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.

Type of Data

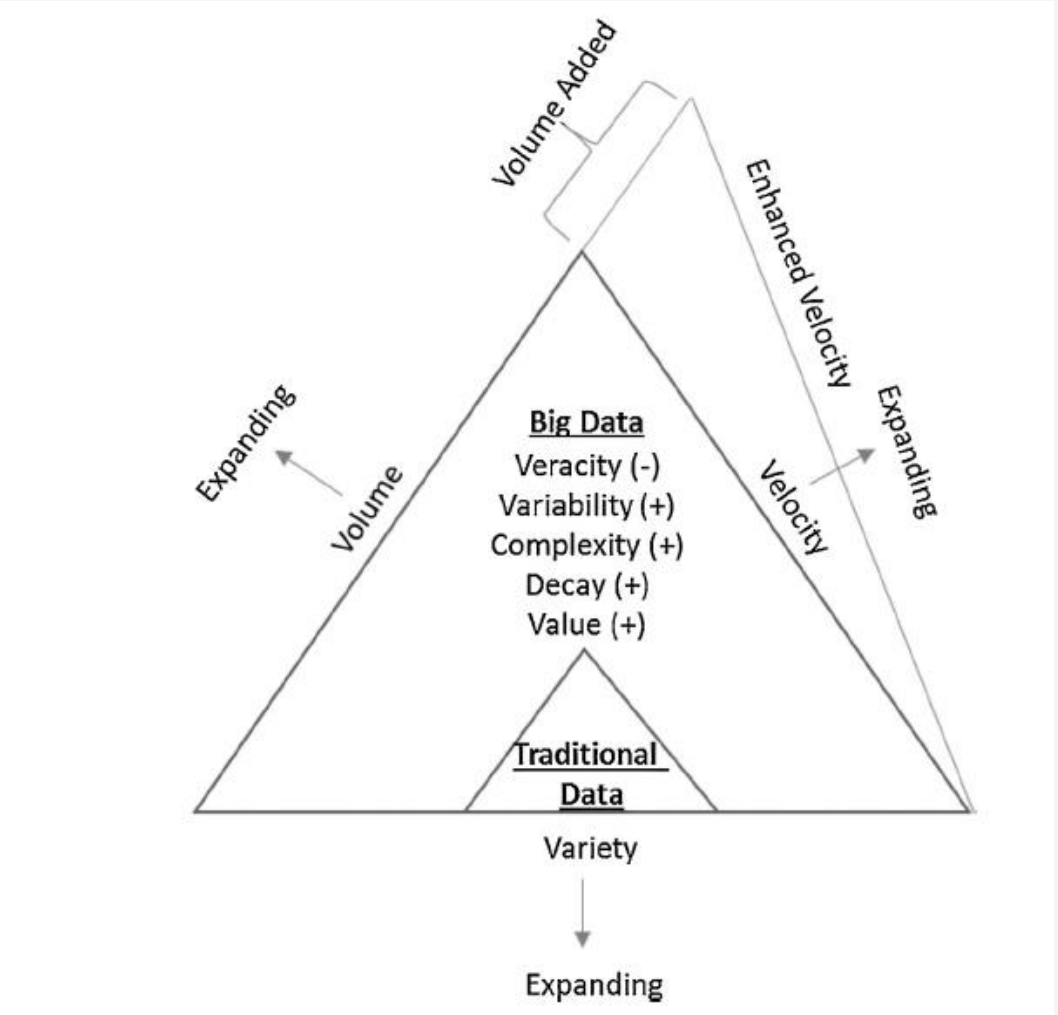
- **Structured**
  - Tables
- **Semi-Structured**
  - XML, JSON
  - Clickstream data, Ecommerce, APIs
- **Unstructured**
  - Audio, Video, Images, Text, Sensory Data
  - IoT Devices, Social Media

Data

- Storage
  - SQL, NoSQL
- Processing
  - Hadoop

Dimensions of big data

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



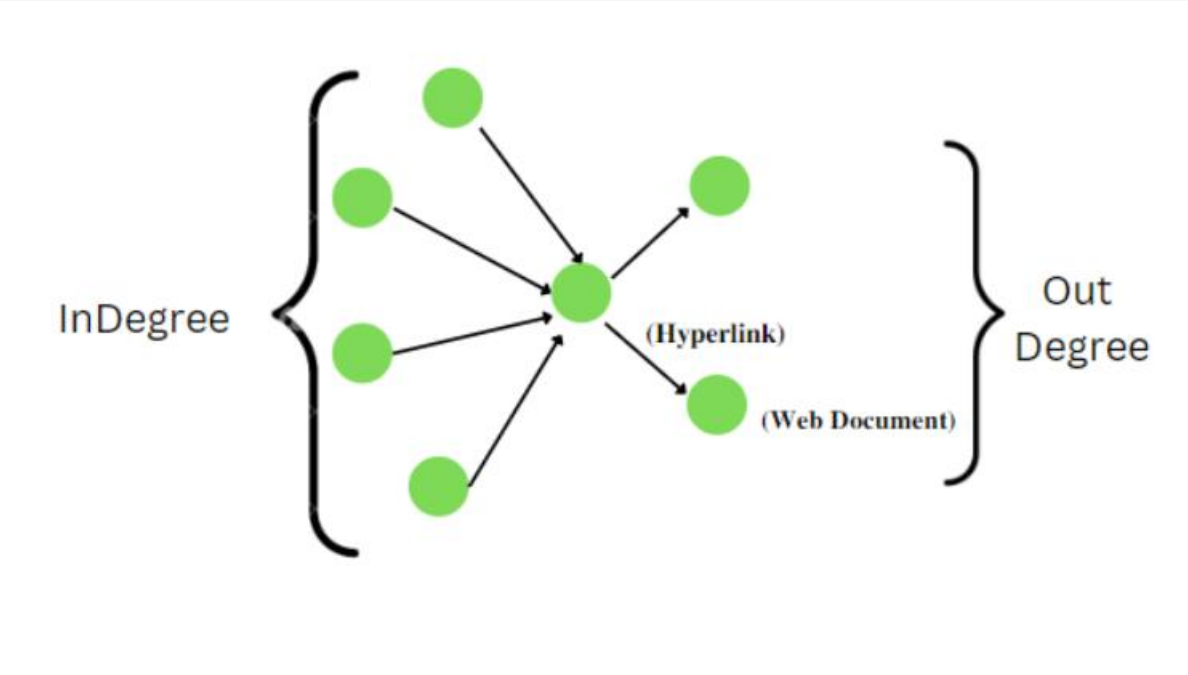
Evolution of big data and data analytics

- 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

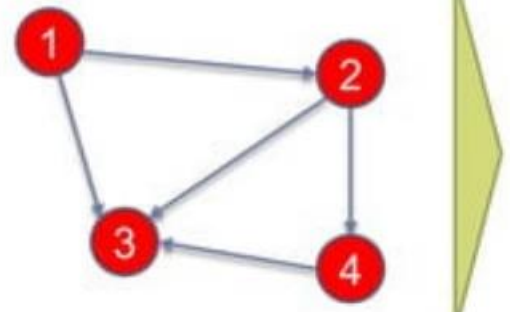


Big Data 2.0

- 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)



