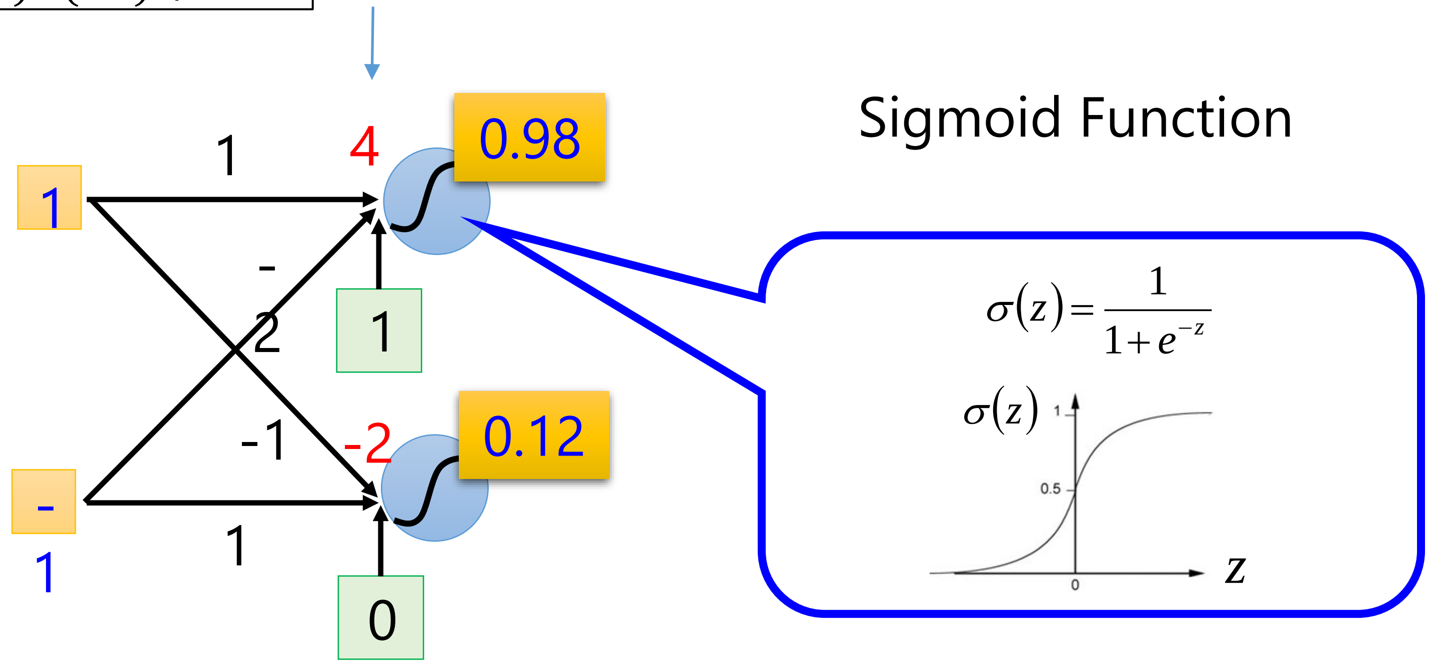
- Deep NNs have many hidden layers
  - Fully-connected (dense) layers (a.k.a. Multi-Layer Perceptron or MLP)
  - Each neuron is connected to all neurons in the succeeding layer



