https://ttpoll.com/p/306471

COSC 325: Introduction to Machine Learning

Dr. Hector Santos-Villalobos

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Lecture 20: Introduction to Artificial Neural Networks

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

Class Announcements

Homework

Homework #6 is due Tomorrow Wednesday.

Course Project:

Course Project Presentation Poster

- In-class 12/03
- Template available in Canvas
  - 24” Tall x 36” Wide
- Stitched copy paper is fine.

Lectures:

Next Lecture: Tenure Teaching Evaluation

Quizzes:

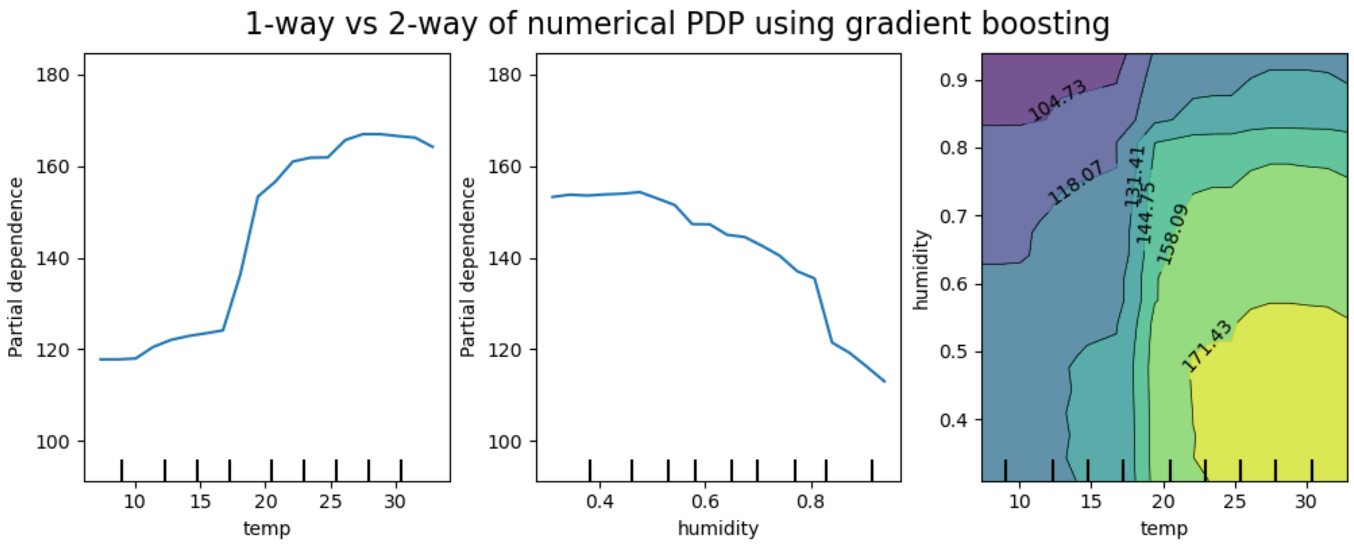
Weekly quiz as usual.

Exams:

The next exam is on Thursday, 11/21—same format.

Review

- Model Explainability
  - Partial Difference Plot (PDP)

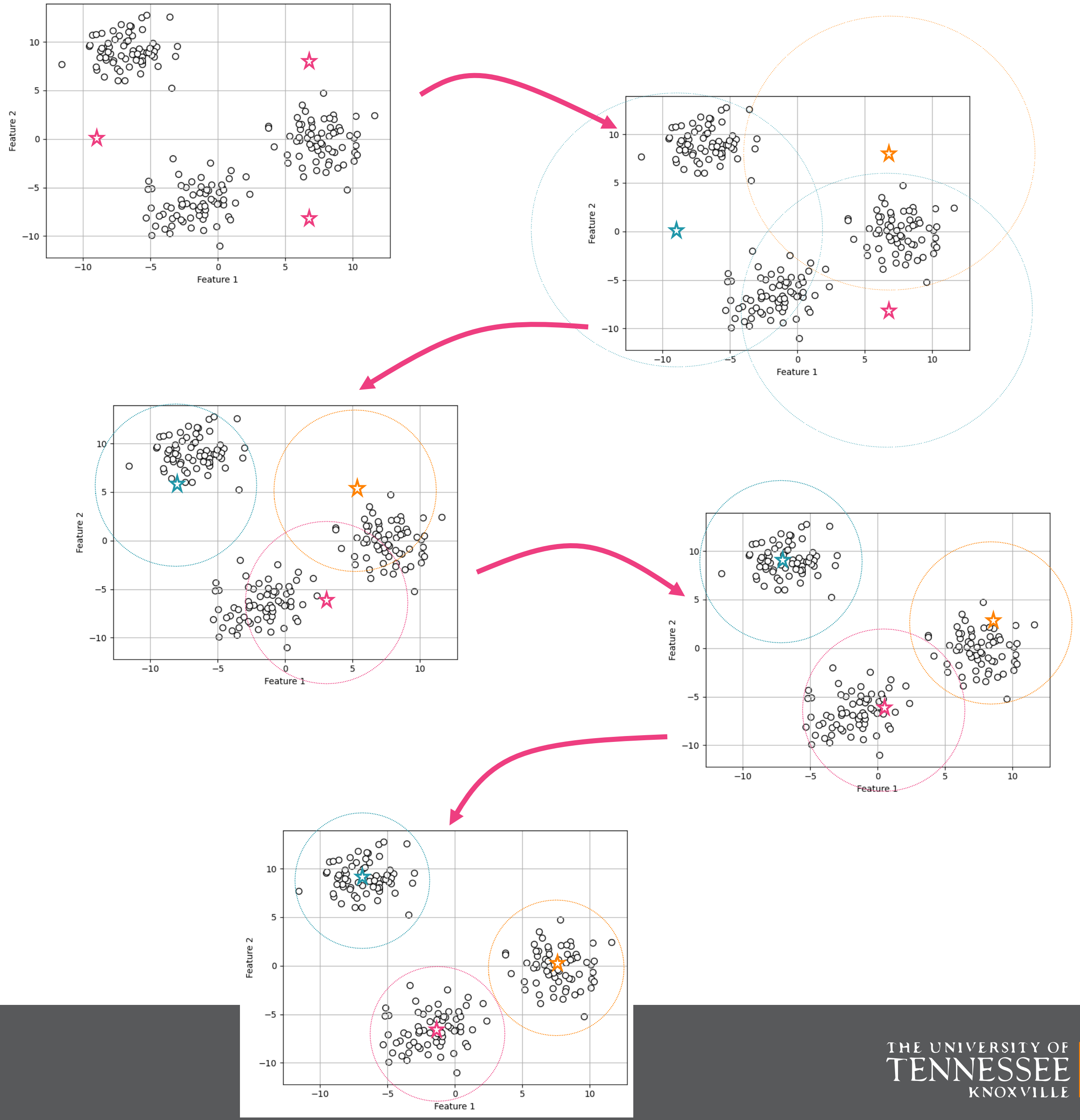


THE UNIVERSITY OF TENNESSEE KNOXVILLE

https://ttpoll.com/p/306471

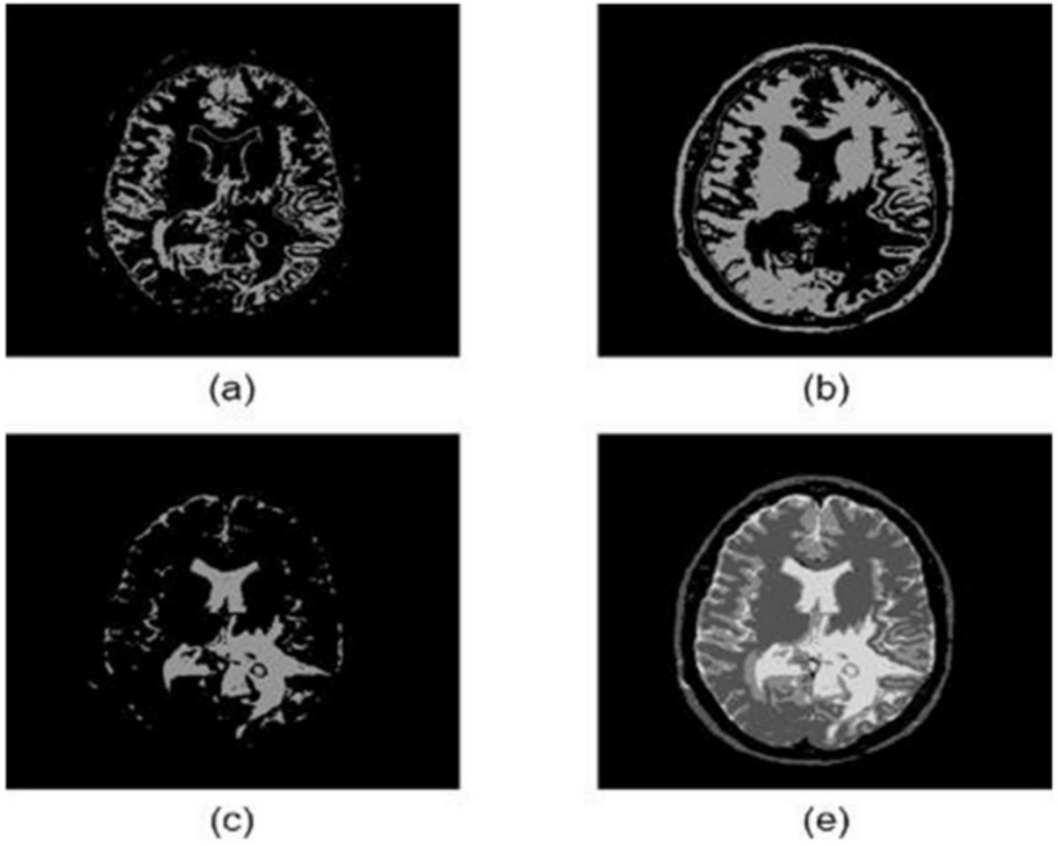
Review

- Model Explainability
  - Partial Difference Plot (PDP)
- Unsupervised Learning – Clustering algorithms
  - k-means
    - K-means++
    - Elbow method
    - Silhouette analysis



Review

- Model Explainability
  - Partial Difference Plot (PDP)
- Unsupervised Learning – Clustering algorithms
  - k-means
    - K-means++
    - Elbow method
    - Silhouette analysis
  - Fuzzy-c-means clustering
    - Assigns samples to a cluster and a probability of cluster membership [0,1] based on the distance of the samples to the cluster's centroid.



FCM Segmentation of (a) Gray Matter, (b) White Matter, (c) Cerebrospinal Fluid, (d) MRI

Dhanachandra and Chanu, "An image segmentation approach based on fuzzy c-means and dynamic particle swarm optimization algorithm," 2020.

Atakishiyev and Reformat, “Analysis of Word Embeddings using Fuzzy Clustering,” 2019.

Review

- Model Explainability
  - Partial Difference Plot (PDP)
- Unsupervised Learning – Clustering algorithms
  - k-means
    - K-means++
    - Elbow method
    - Silhouette analysis
  - Fuzzy-c-means clustering
    - Assigns samples to a cluster and a probability of cluster membership [0,1] based on the distance of the samples to the cluster’s centroid.

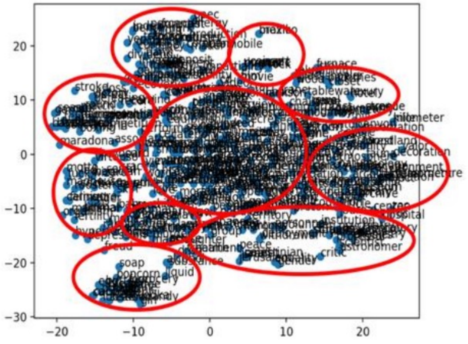


Fig 1. t-SNE Visualization of Word Vectors

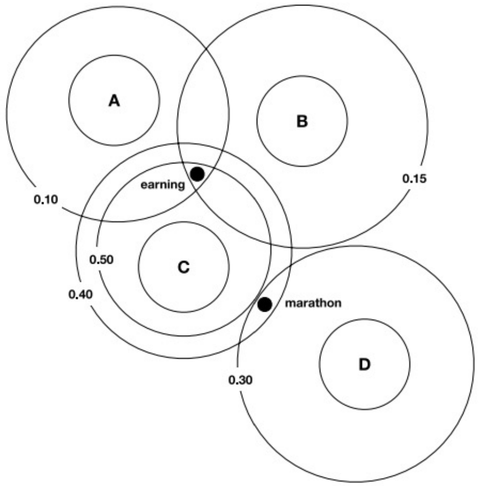


Fig 2. Visualization of words: **earning** and **marathon** that belong to multiple clusters