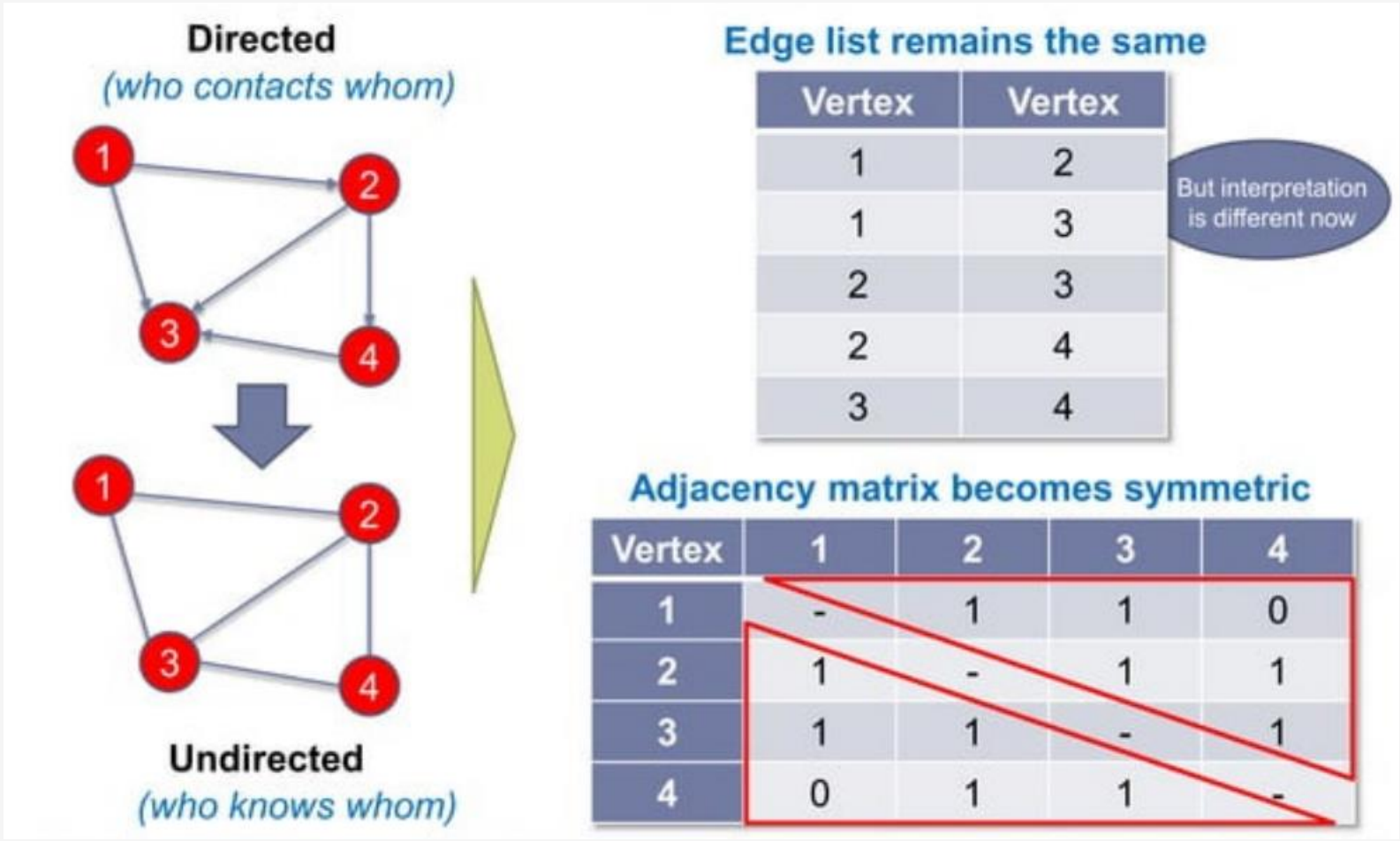
Edge list

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

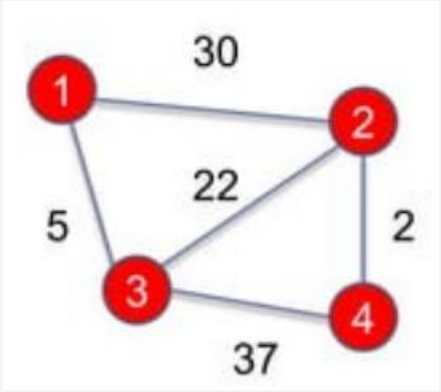
Adjacency matrix

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



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, 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.

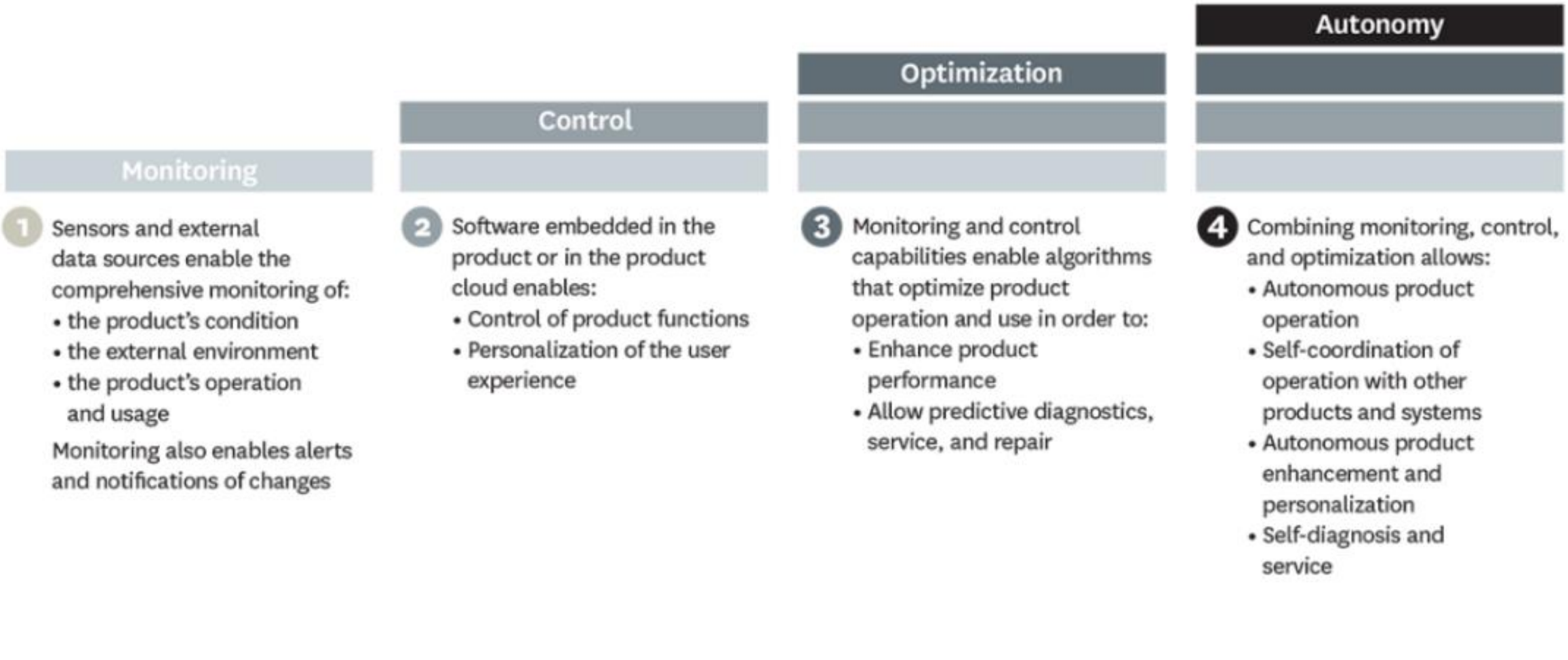
Centrality Measures

- Degree
- Closeness
- Harmonic
- Between

Degree	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
Control
Optimization
Autonomy
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
2
Software embedded in the product or in the product cloud enables:
• Control of product functions
• Personalization of the user experience
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
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

Impact of Big Data

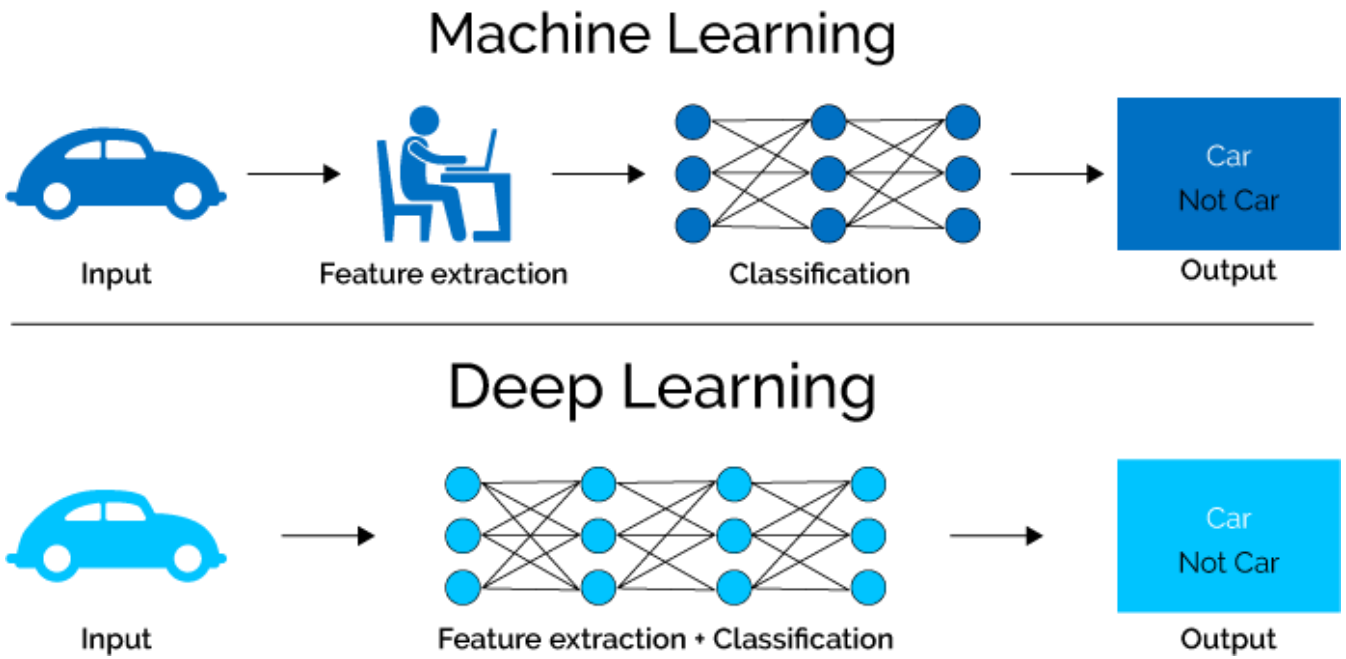
- Personalization marketing
- Better Pricing
- Cost Reduction
- Improved Customer Service