Elements of Neural Networks

- A simple network, toy example

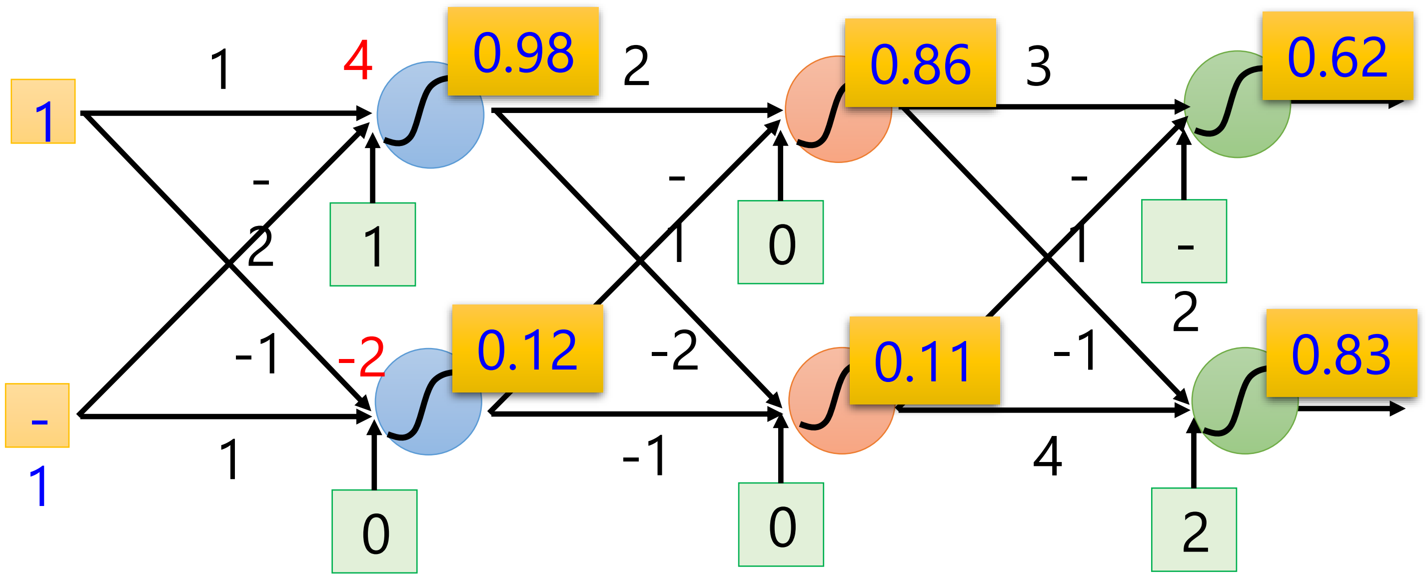
$(1 \cdot 1) + (-1) \cdot (-2) + 1 = 4$



$1 \cdot (-1) + (-1) \cdot 1 + 0 = -2$

Elements of Neural Networks

- A simple network, toy example (cont'd)
  - For an input vector $[1 \quad -1]^T$, the output is $[0.62 \quad 0.83]^T$



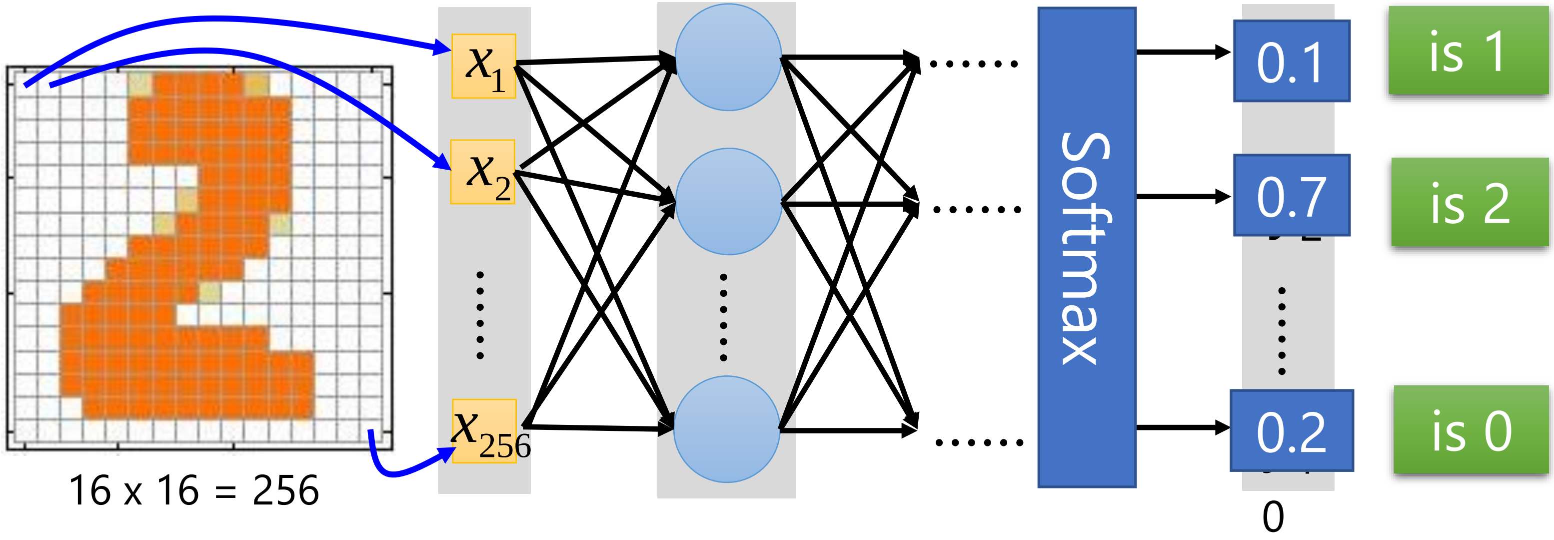
$$f: R^2 \rightarrow R^2 \qquad f\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 0.62 \\ 0.83 \end{bmatrix}$$

Training NNs

- The network ***parameters*** θ include the weight matrices and bias vectors from all layers