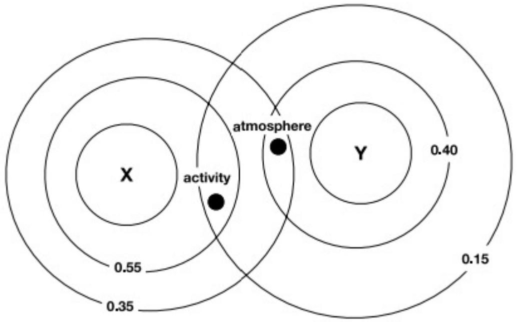


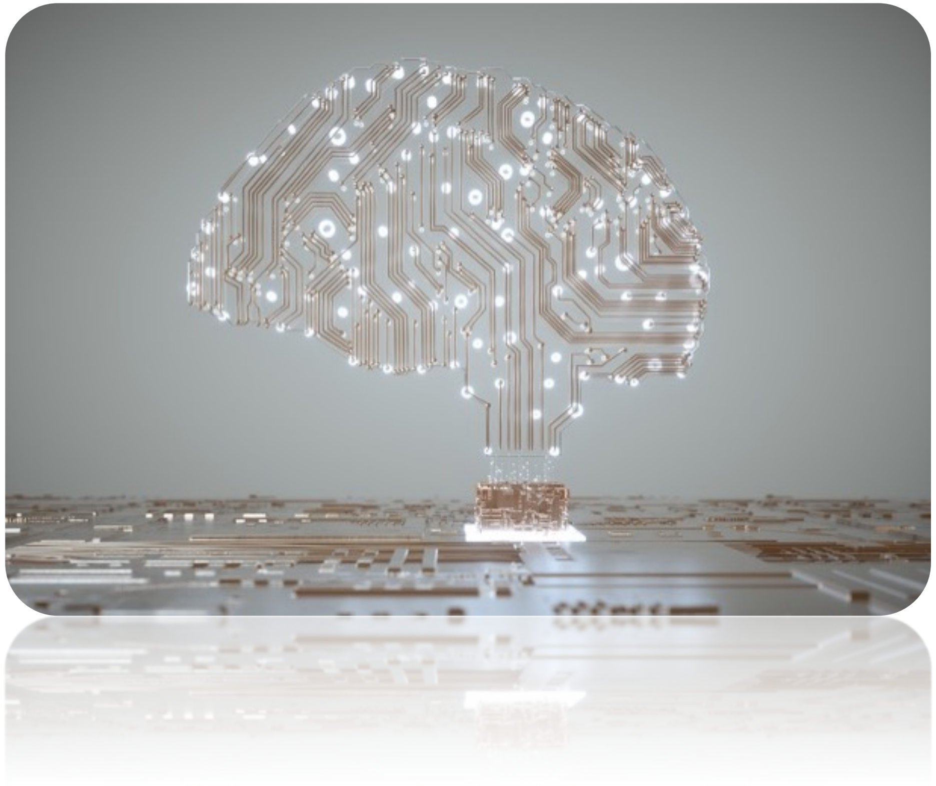
Fig 3. Visualization of words: **activity** and **atmosphere** that belong to two clusters

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Today's Topics

Wrap-up Unsupervised Learning

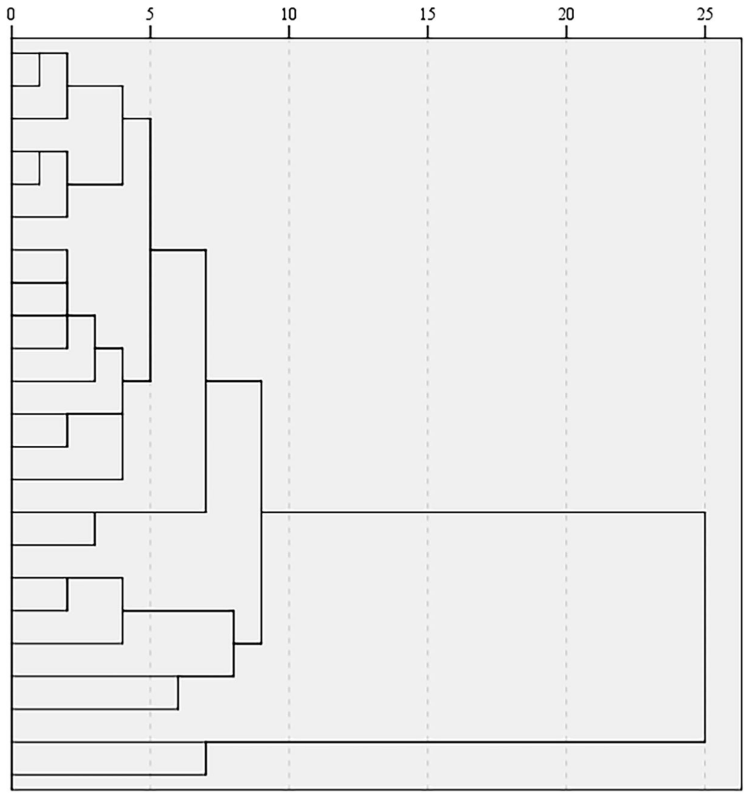
Neural Networks



Hierarchical Clustering

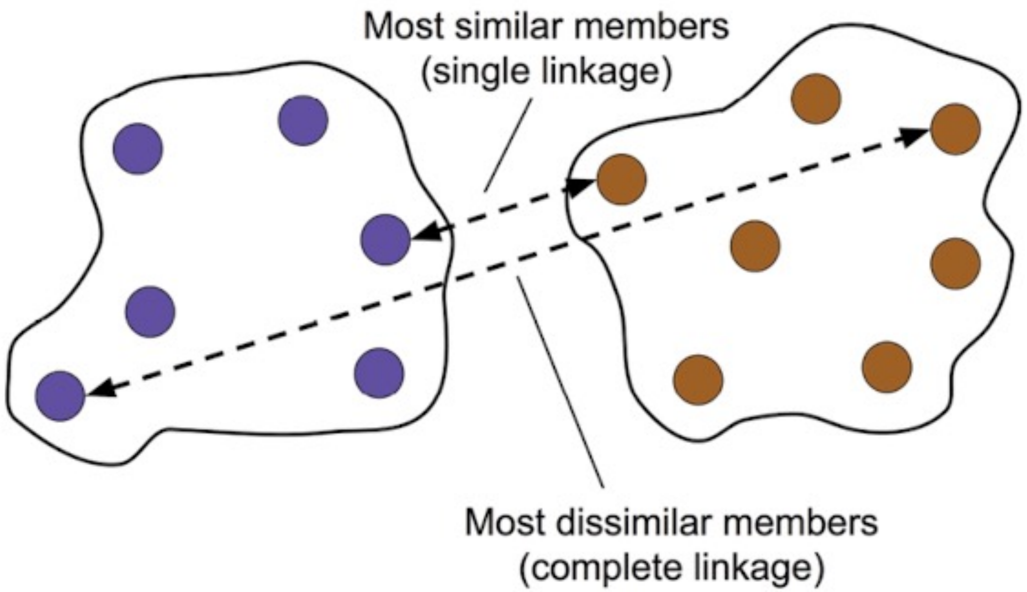
- Bottom-up approach
- Allows plot of dendrograms (Binary hierarchical clustering)
- No need to specify the number of clusters k
- Two main approaches
  - Divisive method:
    - The whole dataset belongs to a single cluster
    - Split cluster
    - Stop when each sample is a cluster
  - **Agglomerative method:**
    - Starts with each sample as a cluster
    - Merge clusters until all samples belong to a single cluster

Dendrogram example



Agglomerative Method

- Single linkage
  - Compares the most similar samples
- **Complete linkage**
  - Compares the most different samples



Raschka, et. al., "Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python"

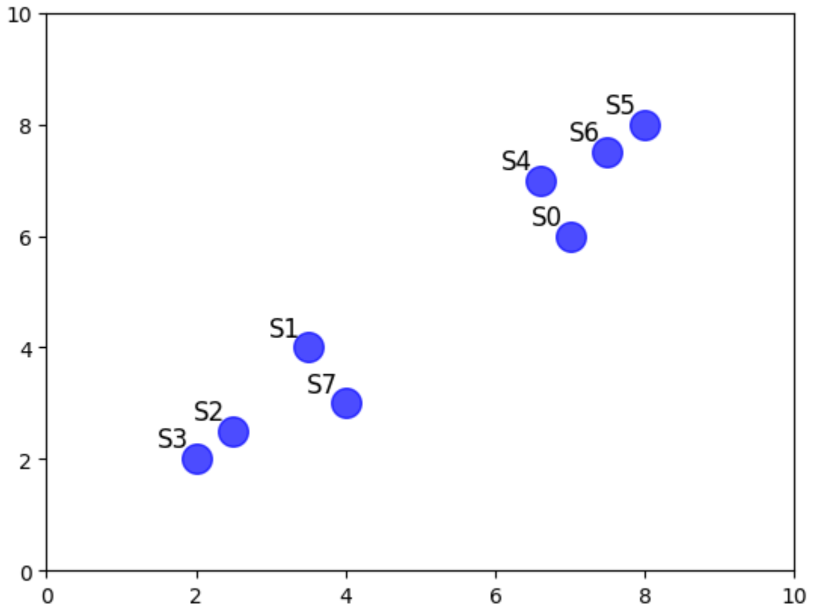
Agglomerative Complete Linkage Algorithm

1. Compute a pair-wise distance matrix of all samples.
2. Represent each data point as a singleton cluster.
3. Merge the two closest clusters based on the distance between the ***most dissimilar (distant) members***.
4. Update the cluster linkage matrix.
5. Repeat steps 2-4 until one single cluster remains.

Raschka, et. al., “Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python”

Step-by-Step Example

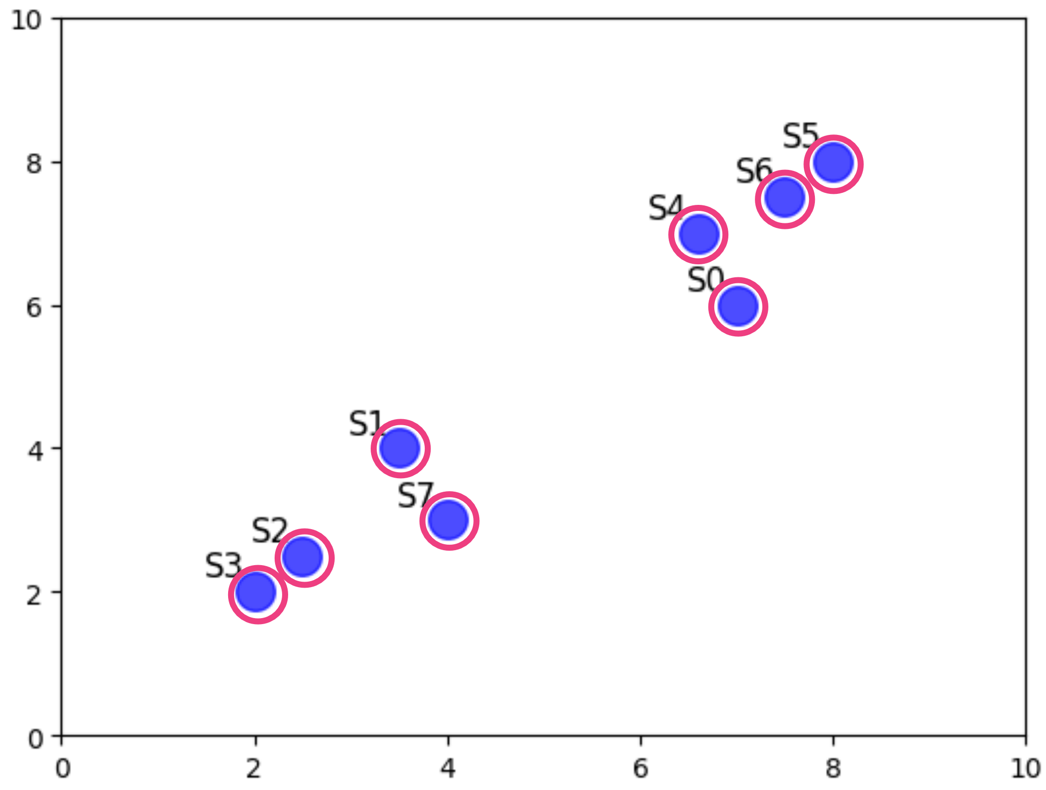
1. **Compute a pair-wise distance matrix of all samples.**
2. Represent each data point as a singleton cluster.
3. Merge the two closest clusters based on the distance between the most dissimilar (distant) members.
4. Update the cluster linkage matrix.
5. Repeat steps 2-4 until one single cluster remains.



	S0	S1	S2	S3	S4	S5	S6	S7
S0	0.0	4.0	5.7	6.4	1.1	2.2	1.6	4.2
S1	4.0	0.0	1.8	2.5	4.3	6.0	5.3	1.1
S2	5.7	1.8	0.0	0.7	6.1	7.8	7.1	1.6
S3	6.4	2.5	0.7	0.0	6.8	8.5	7.8	2.2
S4	1.1	4.3	6.1	6.8	0.0	1.7	1.0	4.8
S5	2.2	6.0	7.8	8.5	1.7	0.0	0.7	6.4
S6	1.6	5.3	7.1	7.8	1.0	0.7	0.0	5.7
S7	4.2	1.1	1.6	2.2	4.8	6.4	5.7	0.0

Step-by-Step Example

1. Compute a pair-wise distance matrix of all samples.
2. **Represent each data point as a singleton cluster.**
3. Merge the two closest clusters based on the distance between the most dissimilar (distant) members.
4. Update the cluster linkage matrix.
5. Repeat steps 2-4 until one single cluster remains.