Challenges in big data

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

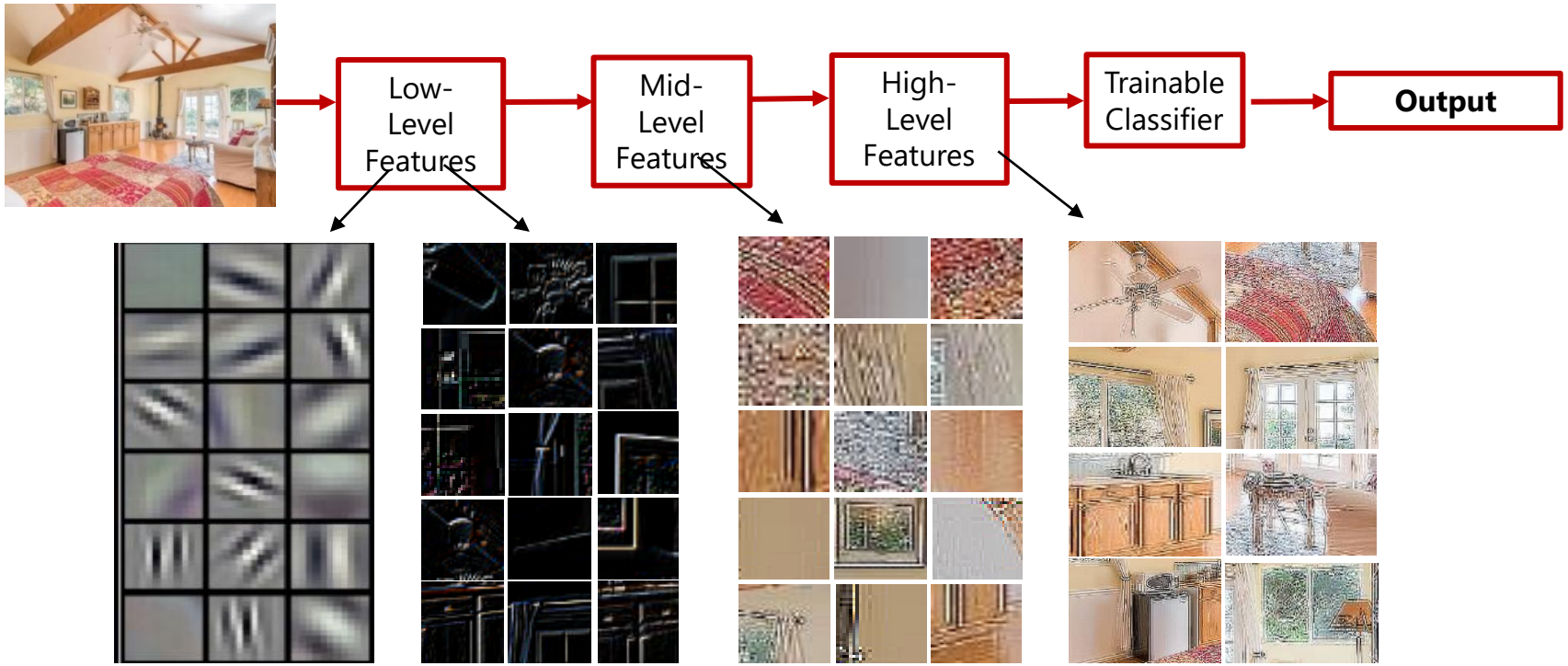
ML vs. Deep Learning

- ***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

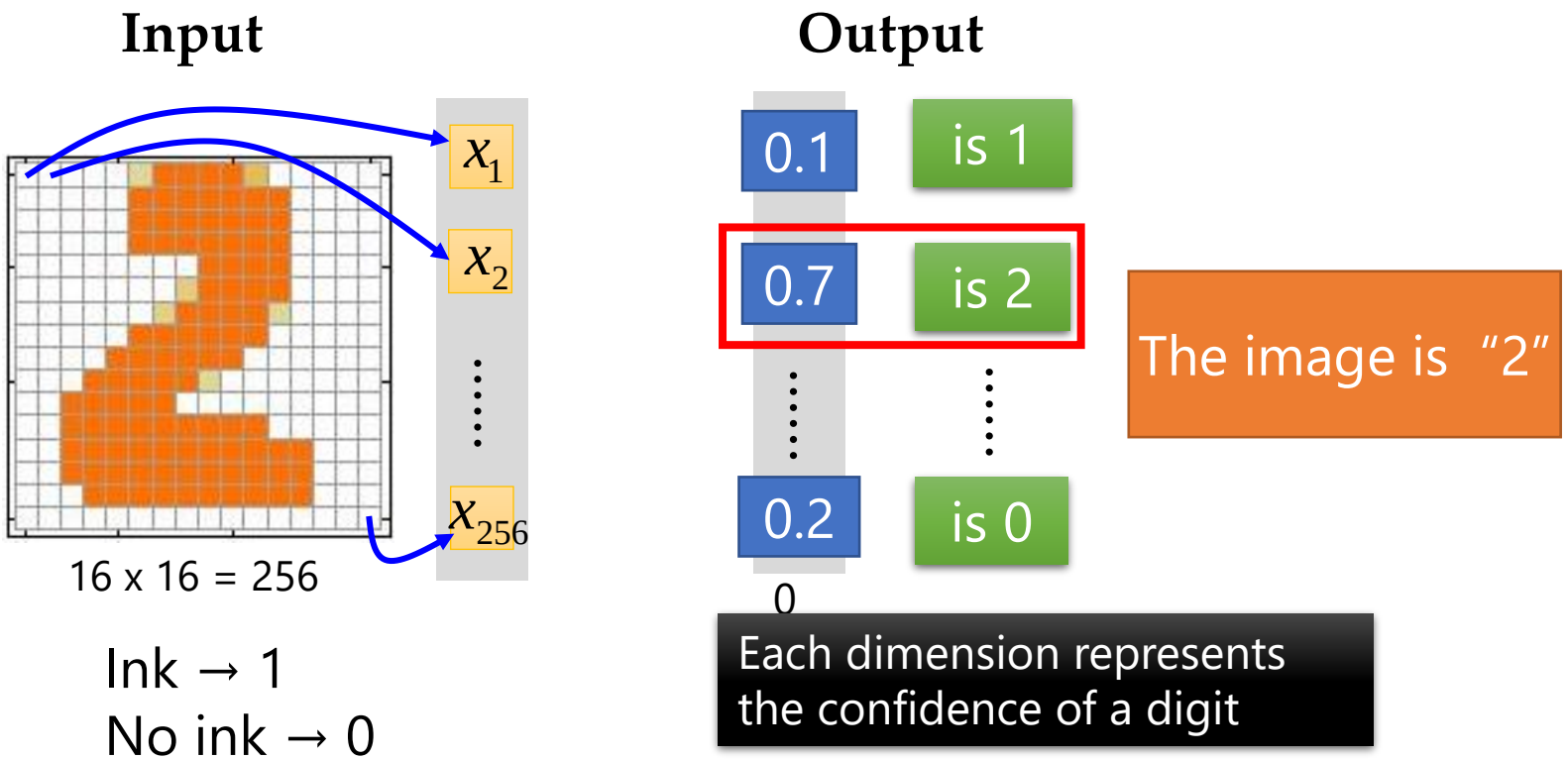


Why is DL Useful?