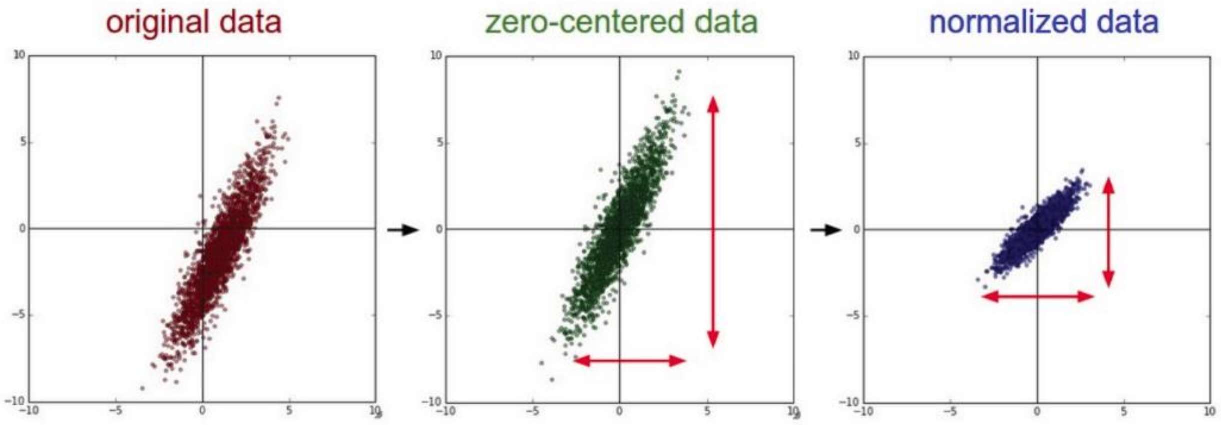
$$\theta = \{W^1, b^1, W^2, b^2, \cdots W^L, b^L\}$$

  - Often, the model parameters θ are referred to as weights
- Training a model to learn a set of parameters θ that are optimal (according to a criterion) is one of the greatest challenges in ML



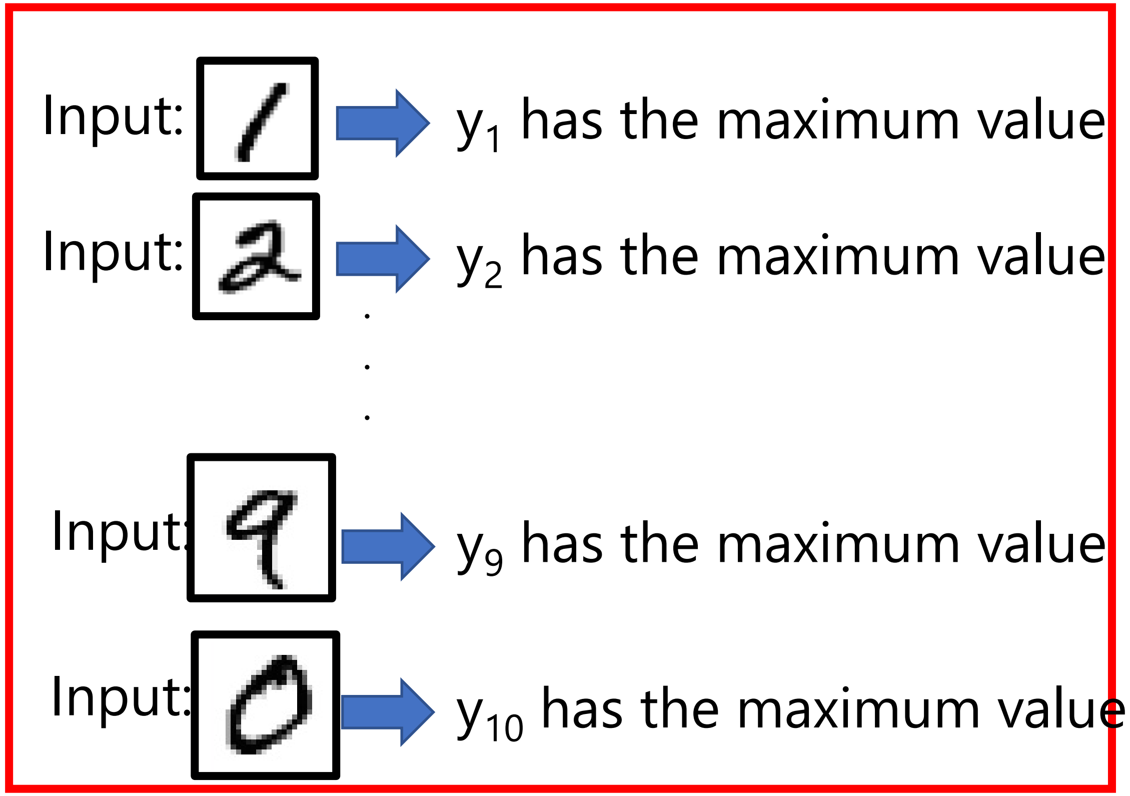
Training NNs

- ***Data preprocessing*** - helps convergence during training
  - Mean subtraction, to obtain zero-centered data
    - Subtract the mean for each individual data dimension (feature)
  - Normalization
    - Divide each feature by its standard deviation
      - To obtain standard deviation of 1 for each data dimension (feature)
    - Or, scale the data within the range [0,1] or [-1, 1]
      - E.g., image pixel intensities are divided by 255 to be scaled in the [0,1] range