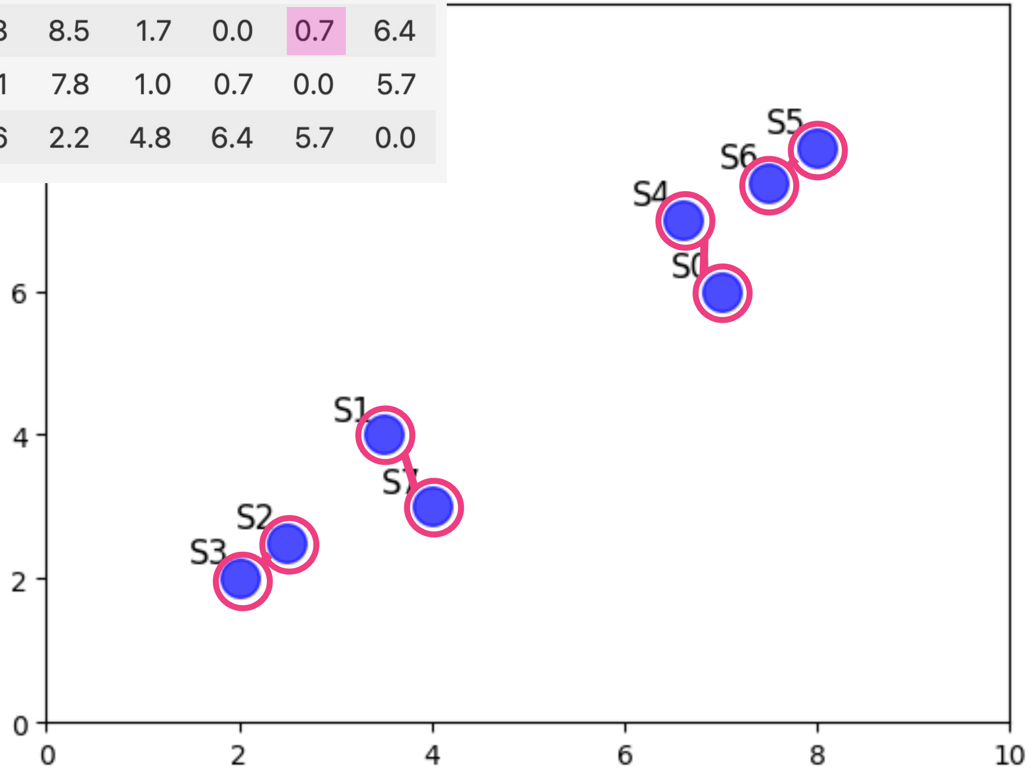


https://ttpoll.com/p/306471

Step-by-Step Exa

	S0	S1	S2	S3	S4	S5	S6	S7
S0	0.0	4.0	5.7	6.4	1.1	2.2	1.6	4.2
S1	4.0	0.0	1.8	2.5	4.3	6.0	5.3	1.1
S2	5.7	1.8	0.0	0.7	6.1	7.8	7.1	1.6
S3	6.4	2.5	0.7	0.0	6.8	8.5	7.8	2.2
S4	1.1	4.3	6.1	6.8	0.0	1.7	1.0	4.8
S5	2.2	6.0	7.8	8.5	1.7	0.0	0.7	6.4
S6	1.6	5.3	7.1	7.8	1.0	0.7	0.0	5.7
S7	4.2	1.1	1.6	2.2	4.8	6.4	5.7	0.0

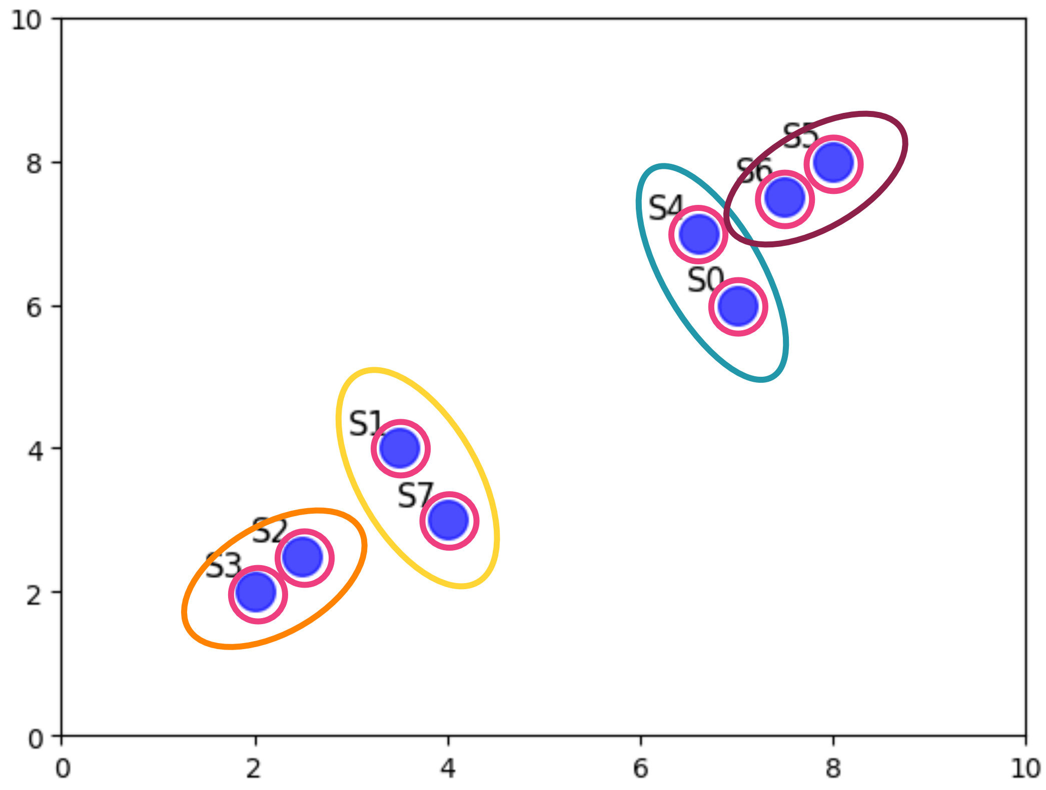
1. Compute a pair-wise distance matrix of all samples.
2. Represent each data point as a singleton cluster.
3. **Merge the two closest clusters based on the distance between the most dissimilar (distant) members.**
4. Update the cluster linkage matrix.
5. Repeat steps 2-4 until one single cluster remains.



THE UNIVERSITY OF TENNESSEE KNOXVILLE

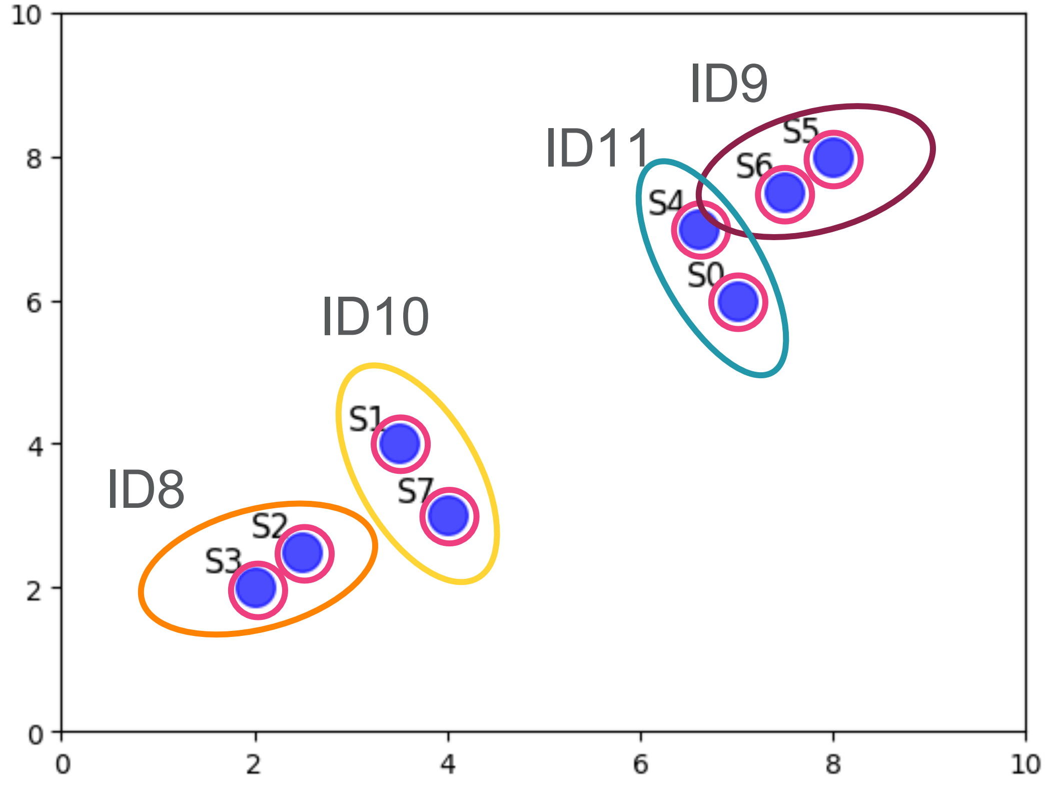
Step-by-Step Example

1. Compute a pair-wise distance matrix of all samples.
2. Represent each data point as a singleton cluster.
3. **Merge the two closest clusters based on the distance between the most dissimilar (distant) members.**
4. Update the cluster linkage matrix.
5. Repeat steps 2-4 until one single cluster remains.



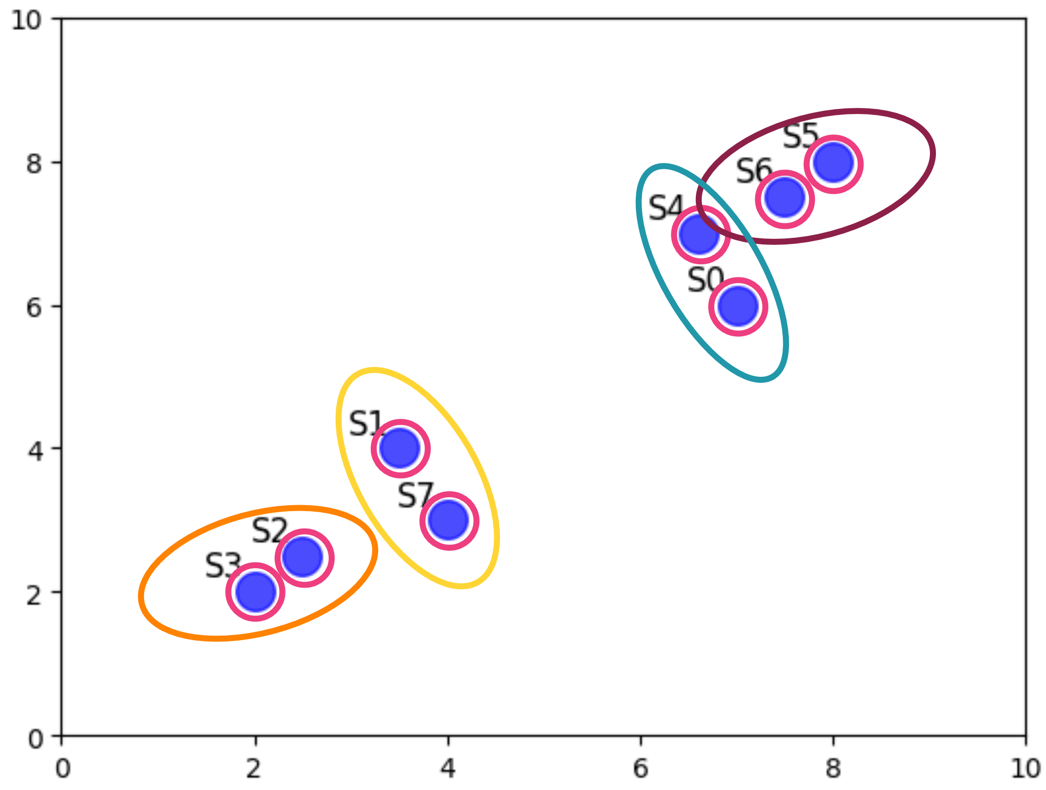
Linkage Matrix Update

Cluster	Label 1	Label 2	Distance	# of samples
Cluster 8	2	3	0.7	2
Cluster 9	5	6	0.7	2
Cluster 10	1	7	1.1	2
Cluster 11	0	4	1.1	2



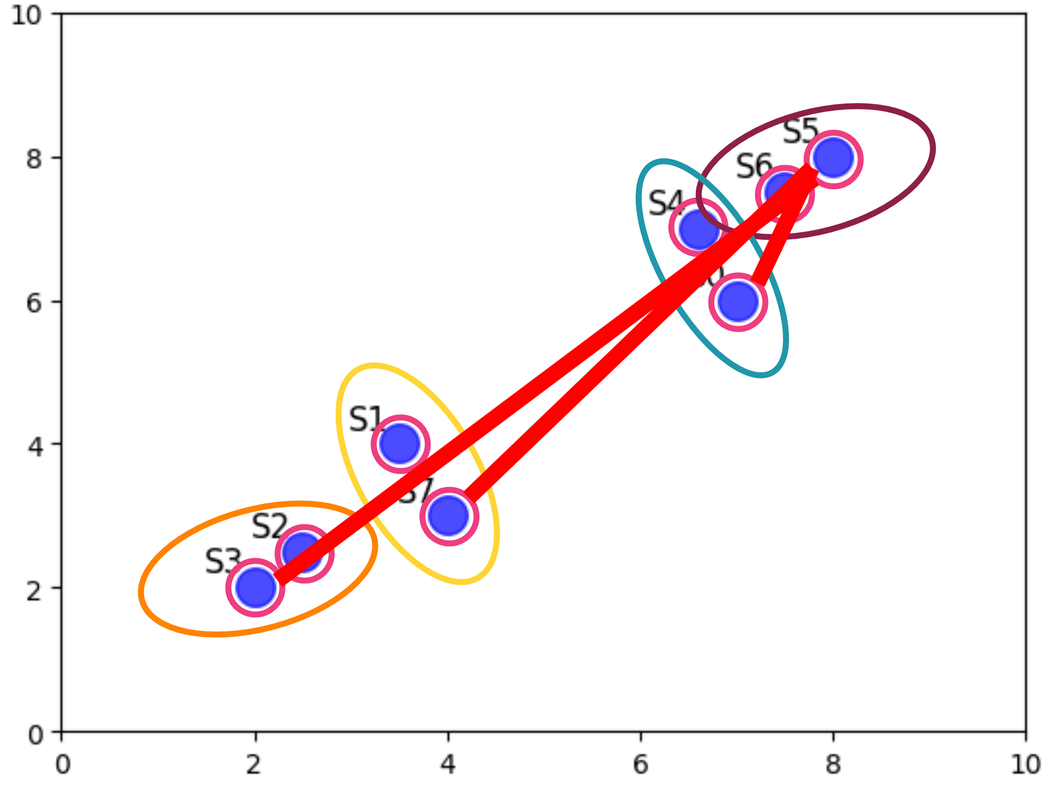
Step-by-Step Example

1. Compute a pair-wise distance matrix of all samples.
2. Represent each data point as a singleton cluster.
3. Merge the two closest clusters based on the **distance between the most dissimilar (distant) members.**
4. Update the cluster linkage matrix.
5. **Repeat steps 2-4 until one single cluster remains.**