- 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

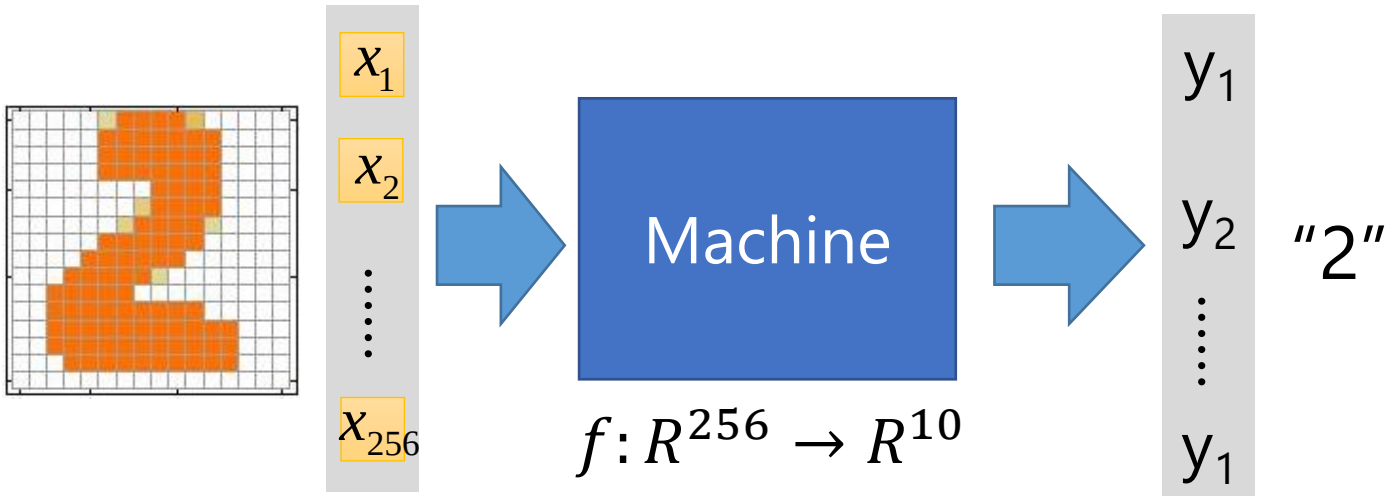
Introduction to Neural Networks

- 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



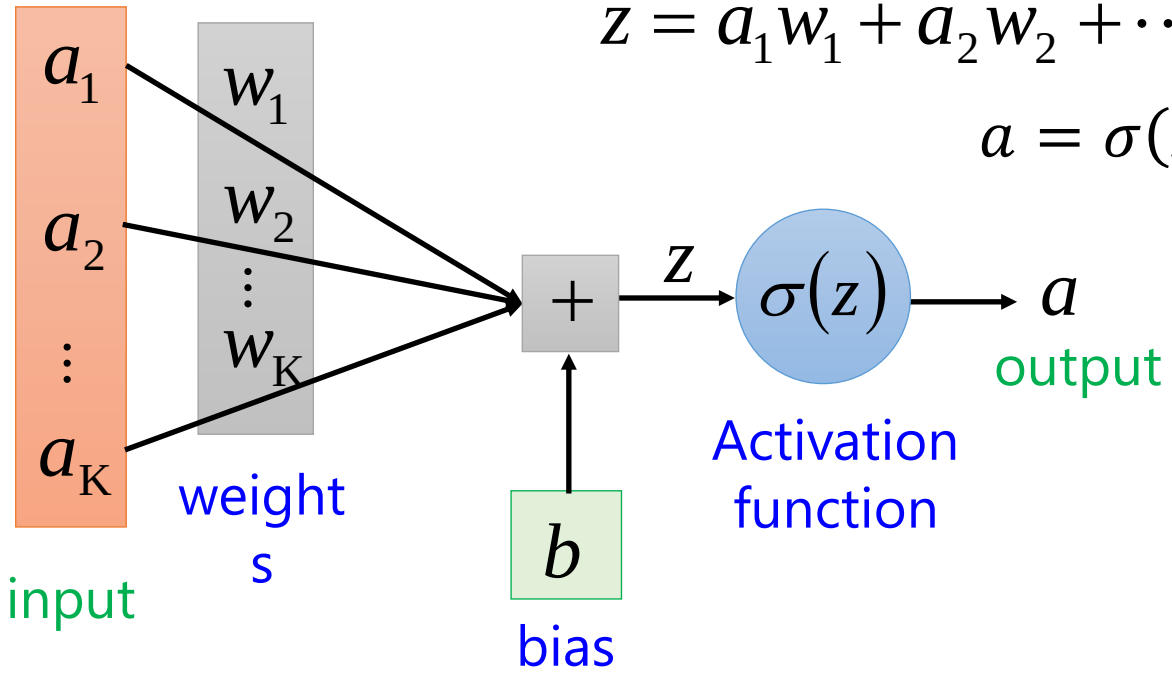
The function f is represented by a neural network

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 $f: R^K \rightarrow R$

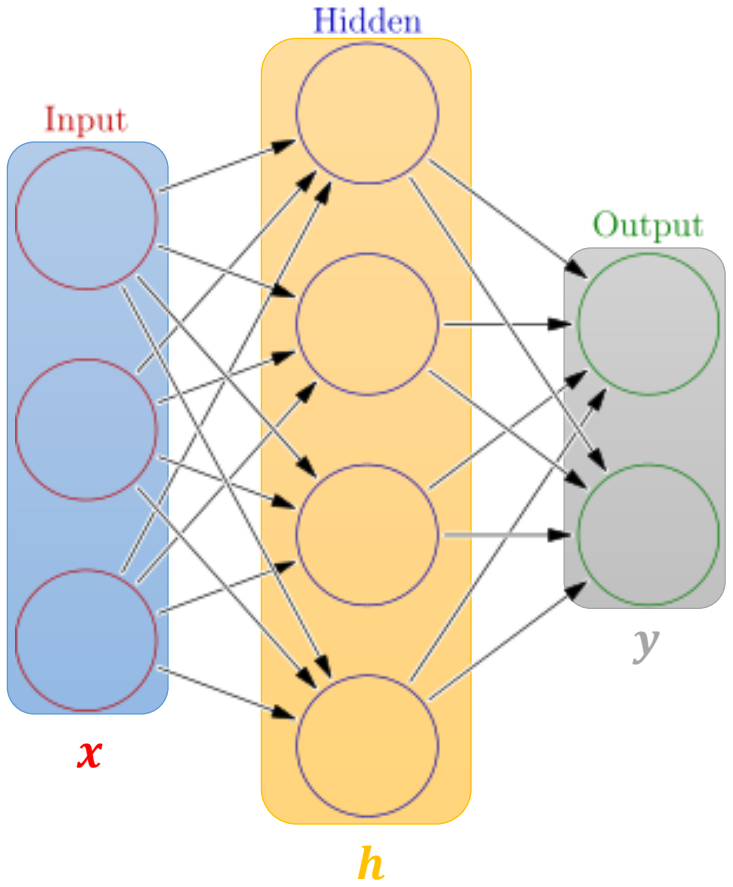
$$z = a_1 w_1 + a_2 w_2 + \cdots + a_K w_K + b$$

$$a = \sigma(z)$$



Elements of Neural Networks

- A NN with one hidden layer and one output layer



Weights Biases

$$hidden\ layer\ h = \sigma(W_1 x + b_1)$$

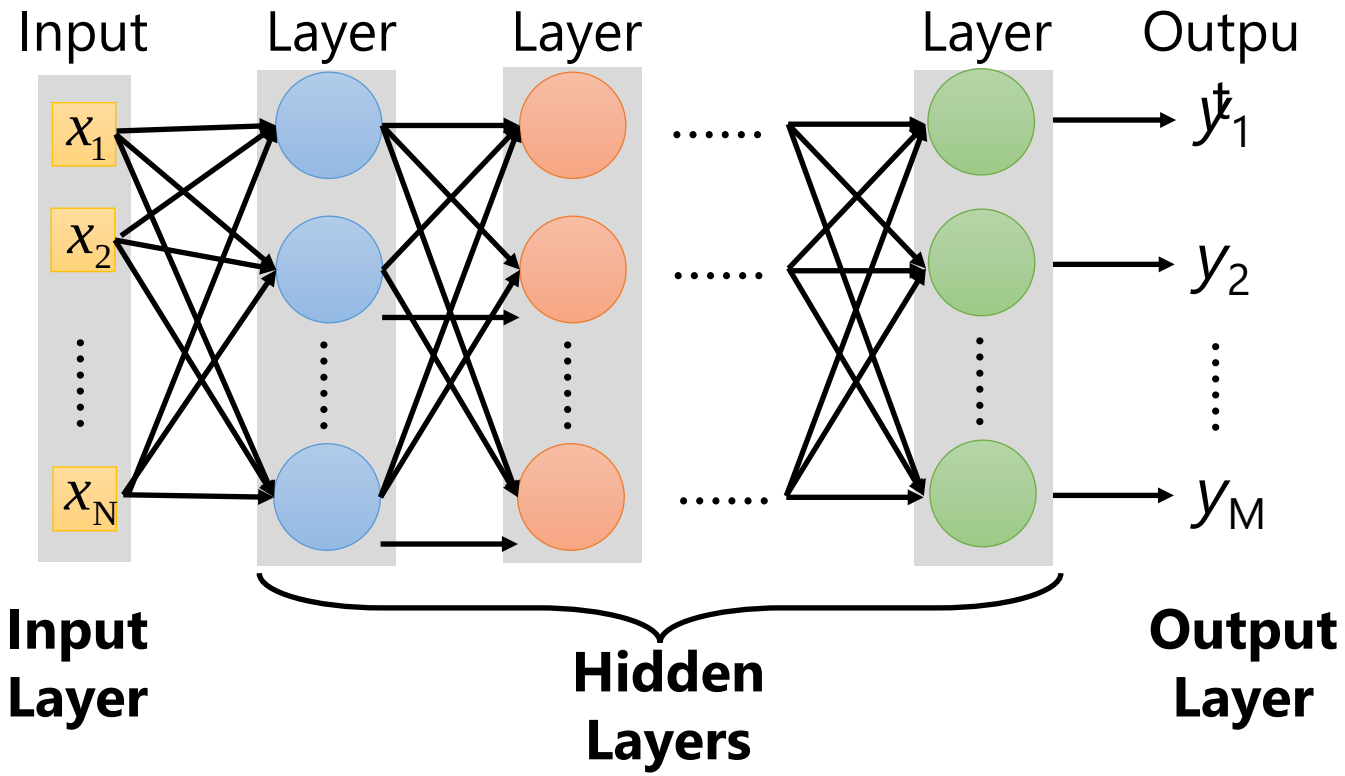
$$output\ layer\ y = \sigma(W_2 h + b_2)$$