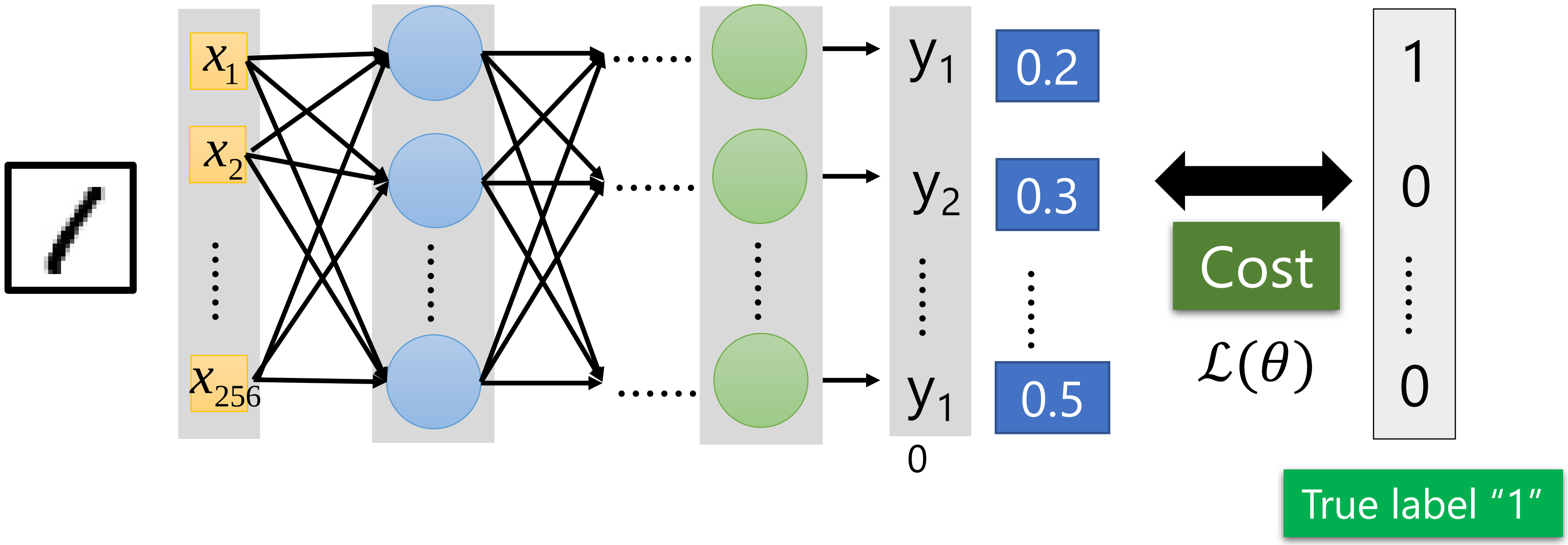
Training NNs

- To train a NN, set the parameters θ such that for a training subset of images, the corresponding elements in the predicted output have maximum values



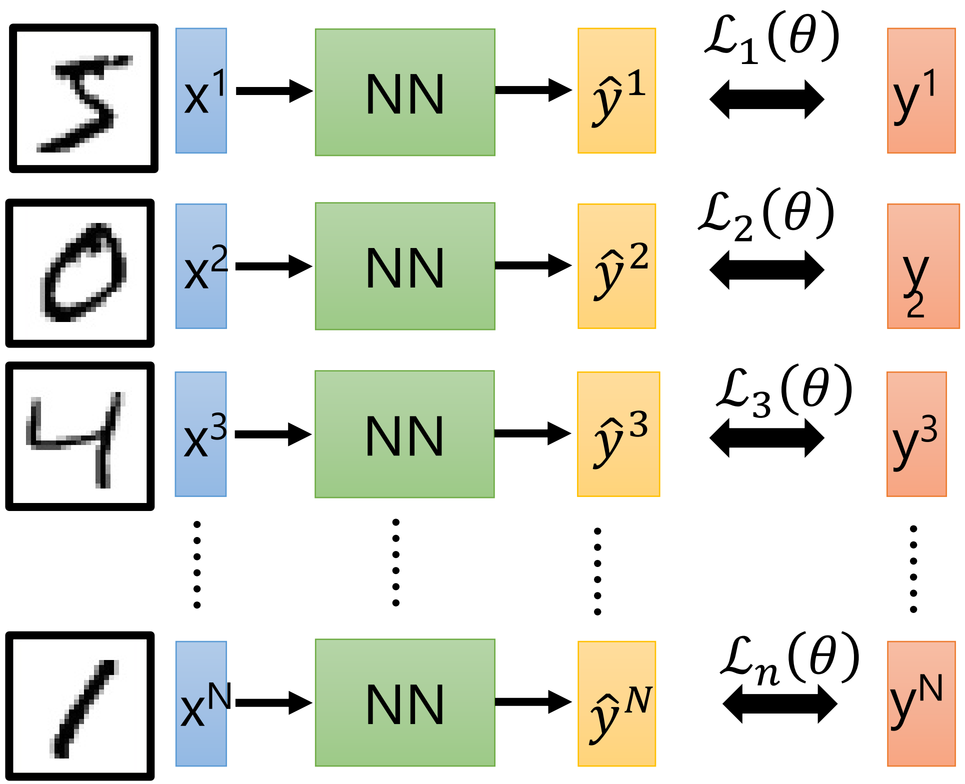
Training NNs

- Define a ***loss function***/objective function/cost function $\mathcal{L}(\theta)$ that calculates the difference (error) between the model prediction and the true label
  - E.g., $\mathcal{L}(\theta)$ can be mean-squared error, cross-entropy, etc.