Step-by-Step Example

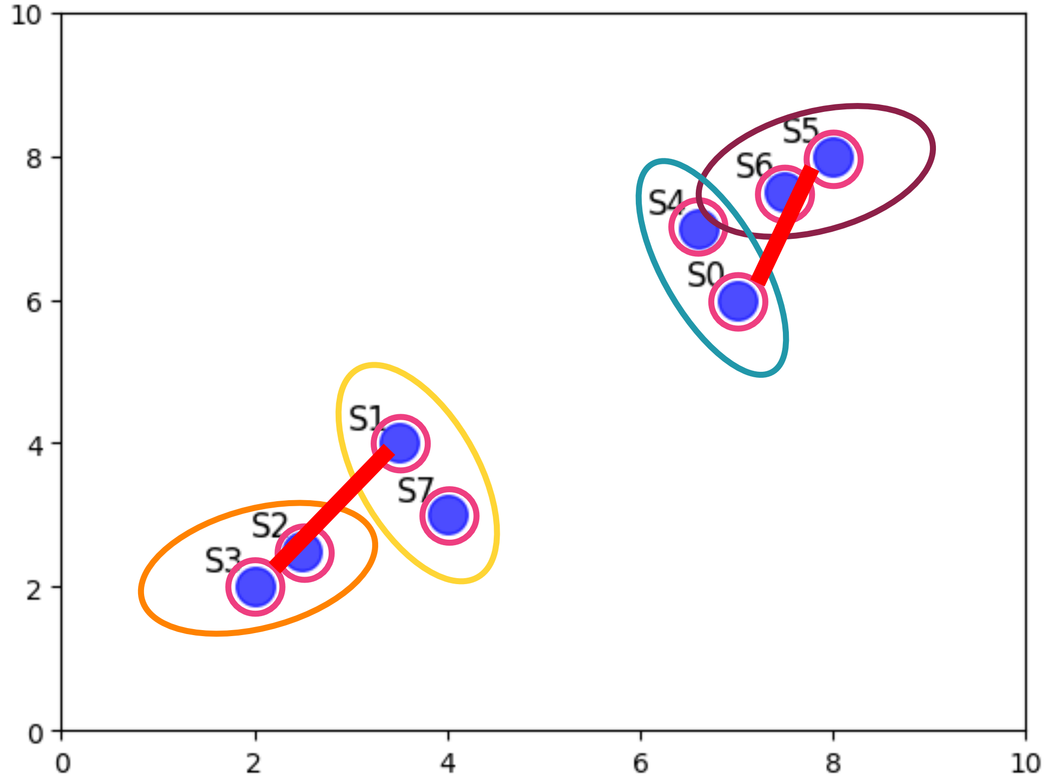
	S0	S1	S2	S3	S4	S5	S6	S7
S0	0.0	4.0	5.7	6.4	1.1	2.2	1.6	4.2
S1	4.0	0.0	1.8	2.5	4.3	6.0	5.3	1.1
S2	5.7	1.8	0.0	0.7	6.1	7.8	7.1	1.6
S3	6.4	2.5	0.7	0.0	6.8	8.5	7.8	2.2
S4	1.1	4.3	6.1	6.8	0.0	1.7	1.0	4.8
S5	2.2	6.0	7.8	8.5	1.7	0.0	0.7	6.4
S6	1.6	5.3	7.1	7.8	1.0	0.7	0.0	5.7
S7	4.2	1.1	1.6	2.2	4.8	6.4	5.7	0.0



THE UNIVERSITY OF TENNESSEE KNOXVILLE

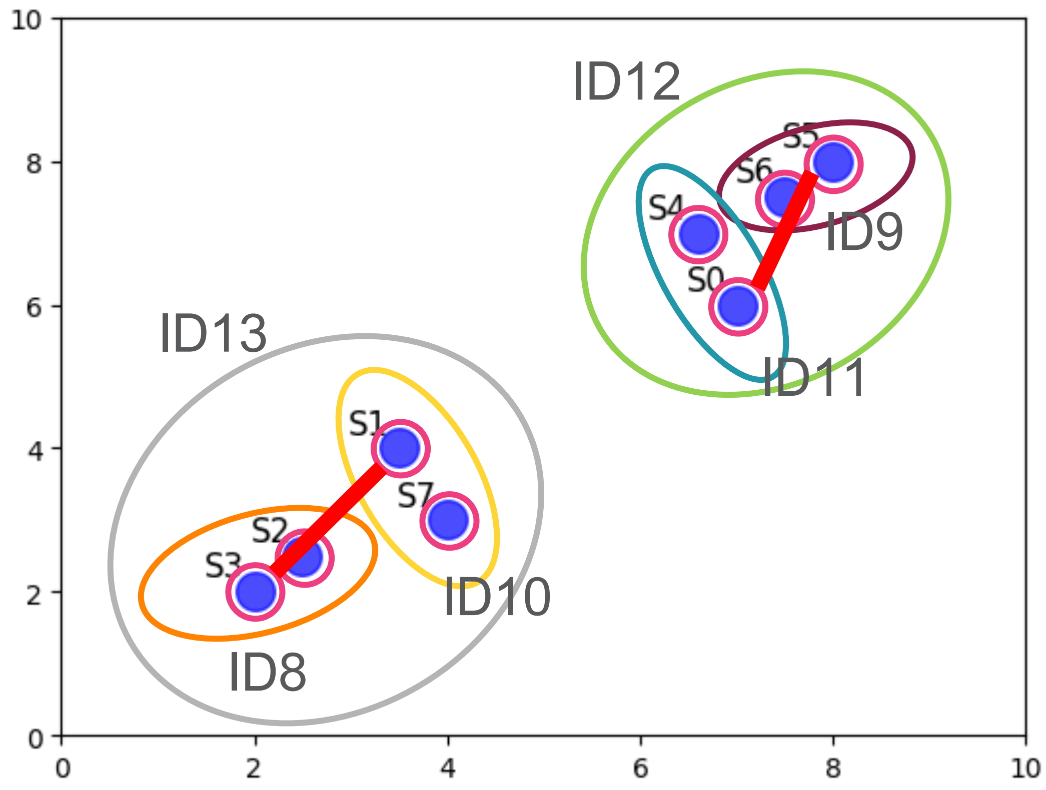
Step-by-Step Example

	S0	S1	S2	S3	S4	S5	S6	S7
S0	0.0	4.0	5.7	6.4	1.1	2.2	1.6	4.2
S1	4.0	0.0	1.8	2.5	4.3	6.0	5.3	1.1
S2	5.7	1.8	0.0	0.7	6.1	7.8	7.1	1.6
S3	6.4	2.5	0.7	0.0	6.8	8.5	7.8	2.2
S4	1.1	4.3	6.1	6.8	0.0	1.7	1.0	4.8
S5	2.2	6.0	7.8	8.5	1.7	0.0	0.7	6.4
S6	1.6	5.3	7.1	7.8	1.0	0.7	0.0	5.7
S7	4.2	1.1	1.6	2.2	4.8	6.4	5.7	0.0



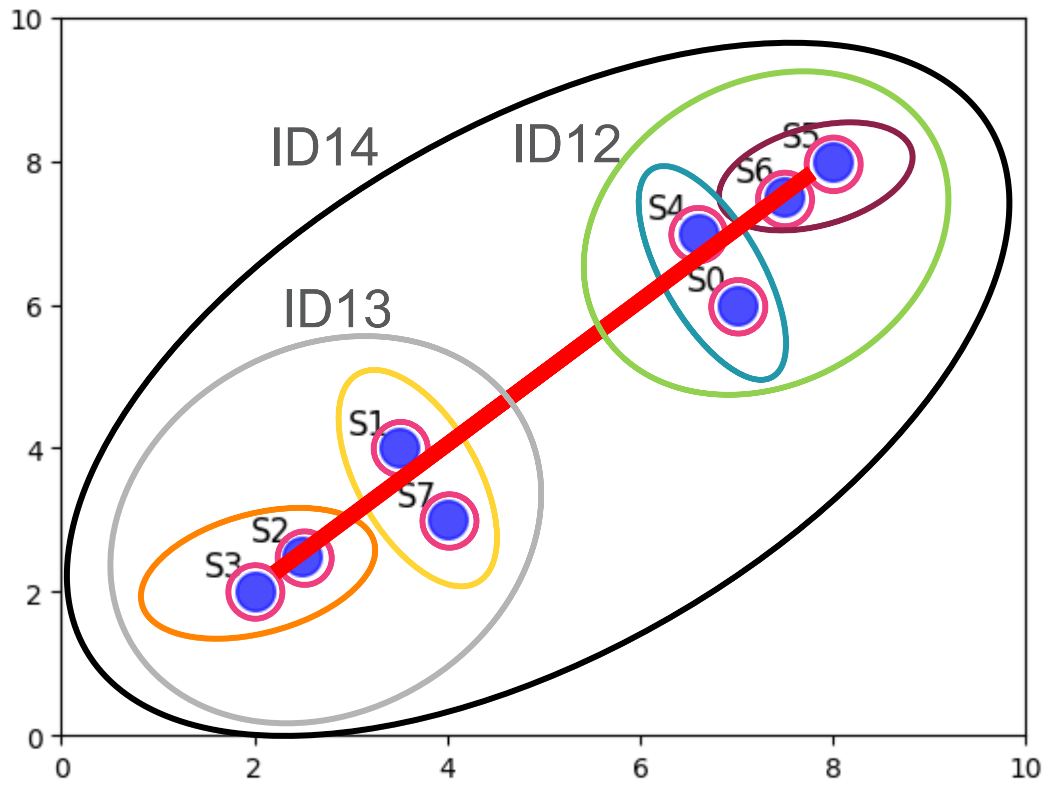
Linkage Matrix Update

Cluster	Label 1	Label 2	Distance	# of samples
Cluster 8	2	3	0.7	2
Cluster 9	5	6	0.7	2
Cluster 10	1	7	1.1	2
Cluster 11	0	4	1.1	2
Cluster 12	9	11	2.2	4
Cluster 13	8	10	2.5	4



Linkage Matrix Update: Final Step

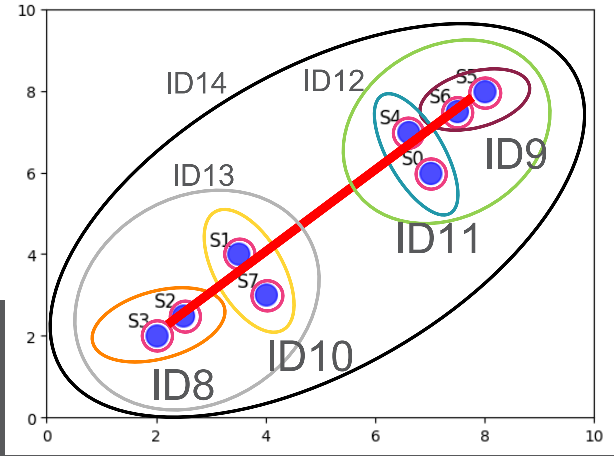
Cluster	Label 1	Label 2	Distance	# of samples
Cluster 8	2	3	0.7	2
Cluster 9	5	6	0.7	2
Cluster 10	1	7	1.1	2
Cluster 11	0	4	1.1	2
Cluster 12	9	11	2.2	4
Cluster 13	8	10	2.5	4
Cluster 14	12	13	8.5	8

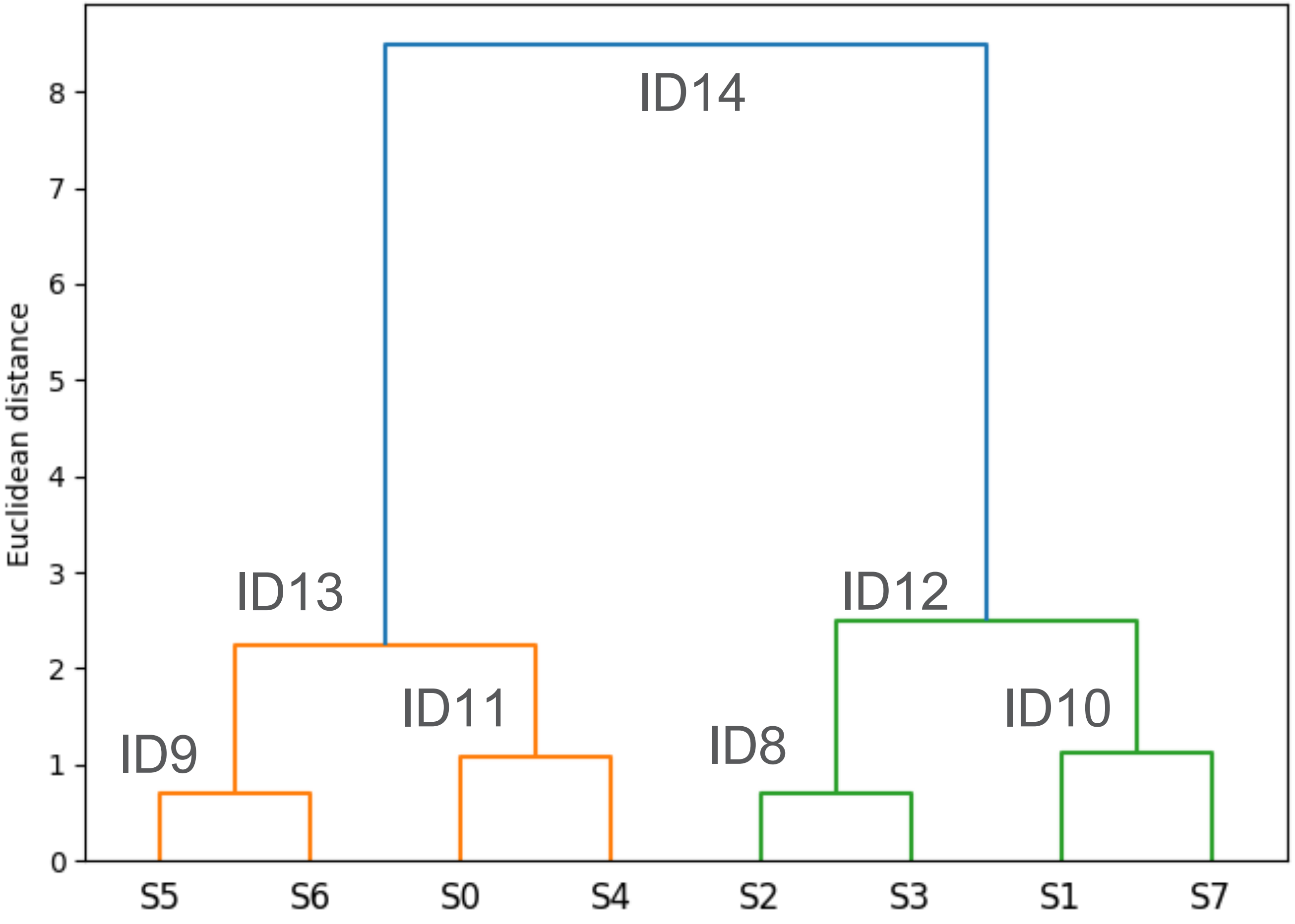


THE UNIVERSITY OF TENNESSEE KNOXVILLE

Dendrogram

Cluster	Label 1	Label 2	Distance	# of samples
Cluster 8	2	3	0.7	2
Cluster 9	5	6	0.7	2
Cluster 10	1	7	1.1	2
Cluster 11	0	4	1.1	2
Cluster 12	9	11	2.2	4
Cluster 13	8	10	2.5	4
Cluster 14	12	13	8.5	8