Activation functions

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

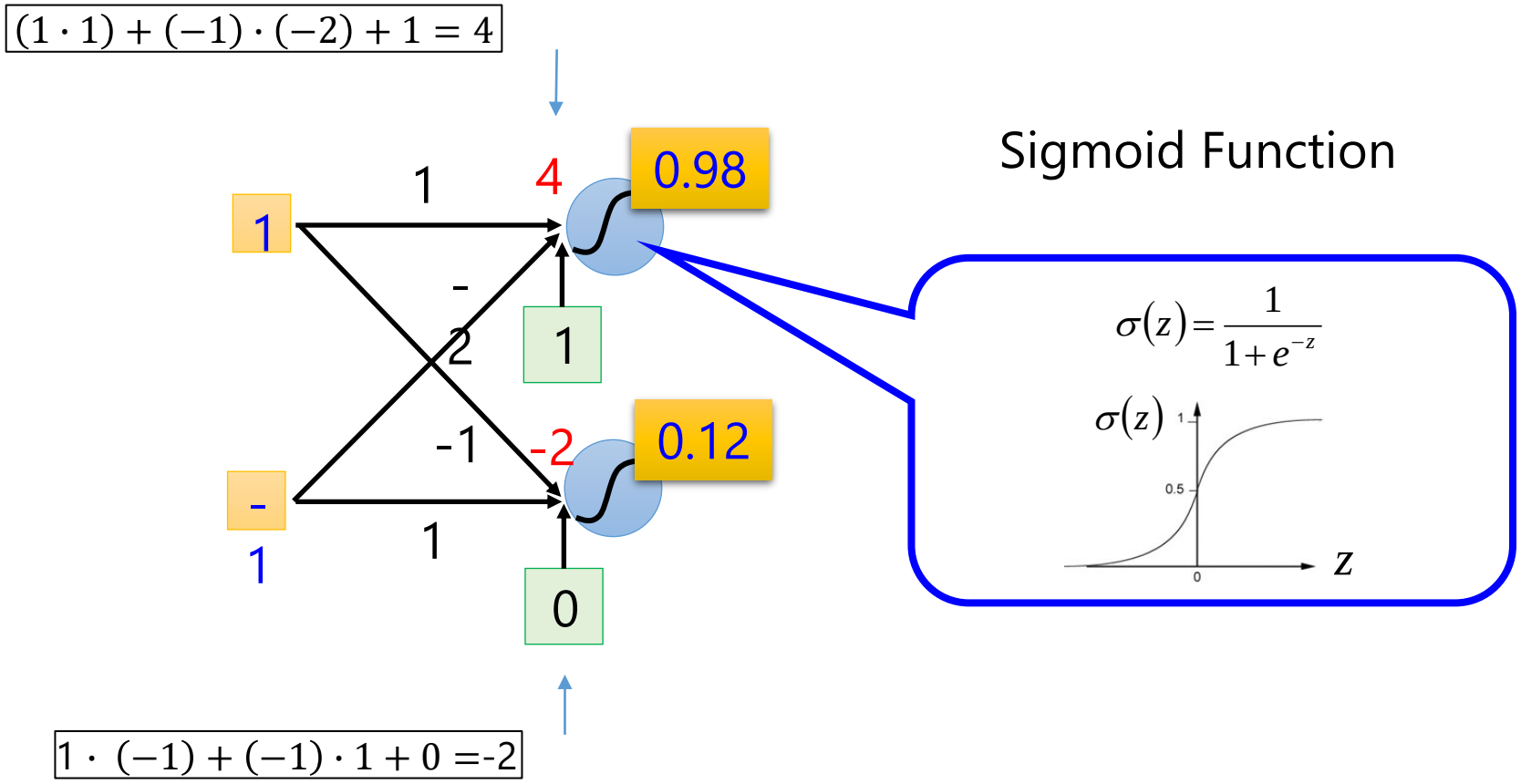
Elements of Neural Networks

- 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



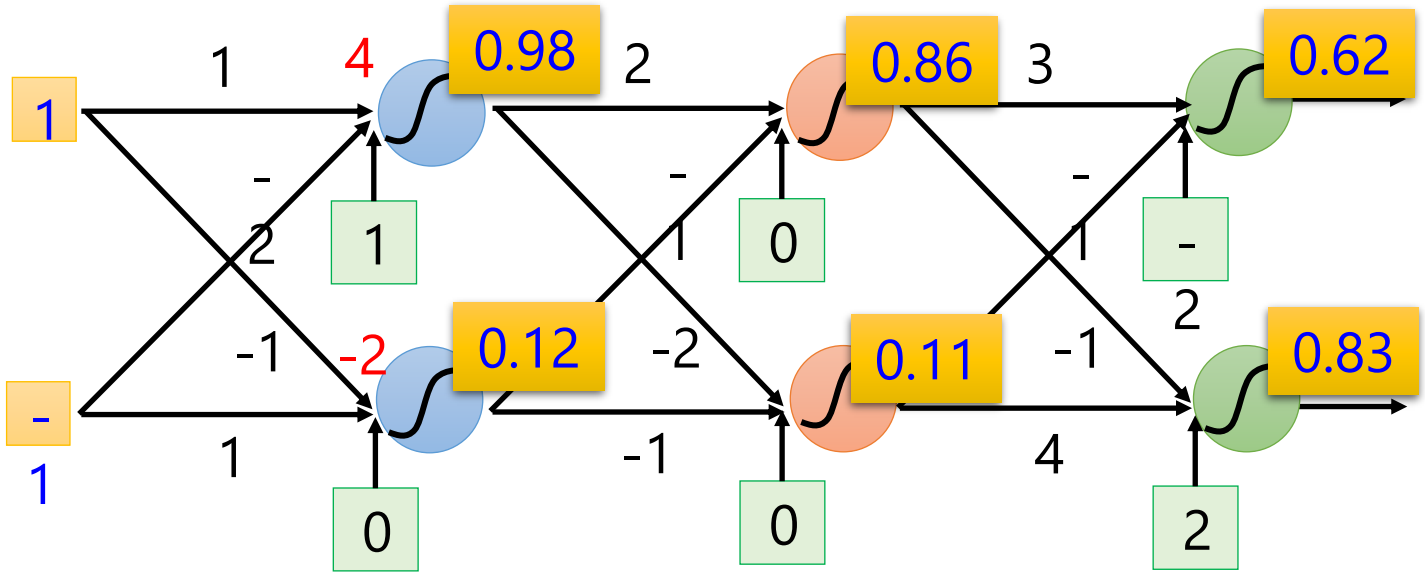
Elements of Neural Networks

- A simple network, toy example



Elements of Neural Networks

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



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

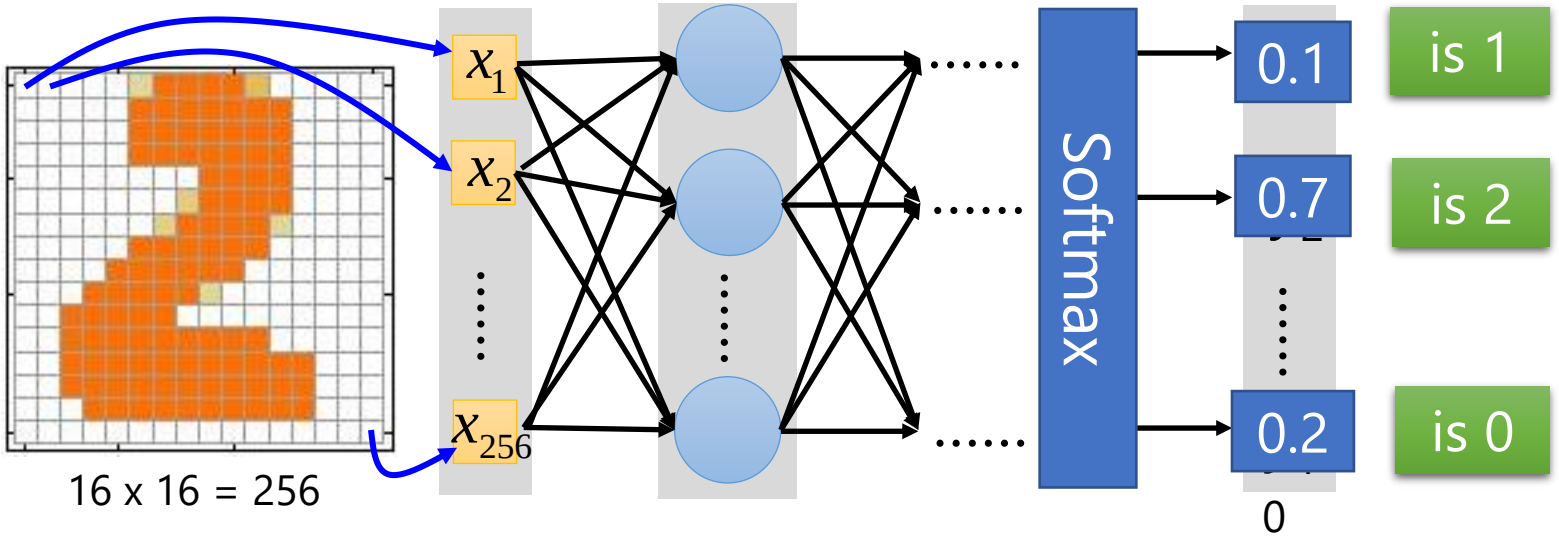
Training NNs

Training Neural Networks

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

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

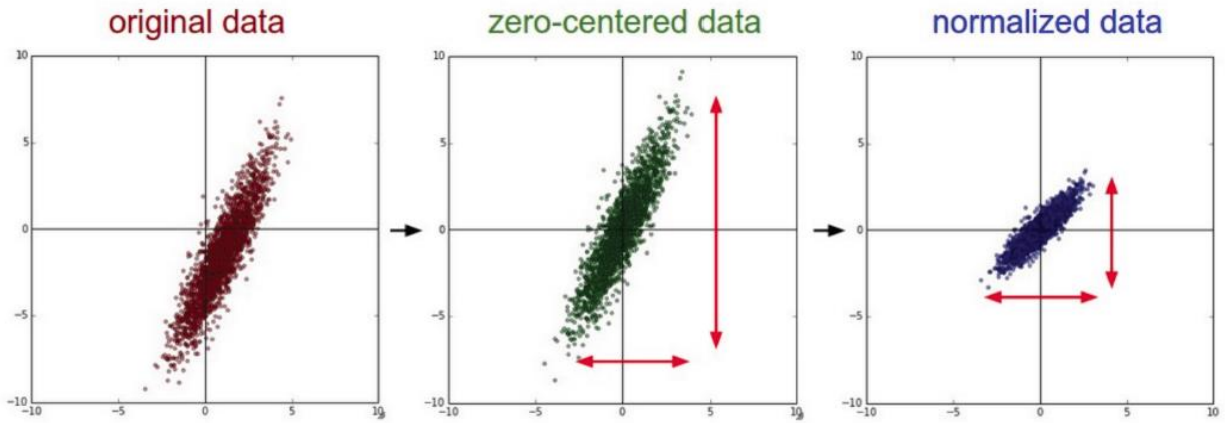
  - Often, the model parameters θ are referred to as weights
- Training a model to learn a set of parameters θ that are optimal (according to a criterion) is one of the greatest challenges in ML