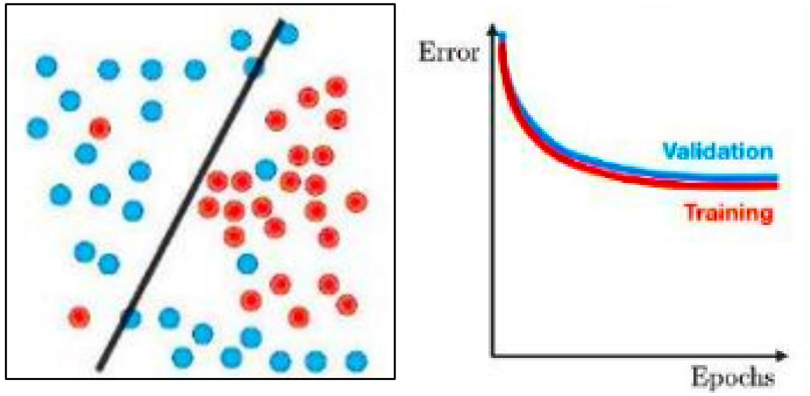
Training NNs

- For a training set of N images, calculate the total loss overall all images: $\mathcal{L}(\theta) = \sum_{n=1}^{N} \mathcal{L}_n(\theta)$
- Find the optimal parameters θ^* that minimize the total loss $\mathcal{L}(\theta)$

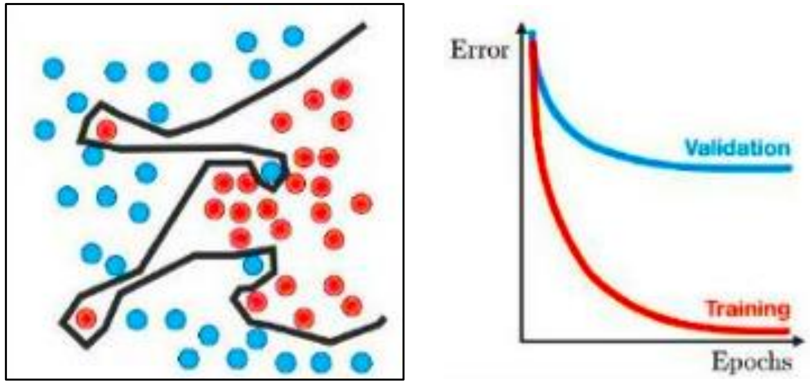


Generalization

- ***Underfitting***
  - The model is too "simple" to represent all the relevant class characteristics
  - E.g., model with too few parameters
  - Produces high error on the training set and high error on the validation set