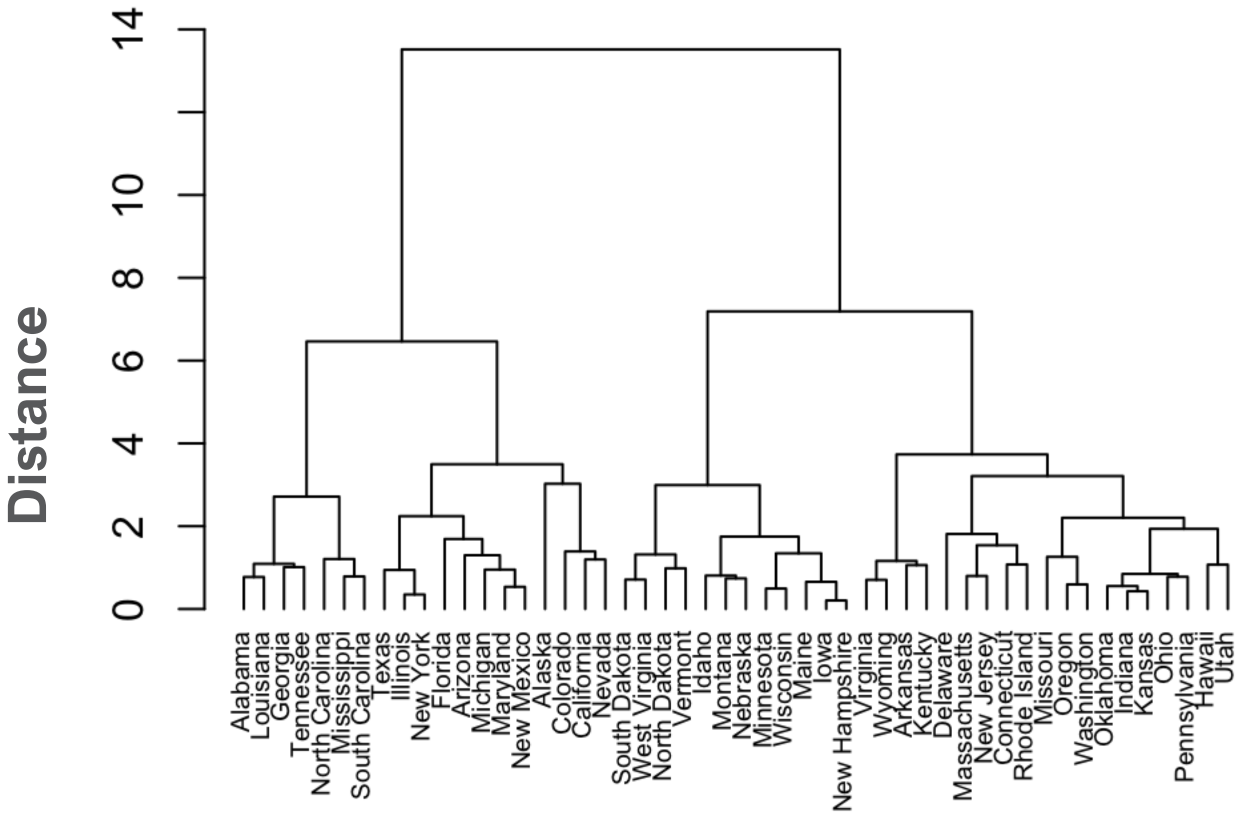




Dendrogram

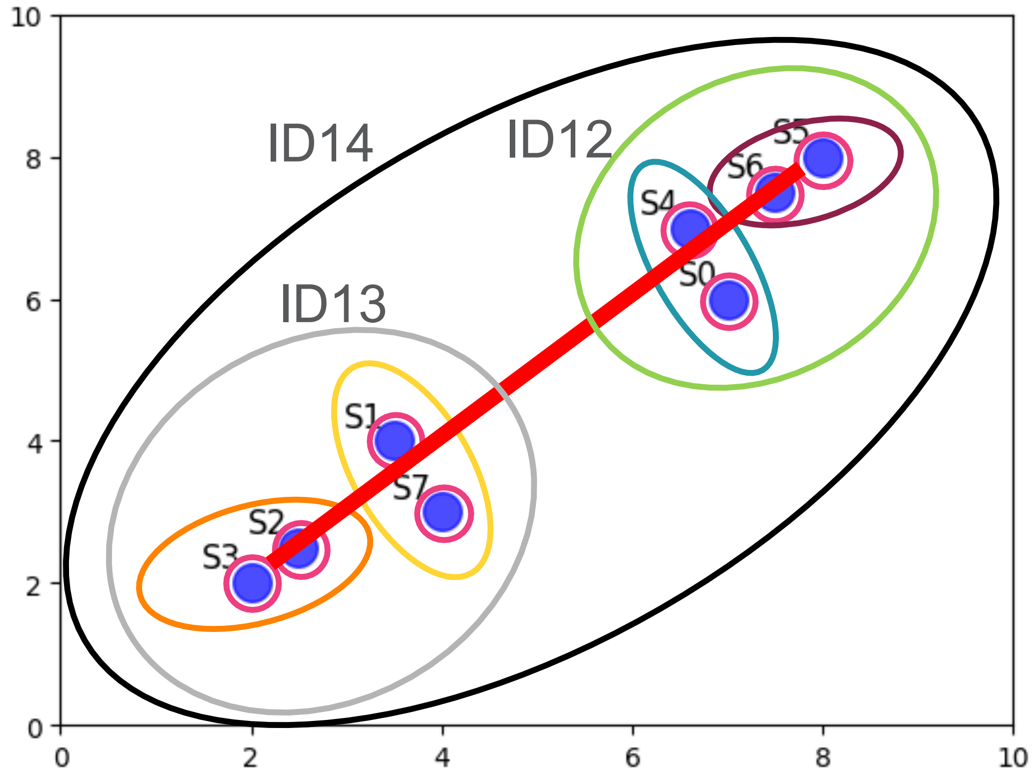
- Hierarchical relationship
- Selecting the number of clusters
- Similarity between clusters
- Identifying outliers
- Cluster stability
  - Multiple runs

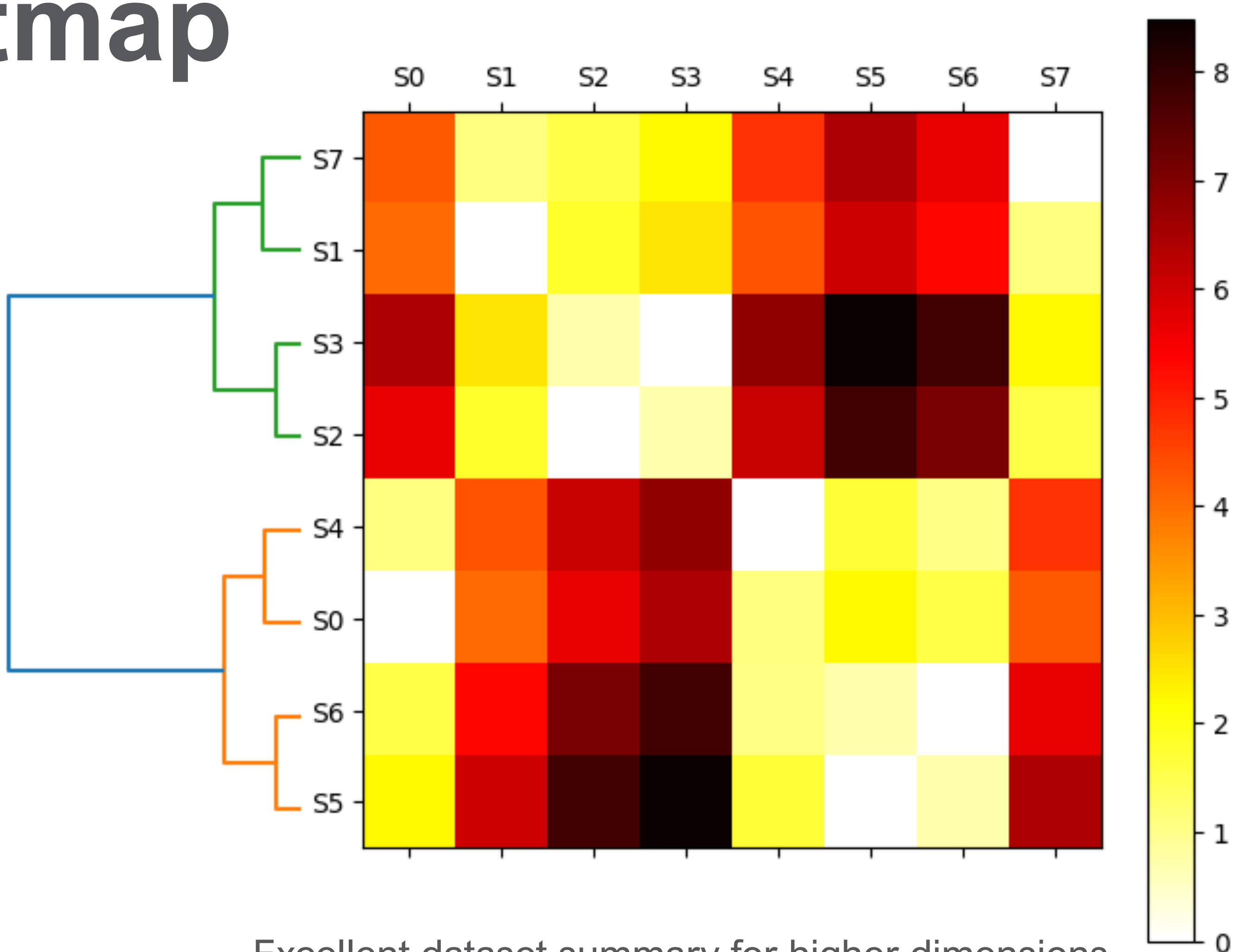
USA Arrests Cluster Dendrogram



https://www.kaggle.com/datasets/halimedogan/usarrests

Dendrogram + Heatmap





Excellent dataset summary for higher dimensions

Applications of Hierarchical Clustering

- Customer segmentation
- Image segmentation
- Anomaly detection
- Recommendation systems
- Document retrieval
- Market Segmentation

Gosal, et. al., “Impact of complete lockdown on total infection and death rates: A hierarchical cluster analysis,” 2020.

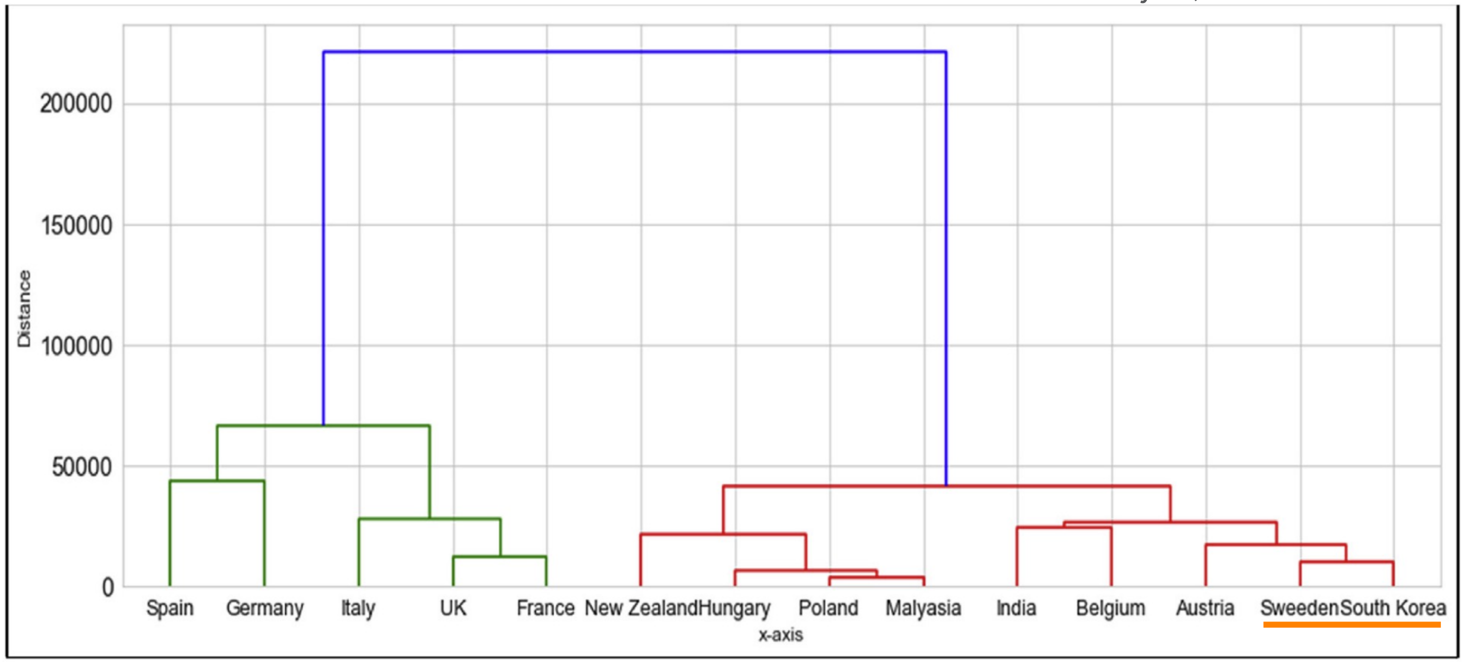


Fig. 4. Hierarchical clustering of death rates of the 17 countries (15 with lockdown & 2 without lockdown divided into two clusters). X-axis: countries; Y-axis: distance from the mean. Cluster 1: Spain, Germany, Italy, UK, and France. Cluster: Belgium, Austria, New Zealand, India, Hungary, Poland, Malaysia, Sweden and South Korea. Blue lines: Division of the two main clusters. Green lines: Individual countries under cluster 1. Red lines: Individual countries under cluster 2.