Training NNs

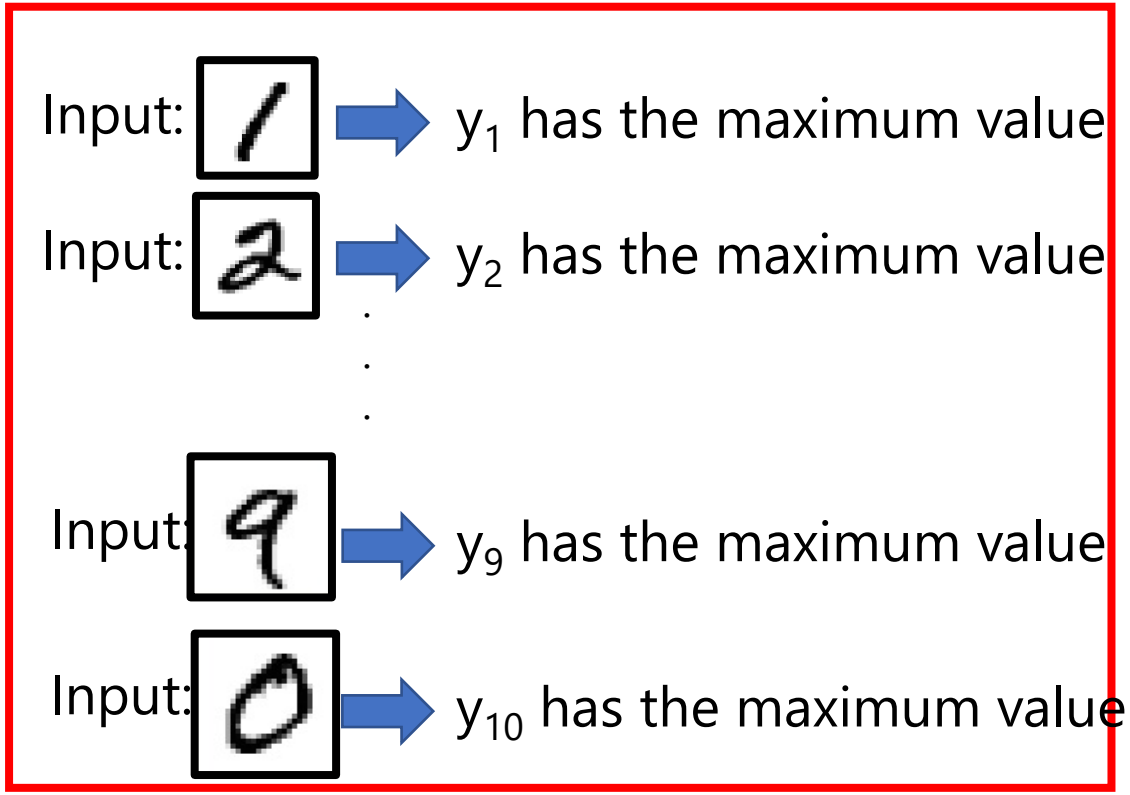
Training Neural Networks

- ***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

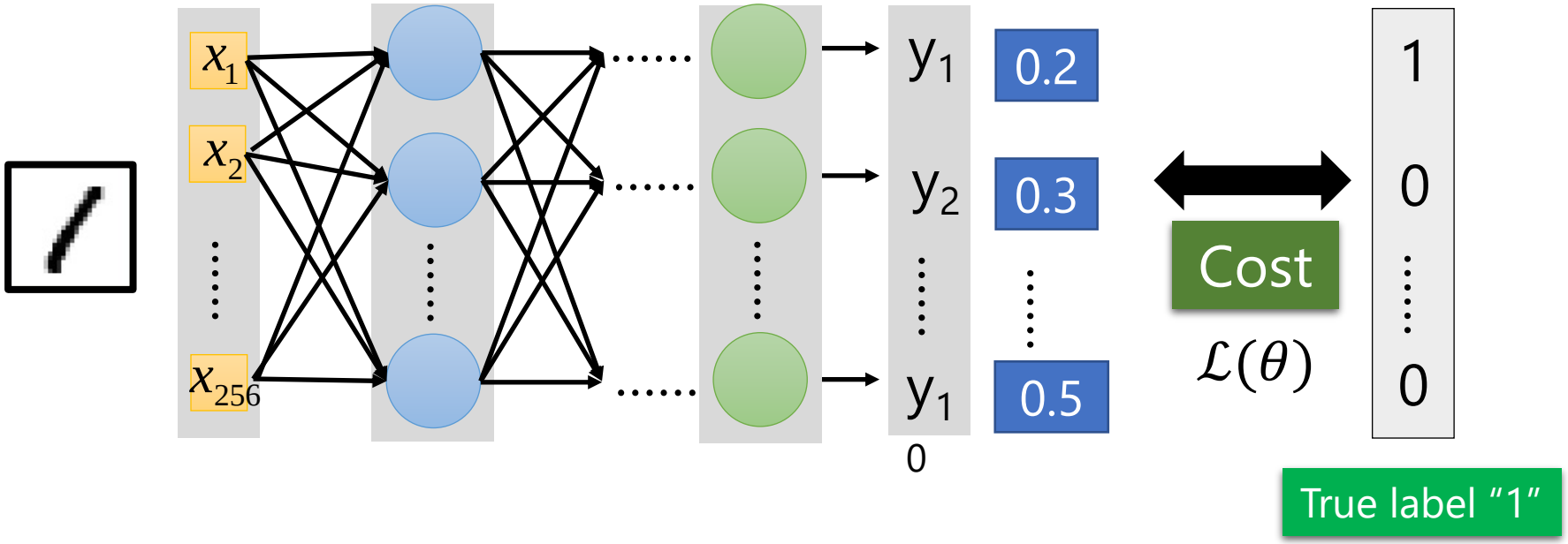
- To train a NN, set the parameters θ such that for a training subset of images, the corresponding elements in the predicted output have maximum values



Training NNs

Training Neural Networks

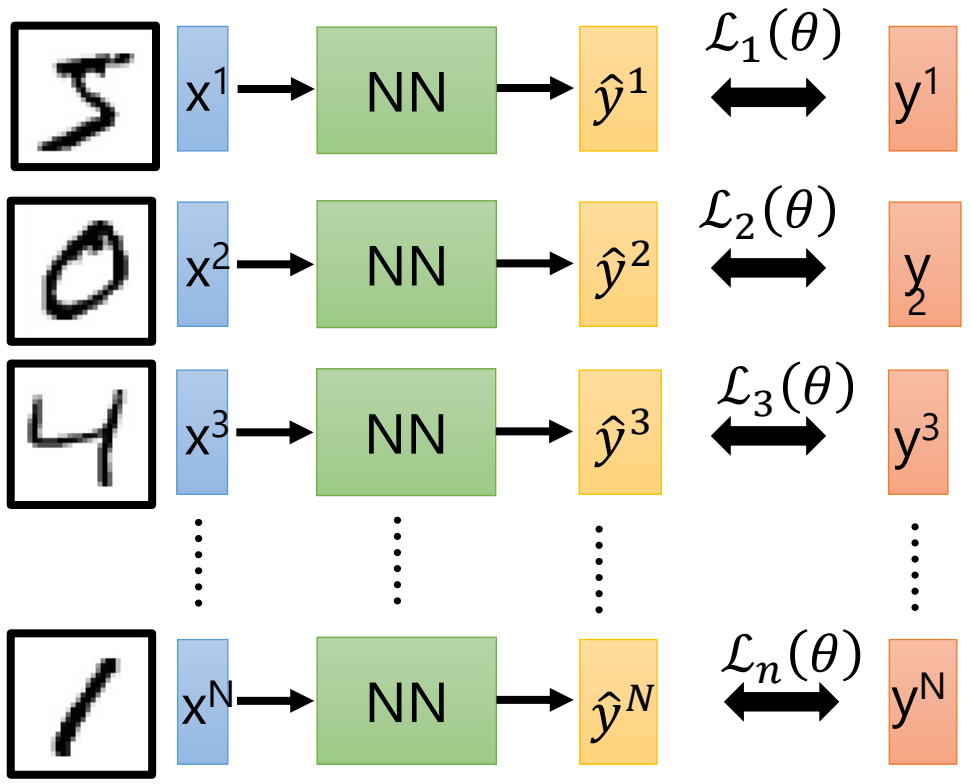
- Define a ***loss function***/objective function/cost function $\mathcal{L}(\theta)$ that calculates the difference (error) between the model prediction and the true label
  - E.g., $\mathcal{L}(\theta)$ can be mean-squared error, cross-entropy, etc.