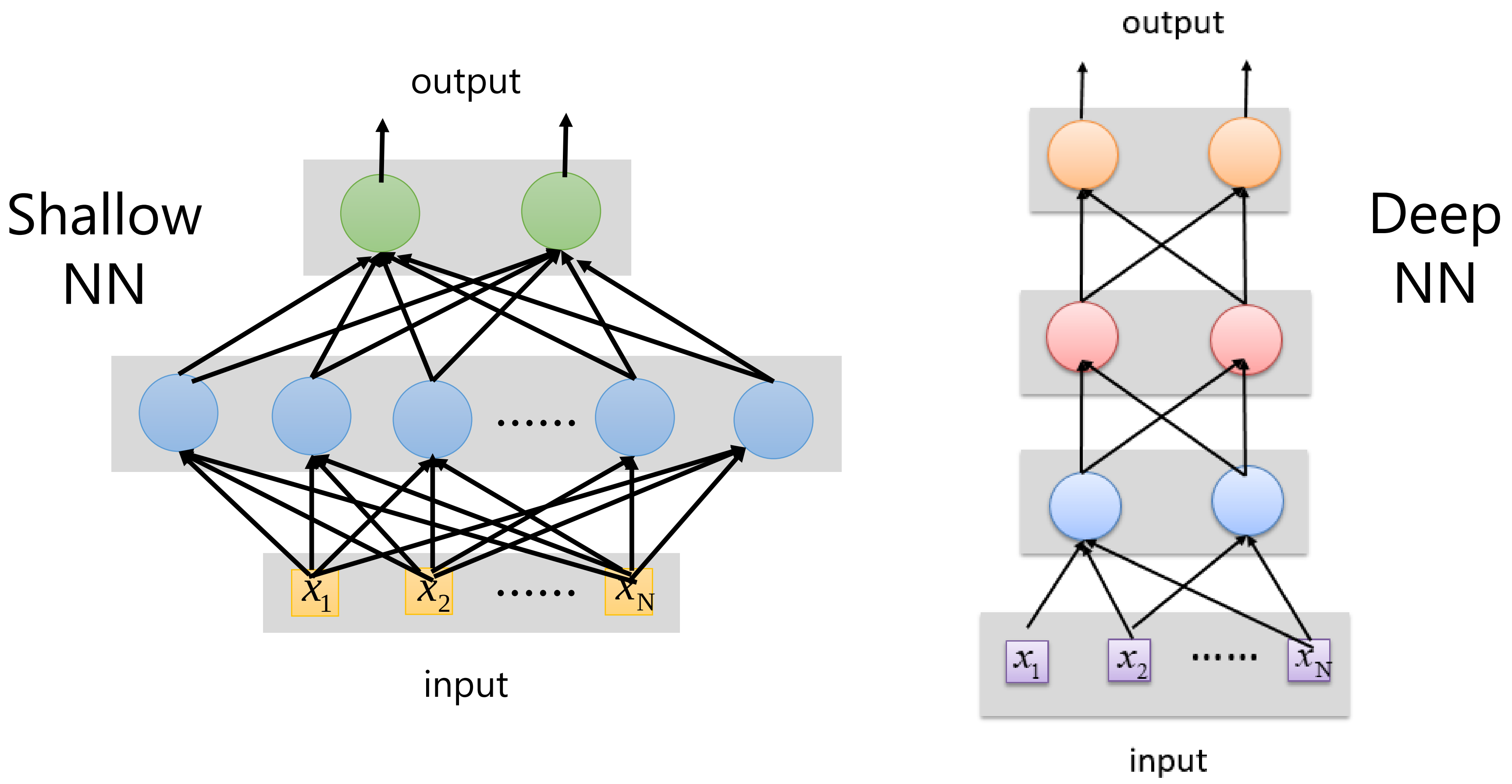


- ***Overfitting***
  - The model is too "complex" and fits irrelevant characteristics (noise) in the data
  - E.g., model with too many parameters
  - Produces low error on the training error and high error on the validation set



Deep vs Shallow Networks

- Deeper networks perform better than shallow networks
  - But only up to some limit: after a certain number of layers, the performance of deeper networks plateaus