Pop Quiz

True or False. A dendrogram visualizes the correlations between data features.

A. True

B. False

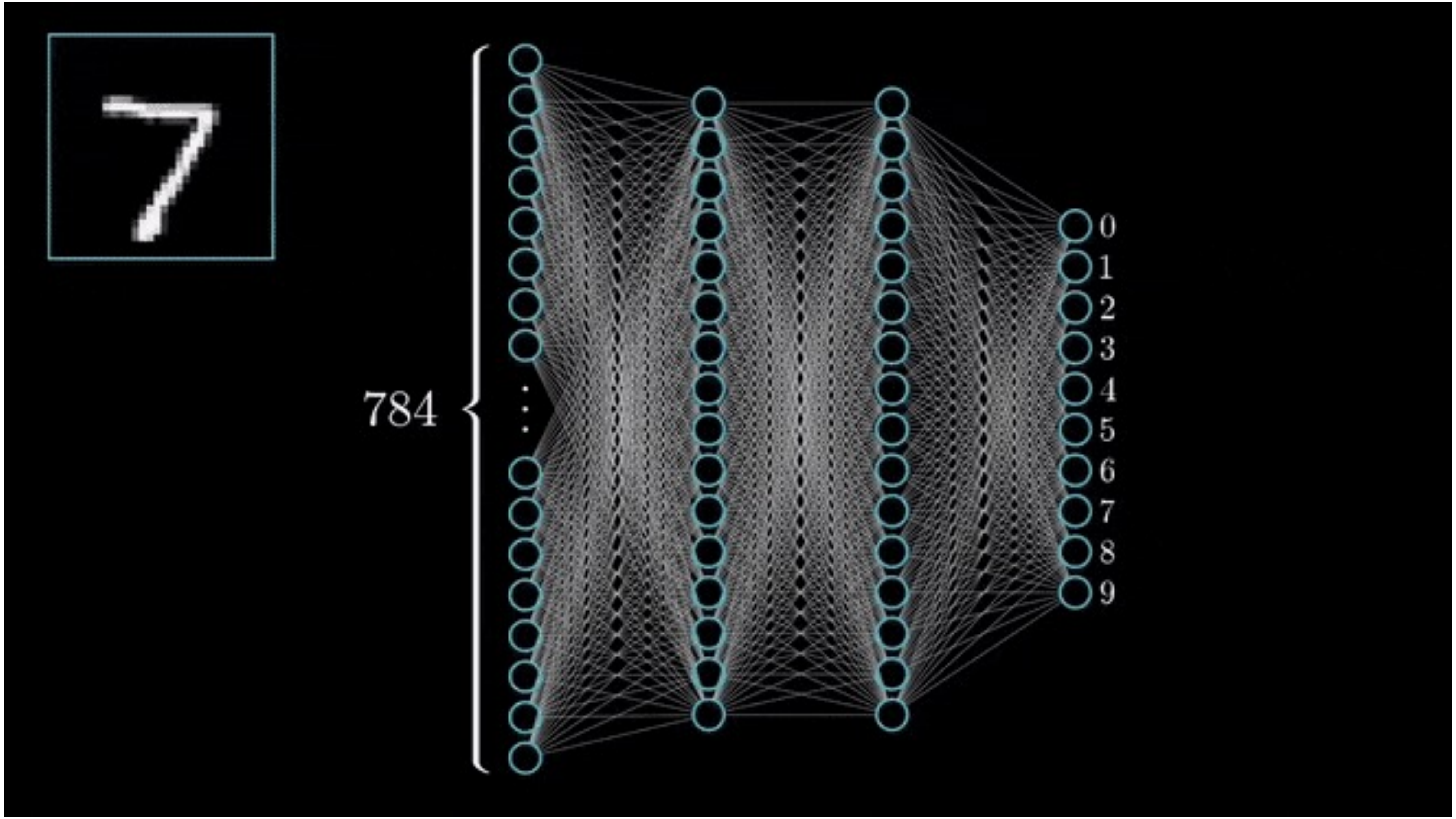
Artificial Neural Networks

THE UNIVERSITY OF TENNESSEE KNOXVILLE

For a deep dive into ANNs and Deep Learning

COSC 425/524 this Spring

https://ttpoll.com/p/306471

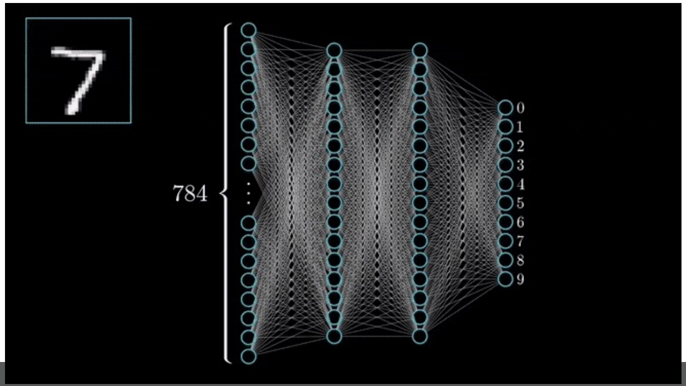


Credit: Ashis Kumer Biswas, UC Denver

Why does it work?

The short answers:

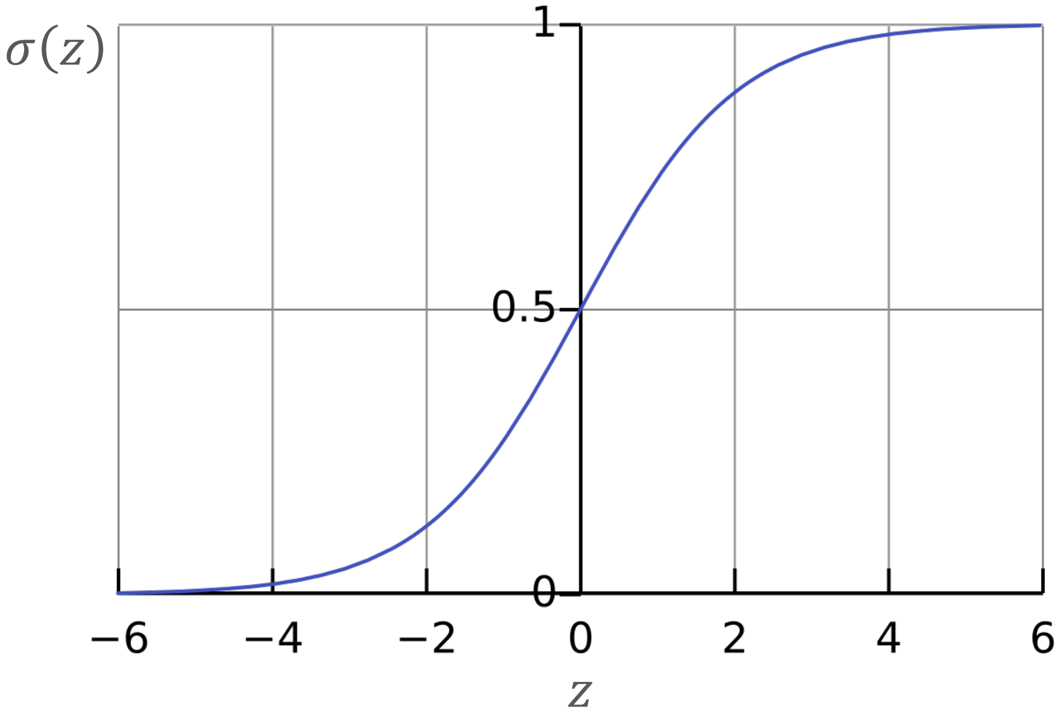
1. Universal approximation theorem – Any continuous function can be represented using a ***feed-forward neural network*** (does not discuss how to learn)
2. The ***backpropagation algorithm*** is an efficient way to learn the weights in a feedforward network.



Source Dr. Amir Sadovnik

Logistic Regression

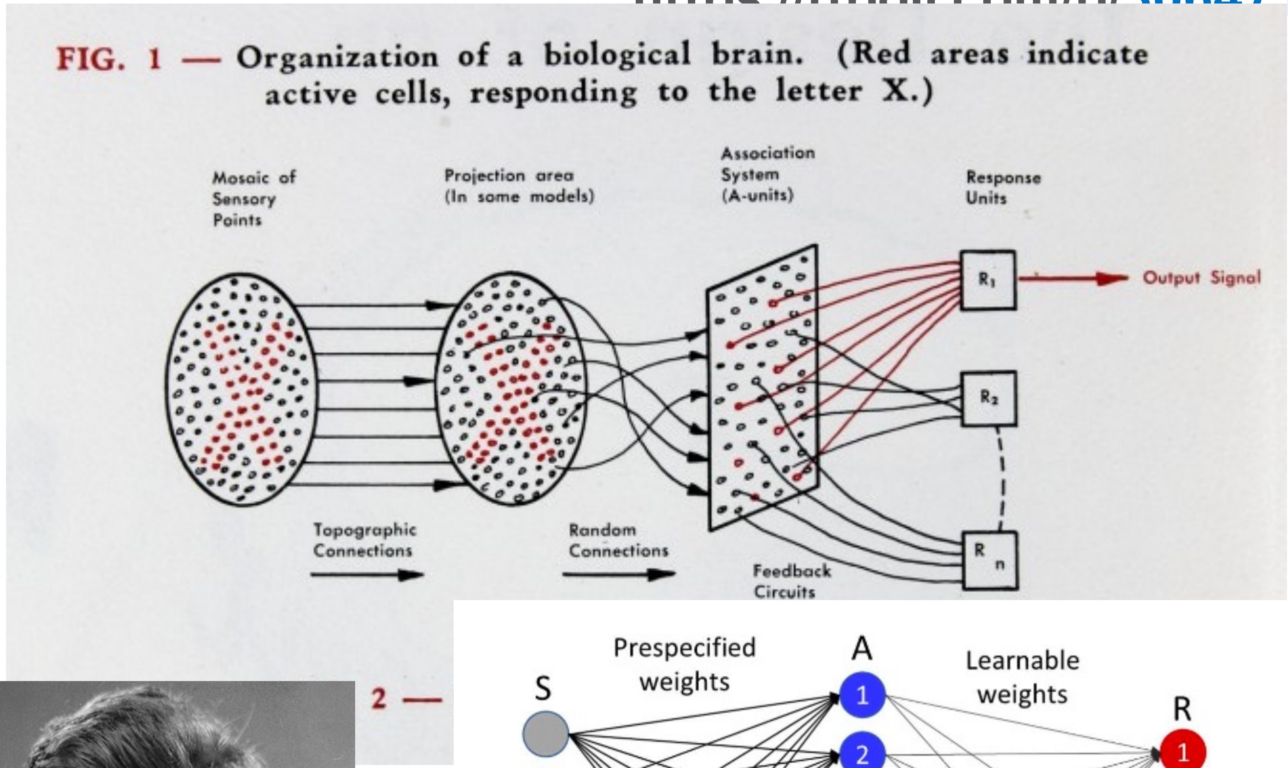
- Linear regression: $\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_m x_m = X\theta$
- Logistic regression:
  - $z = X\theta$
  - $\hat{p} = \sigma(z)$
  - $\sigma(z) = \frac{1}{1+e^{-z}} \Rightarrow \hat{p} = \frac{1}{1+e^{-X\theta}}$
  - $\hat{y} = \begin{cases} 1, \hat{p} \geq 0.5 \\ 0, \hat{p} < 0.5 \end{cases}$