Training NNs

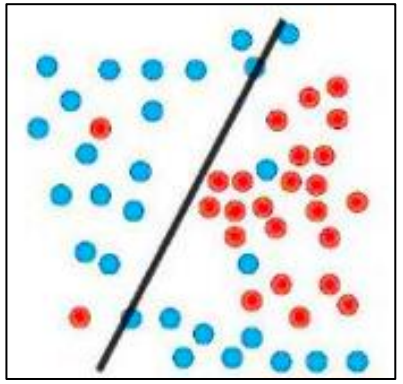
Training Neural Networks

- 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 $\mathcal{L}(\theta)$

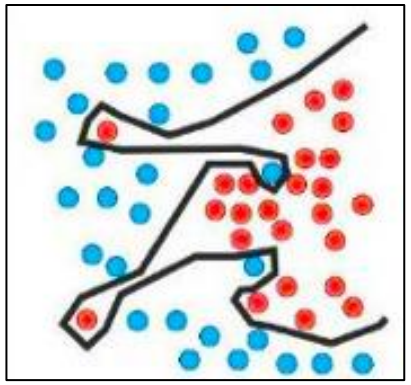


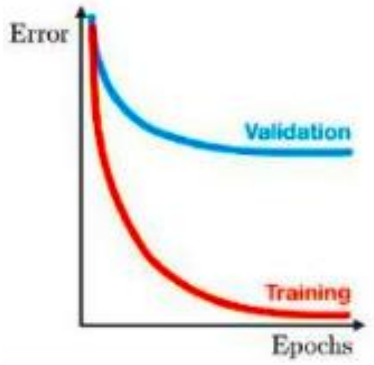
Generalization

- ***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



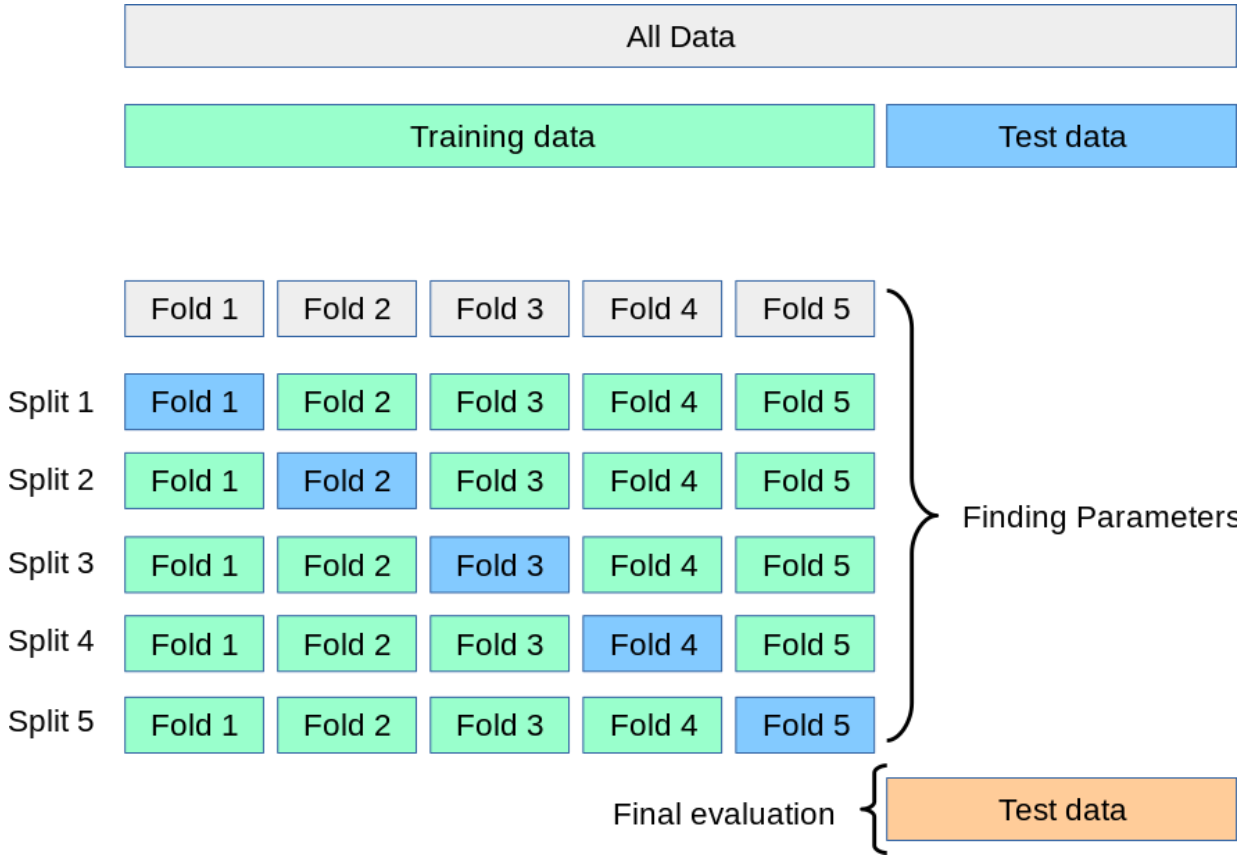
- ***Overfitting***
  - The model is too "complex" and fits irrelevant characteristics (noise) in the data
  - E.g., model with too many parameters
  - Produces low error on the training error and high error on the validation set





k-Fold Cross-Validation

- Illustration of a 5-fold cross-validation

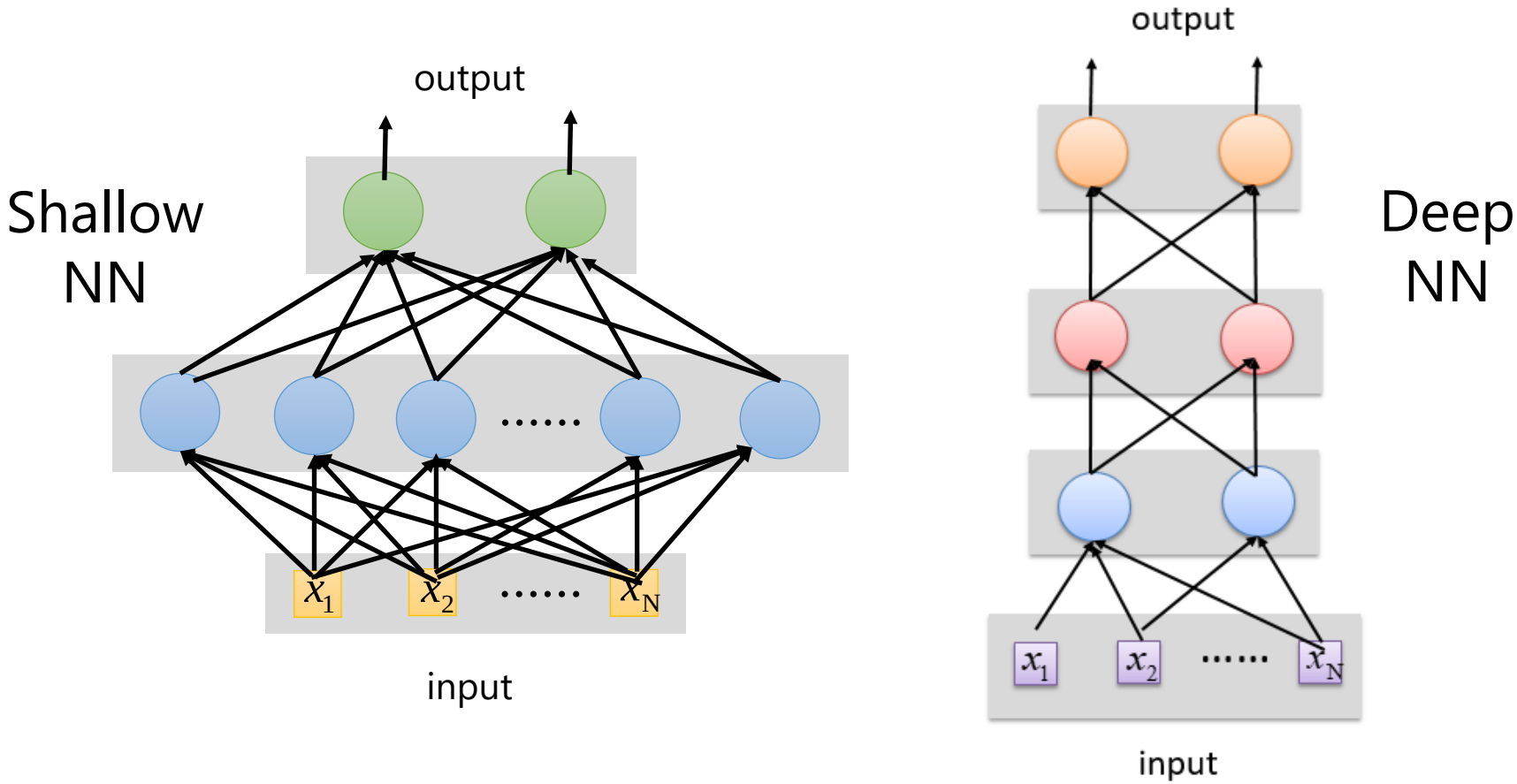


Ensemble Learning

- ***Ensemble learning*** is training multiple classifiers separately and combining their predictions
  - Ensemble learning often outperforms individual classifiers
  - Better results obtained with higher model variety in the ensemble
  - ***Bagging*** (**b**ootstrap **ag**gregating)
    - Randomly draw subsets from the training set (i.e., bootstrap samples)
    - Train separate classifiers on each subset of the training set
    - Perform classification based on the average vote of all classifiers
  - ***Boosting***
    - Train a classifier, and apply weights on the training set (apply higher weights on misclassified examples, focus on "hard examples")
    - Train new classifier, reweight training set according to prediction error
    - Repeat
    - Perform classification based on weighted vote of the classifiers

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

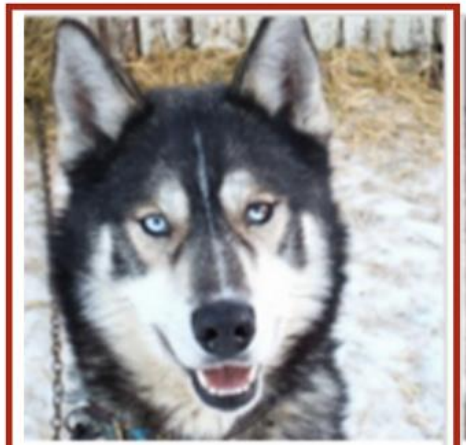
- ***Long Short-Term Memory (LSTM)*** networks are a variant of RNNs
- LSTM mitigates the vanishing/exploding gradient problem
  - Solution: a Memory Cell, updated at each step in the sequence