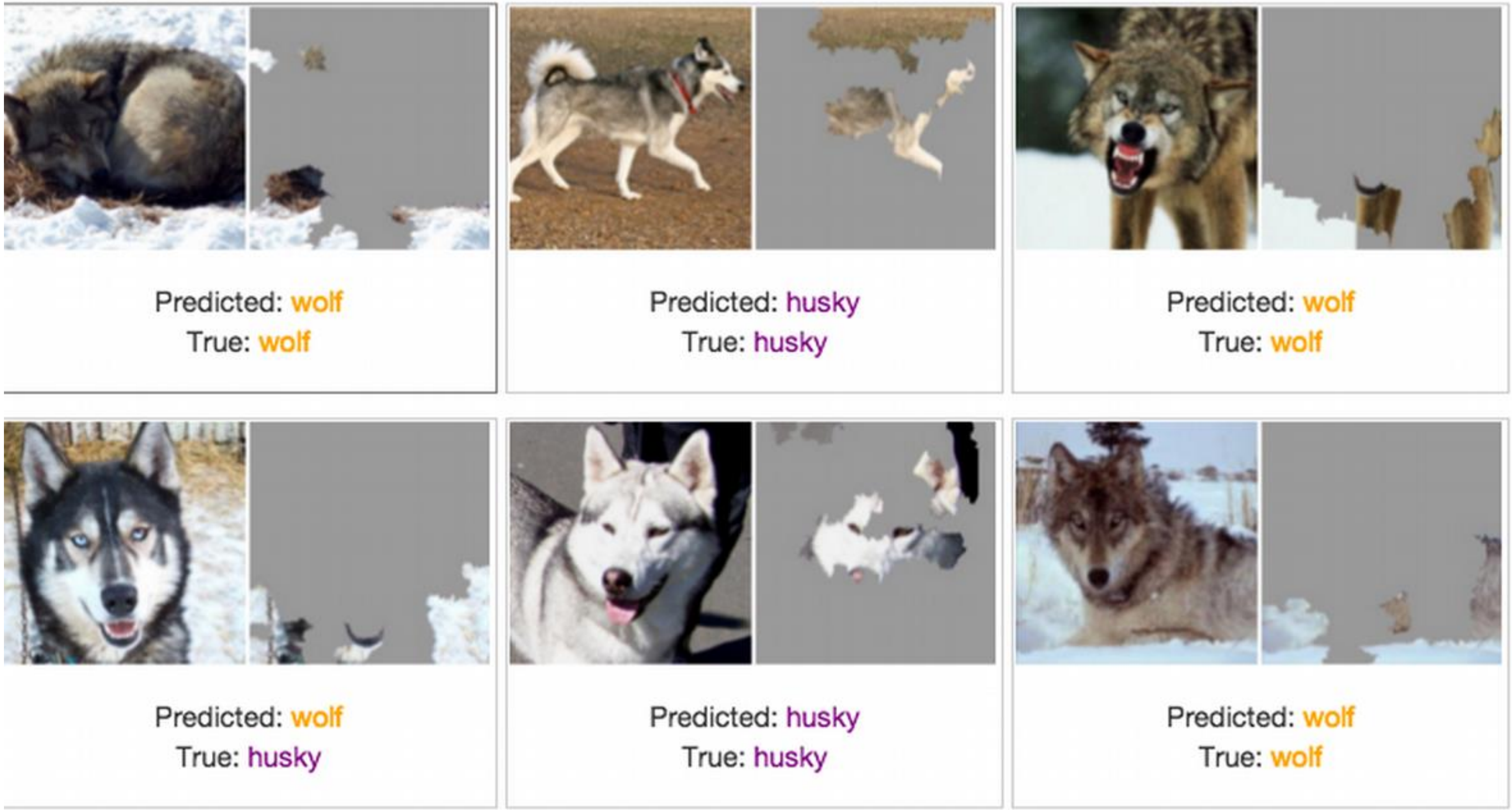


Neural Networks

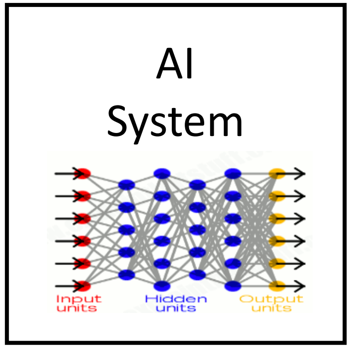
- *Convolutional neural networks* (CNNs) were primarily designed for image data
- *Recurrent NNs* are used for modeling sequential data and data with varying length of inputs and outputs
  - Videos, text, speech, DNA sequences, human skeletal data

- *Long Short-Term Memory (LSTM)* networks are a variant of RNNs
- LSTM mitigates the vanishing/exploding gradient problem
  - Solution: a Memory Cell, updated at each step in the sequence

Incorrect prediction


Predicted: wolf
True: wolf
Predicted: husky
True: husky
Predicted: wolf
True: wolf
Predicted: wolf
True: husky
Predicted: husky
True: husky
Predicted: wolf
True: wolf