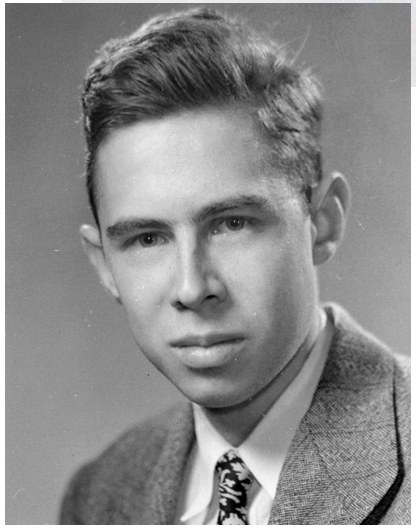


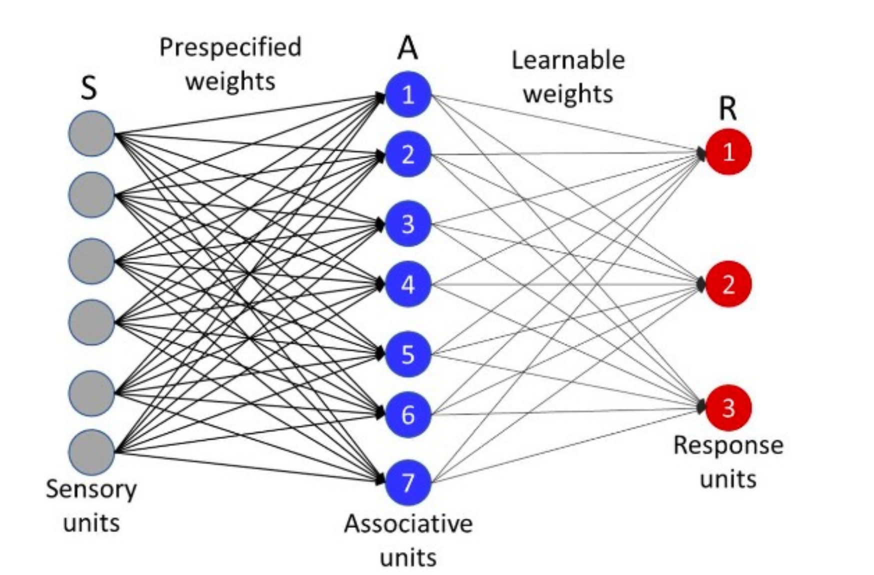
THE UNIVERSITY OF TENNESSEE KNOXVILLE

Rosenblatt’s Mark I: The Perceptron

- Frank Rosenblatt was a psychologist and logician
- Inventor of the perceptron in1957 with the paper “The Perceptron—a perceiving and recognizing automaton”
- Funded by the U.S. Office of Naval Research
- Inspired by McCulloch-Pitts (MCP) neuron





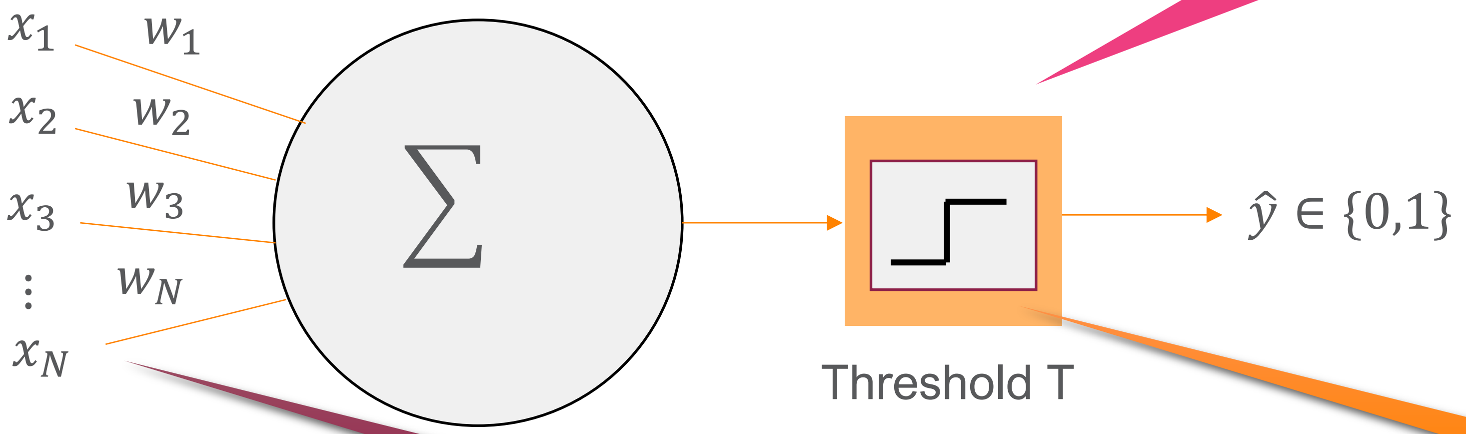


https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon

Simplified Perceptron

Bias

Activation Function

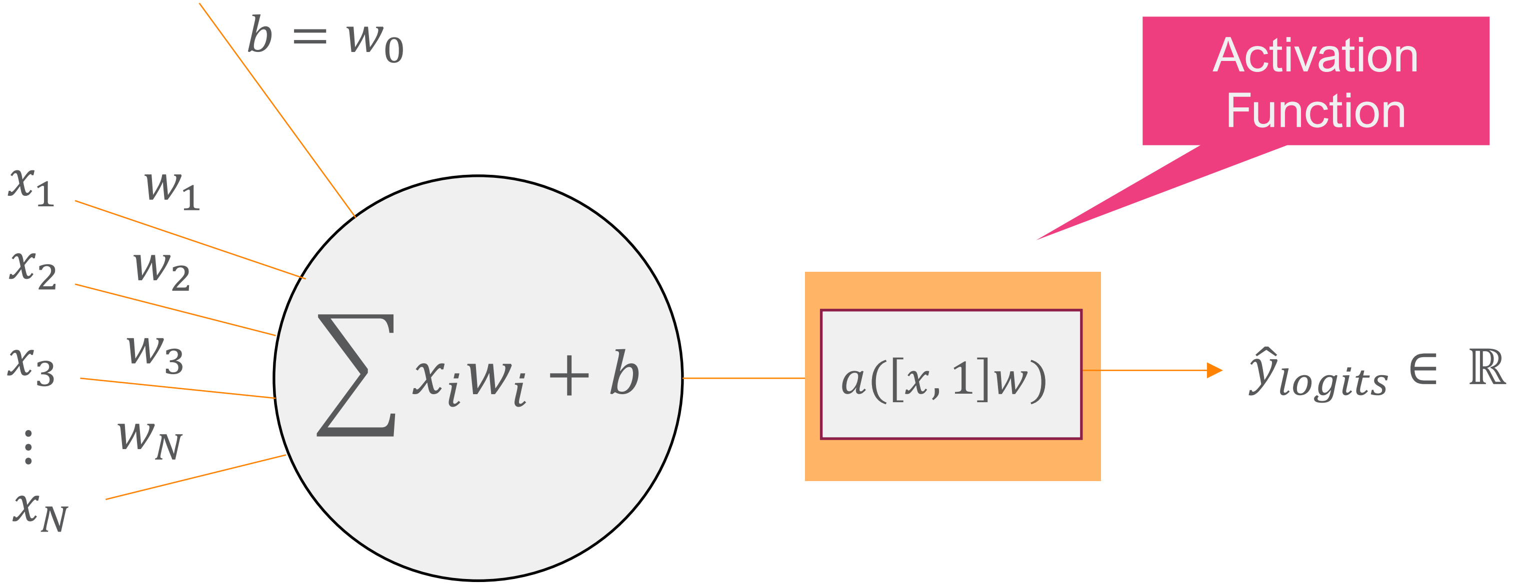


$$\hat{y} = \begin{cases} 1 & if \sum w_i x_i - T \geq 0 \\ 0 & otherwise \end{cases}$$

Binary inputs and outputs.

Step function

Modern Perceptron



$$\hat{y} = \begin{cases} 1 & if\ \hat{y}_{logits} \geq T \\ 0 & otherwise \end{cases}$$

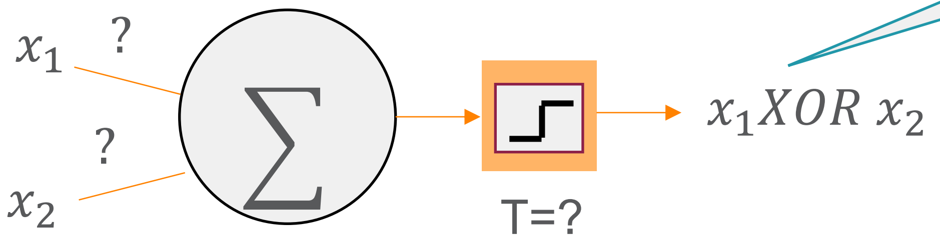
$x_i \in \mathbb{R} \quad w_i \in \mathbb{R}$

Real inputs and outputs.

Limitations

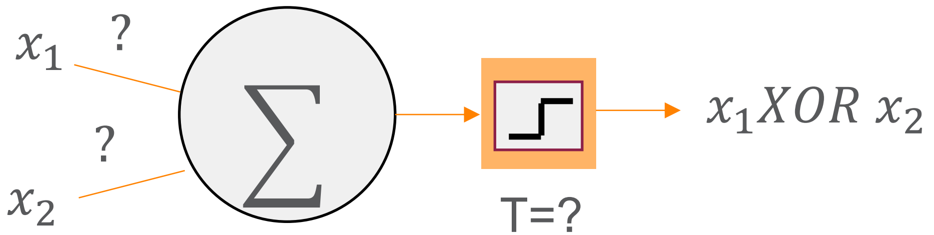
- Still, binary inputs and outputs

Can we compute XOR?

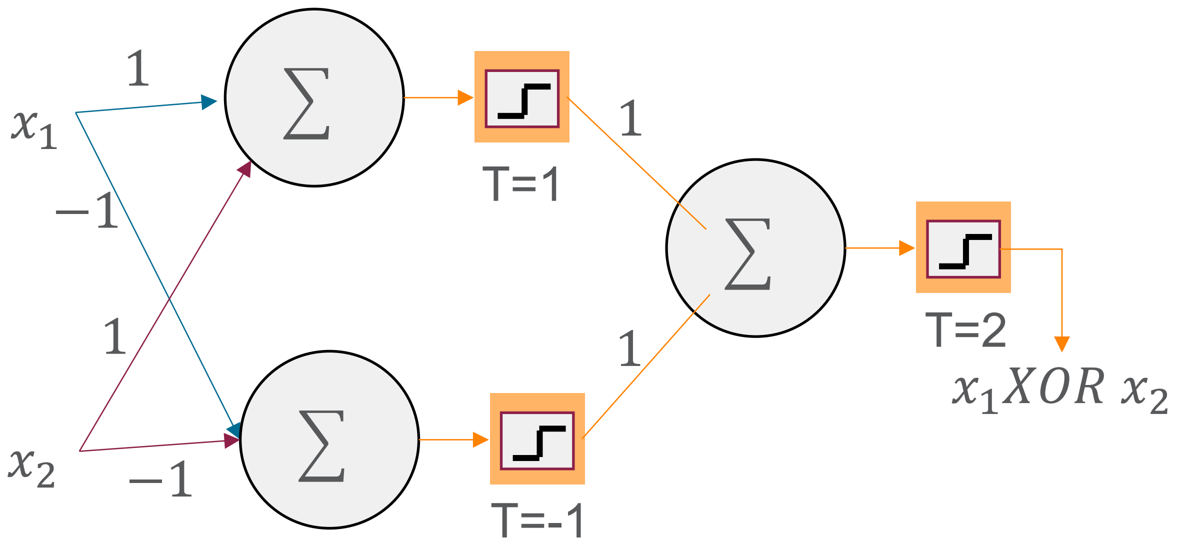


Limitations

- Still, binary inputs and outputs
- No solution for XOR or non-linear functions

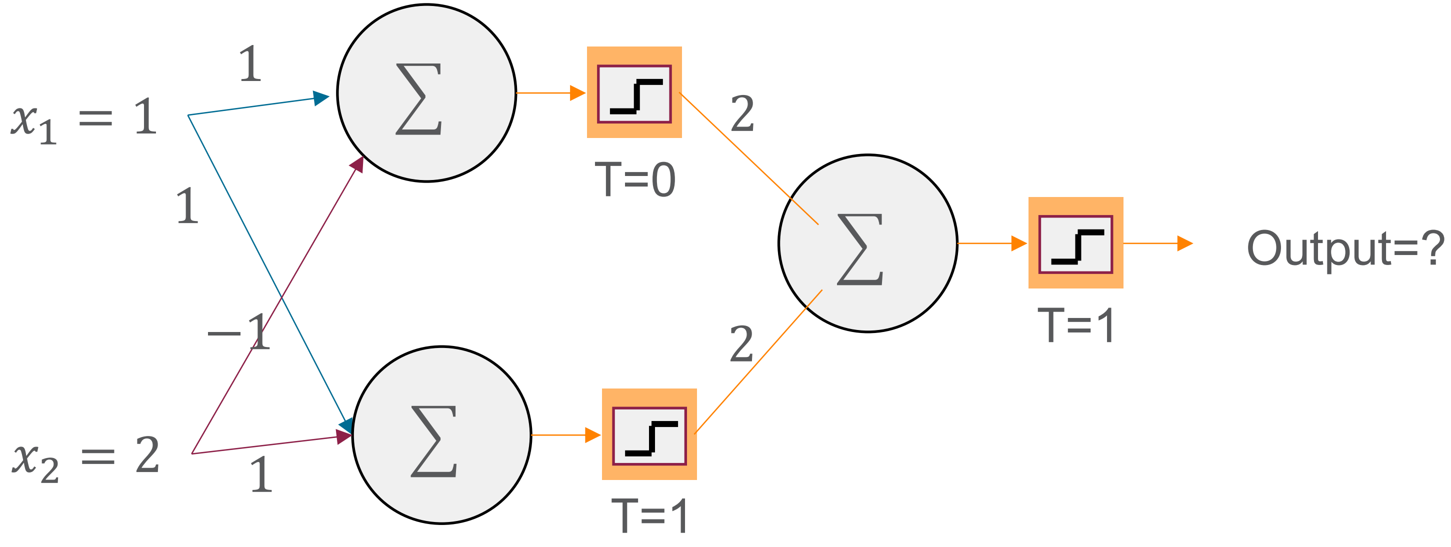


Solution: Multi-layer Perceptron



More on MLP soon.