Predicted: wolf
True: wolf

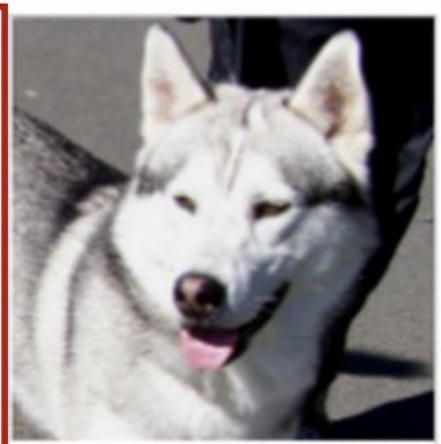
Predicted: husky
True: husky

Predicted: wolf
True: wolf

Incorrect prediction

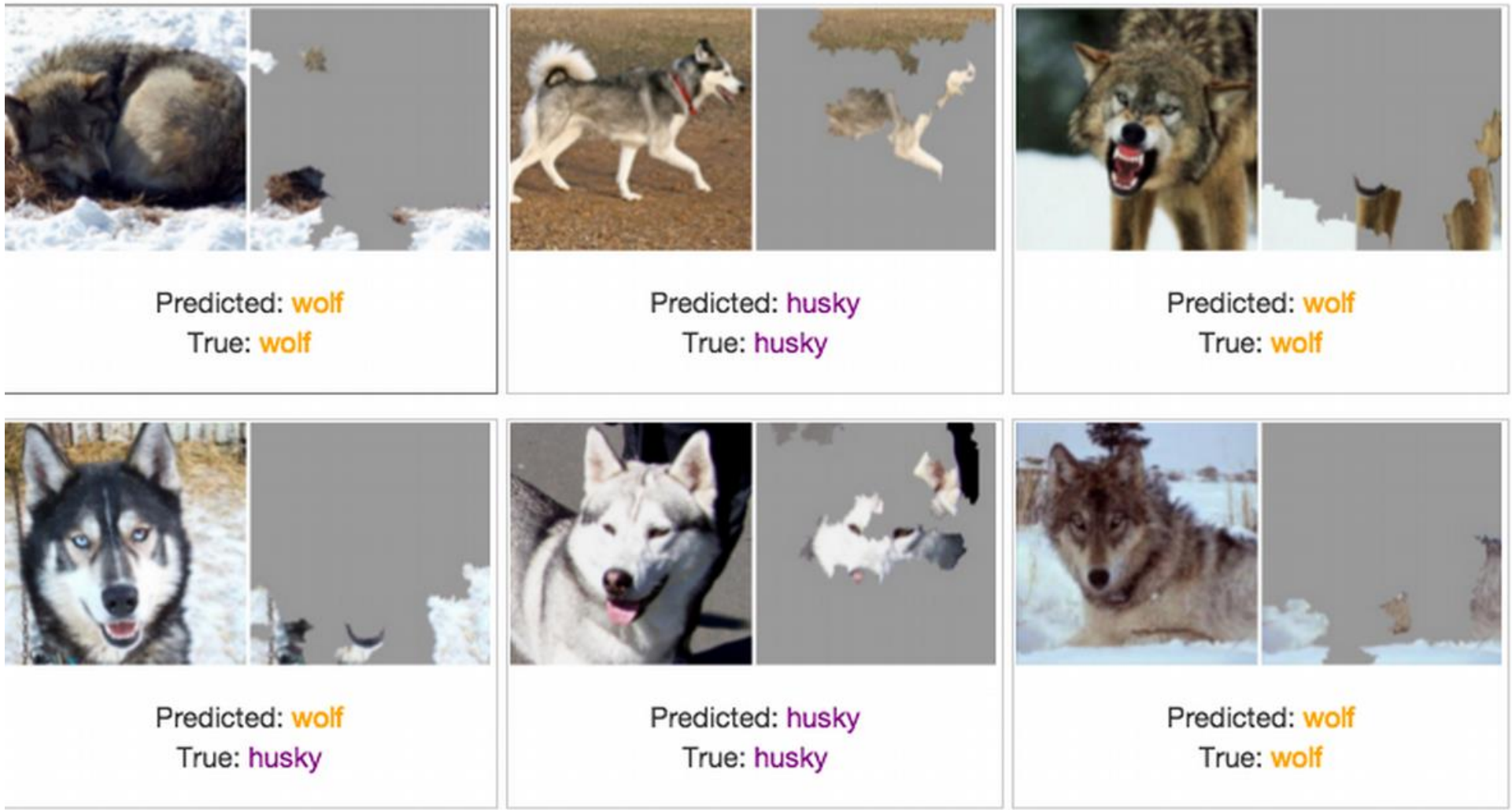


Predicted: wolf
True: husky

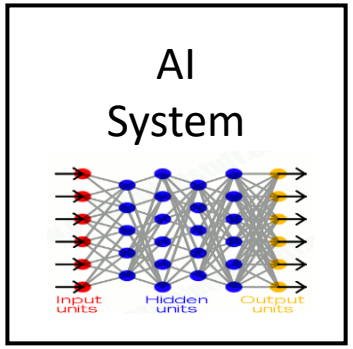


Predicted: husky
True: husky

Predicted: wolf
True: wolf


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

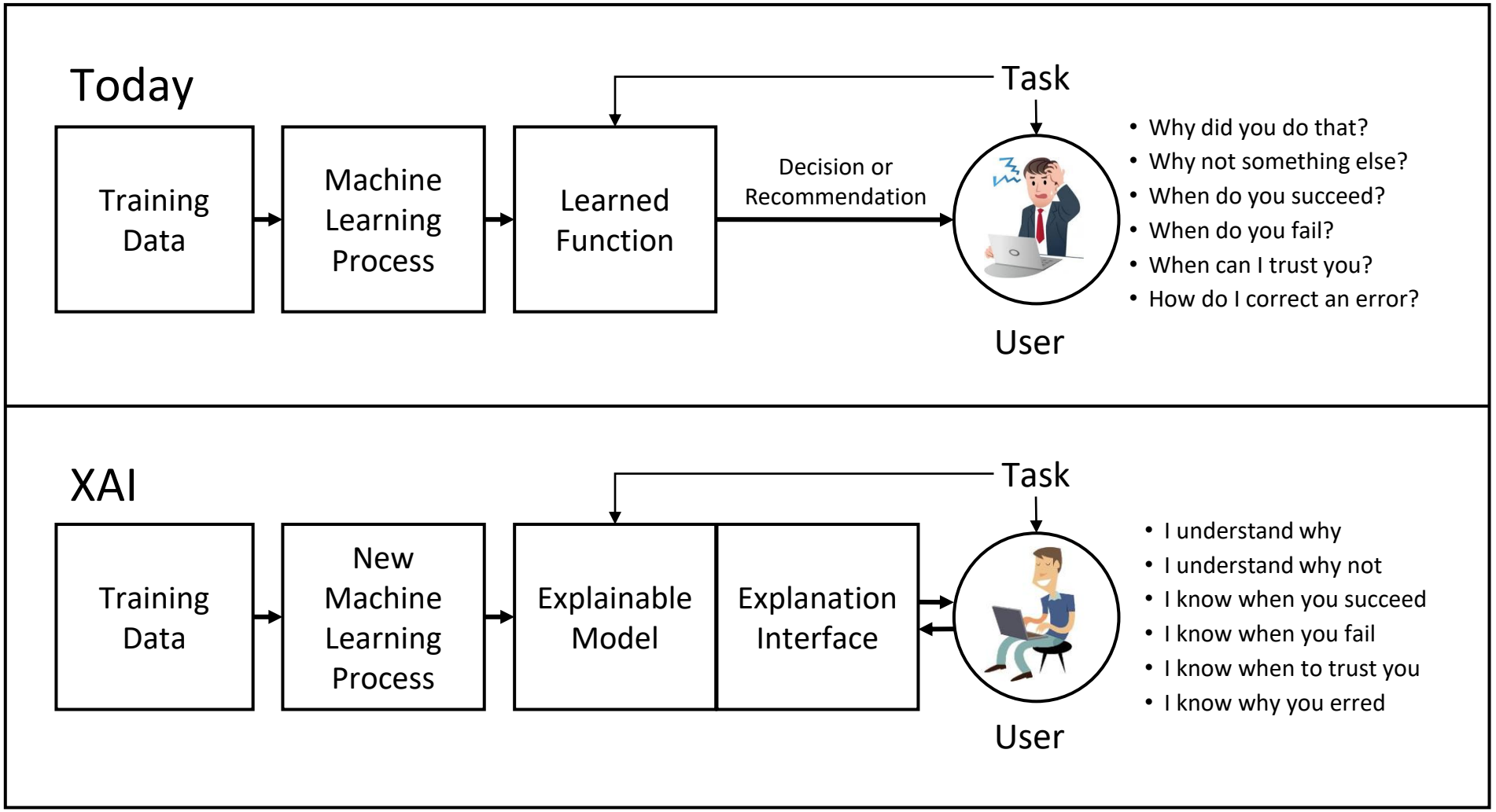


- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

- 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
Decision or Recommendation
Task
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

Type – Capabilities

- Artificial Narrow Intelligence
- Artificial General Intelligence (Strong AI)
- Artificial Super Intelligence