The Need for Explainable AI

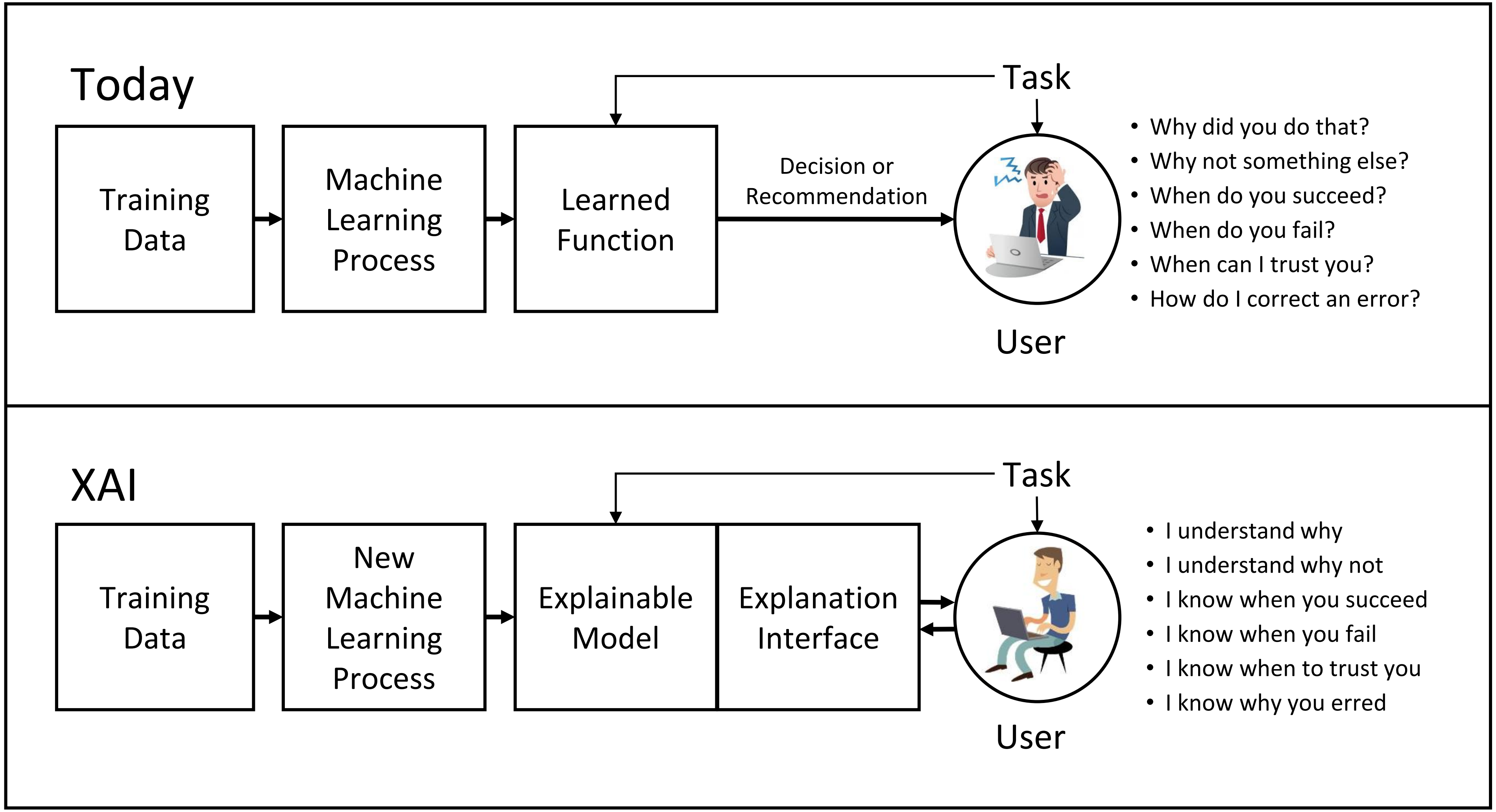


- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

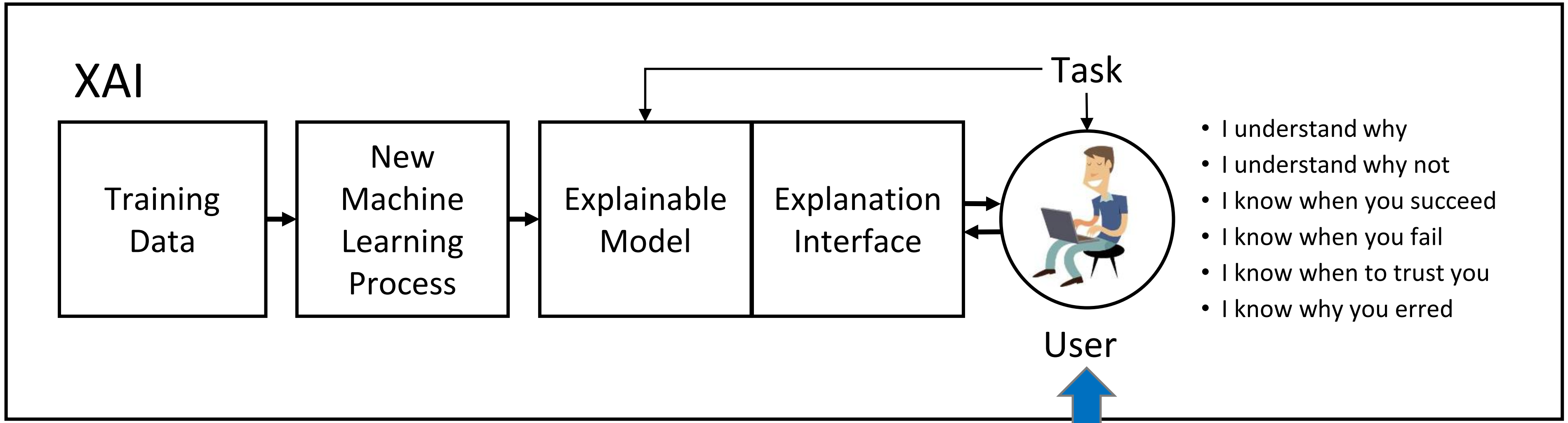
- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

- The current generation of AI systems offer tremendous benefits, but their effectiveness will be limited by the machine's inability to explain its decisions and actions to users.
- Explainable AI will be essential if users are to understand, appropriately trust, and effectively manage this incoming generation of artificially intelligent partners.

XAI Concept


Today
Training Data
Machine Learning Process
Learned Function
Task
Decision or Recommendation
User
• Why did you do that?
• Why not something else?
• When do you succeed?
• When do you fail?
• When can I trust you?
• How do I correct an error?
XAI
Training Data
New Machine Learning Process
Explainable Model
Explanation Interface
Task
User
• I understand why
• I understand why not
• I know when you succeed
• I know when you fail
• I know when to trust you
• I know why you erred

XAI Concept



- The target of XAI is an end user who:
  - depends on decisions, recommendations, or actions of the system
  - needs to understand the rationale for the system's decisions to understand, appropriately trust, and effectively manage the system
- The XAI concept is to:
  - provide an explanation of individual decisions
  - enable understanding of overall strengths & weaknesses
  - convey an understanding of how the system will behave in the future
  - convey how to correct the system's mistakes (perhaps)

Explainable AI

- Reference Books:
- Explainable AI: Foundations, Methodologies and Applications

Emphasis and Scope of XAI Research