Pop Quiz



A. Output=-1

B. Output=0

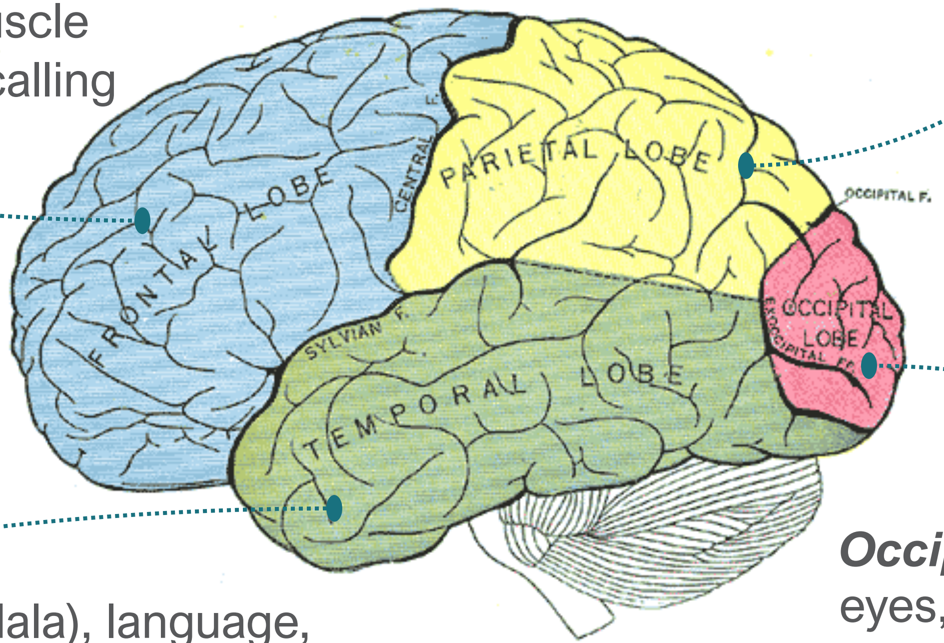
C. Output=1

D. Output=2

Human Brain

Parietal lobe: Integrates the brain processes, writing, self-perception, location awareness, and touch senses such as pressure, heat, cold, vibrations, and pain.

Frontal lobe: Reasoning, social understanding, executive functions, voluntary muscle movements, and learning and recalling information.



Temporal lobe: Emotions (Amygdala), language, memory (hippocampus), sight and sound processing, and object recognition.

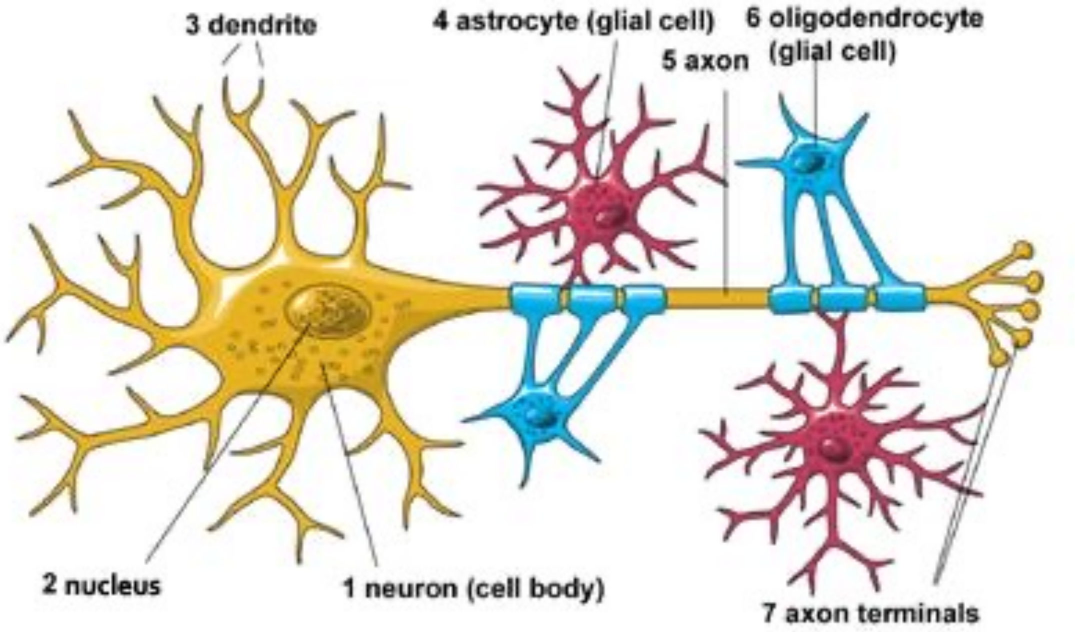
Occipital lobe: direct connection to the eyes, visuospatial processing, distance and depth perception, color determination, object and face recognition.

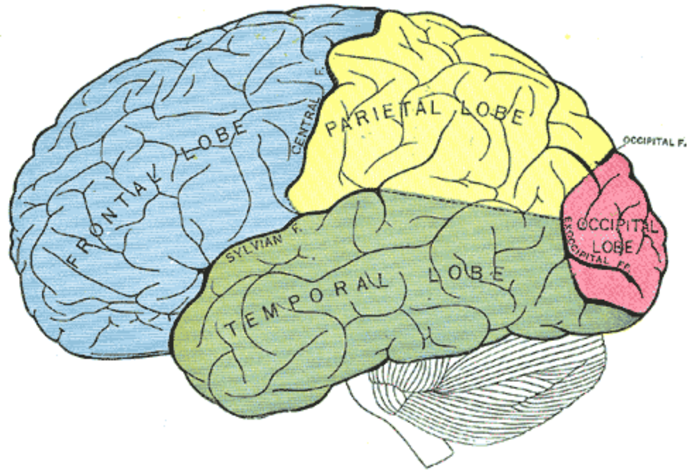
Image: Wikipedia. Sources: Cleveland Clinic.

NIH Brain Basics: Link

A neuron

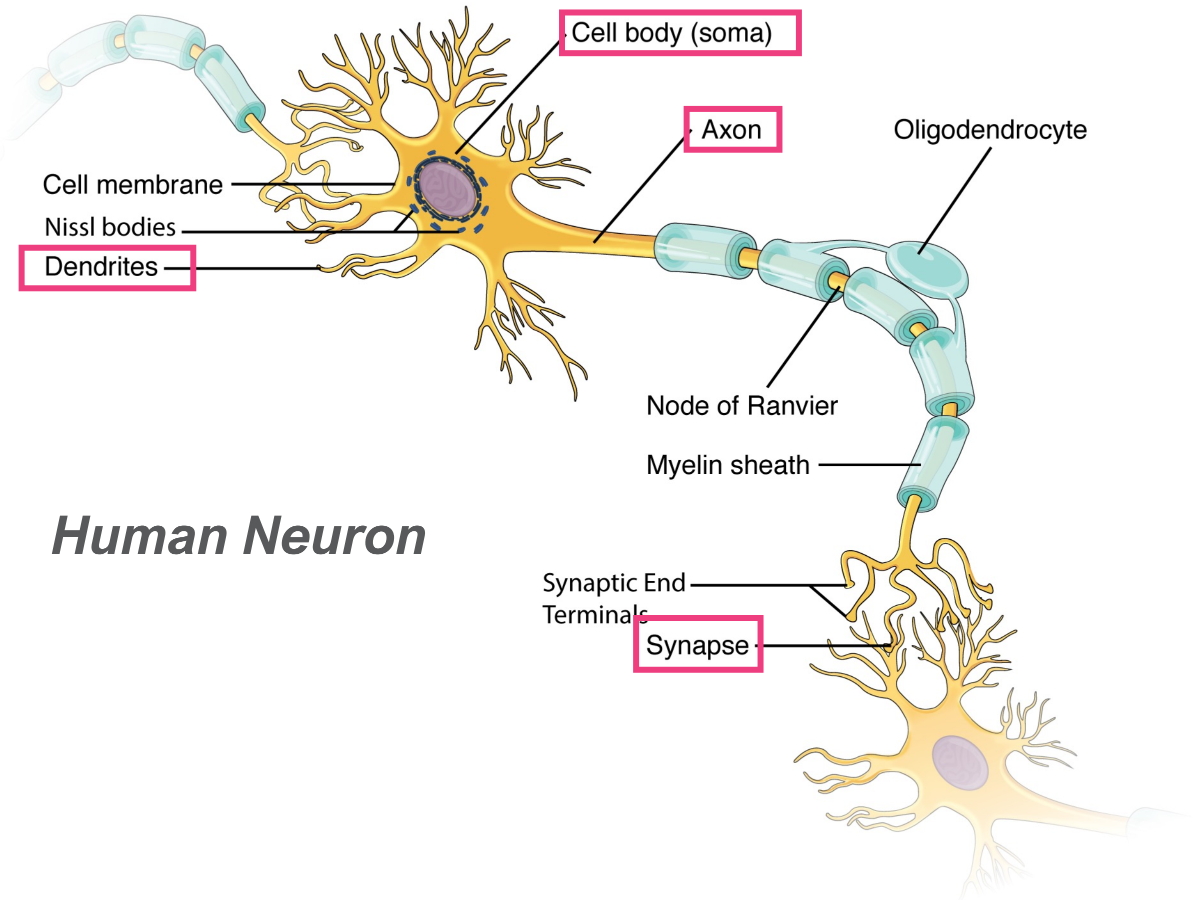
- Dendrites – receive signals from other neurons
- Soma (Cell Body) – Process information received from dendrites
- Axon – Transmit the output of the Soma
- Synapses – Small connections between axon and dendrites





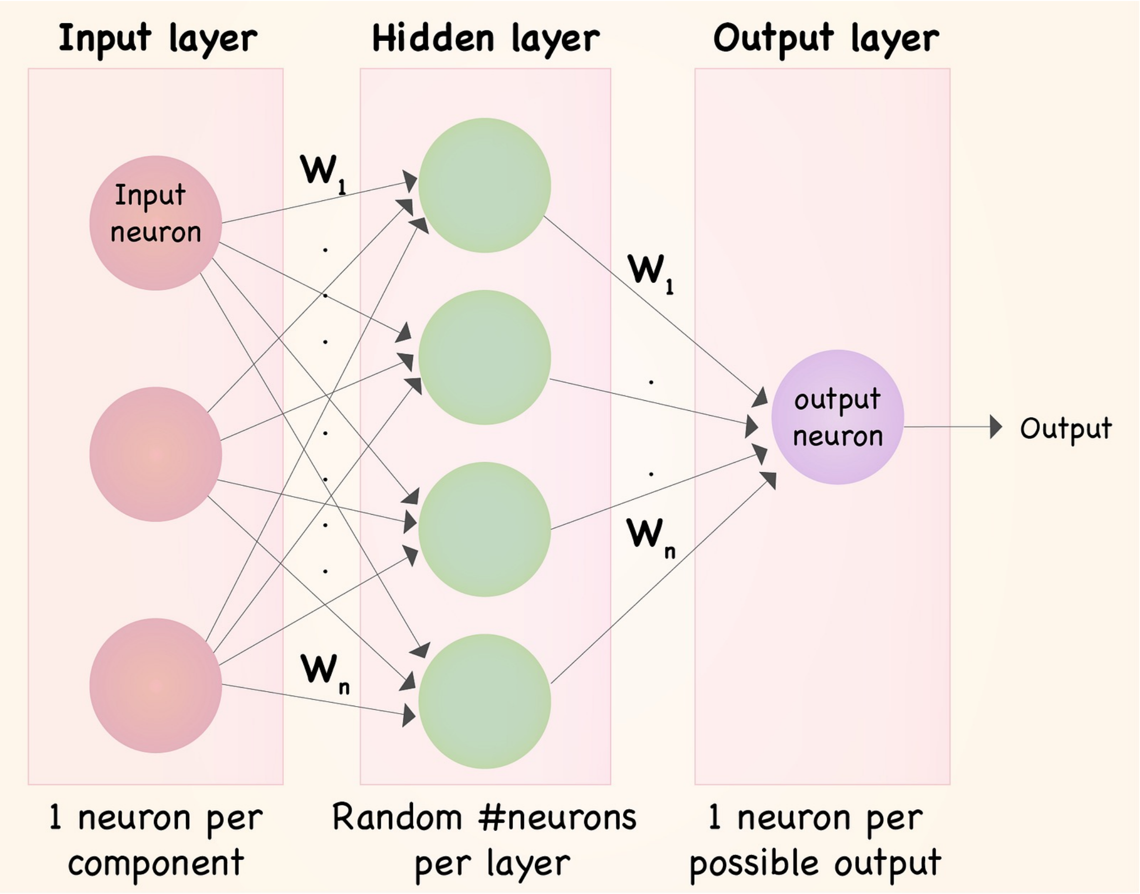
THE UNIVERSITY OF TENNESSEE KNOXVILLE

Deep Learning



Human Neuron

Artificial Neural Network (NN)



https://ttpoll.com/p/306471

Touching a flame.

Response at scale.

Mid 1800s: The brain is comprised of interconnected neurons.
~100 Trillion Connections

Emotions
Movement

Connectionism (1873)

- Alexander Bain: philosopher, psychologist, mathematician, logician, linguist, professor
- Main ideas in the book “Mind and Body”
  - Neural groupings
    - Neurons excite and stimulate each other
    - Different input combinations can result in different outputs
    - Activation intensity influences the activation of connected neurons
  - Making memories
    - Neurons connections strengthen with repetitive inputs (Before Hebb’s Law 1949)

https://ttpoll.com/p/306471

Hebb's Law: Model for Neural Plasticity

- Novelist, schoolteacher, psychologist
- Main idea in book "The Organization of Behavior" (1949):
  - If neuron A repeatedly triggers neuron B, the synapses connecting these neurons get larger.
  - Hebb's Law: "Neurons that fire together wire together."

Response of a neuron x_j

$$x_j = f\left(\sum_{i=1}^{N} w_i x_i\right)$$

Weights can be different now.

Evolution of synaptic weight

$$\Delta w_{ij} = \eta x_i x_j$$

Weight Update

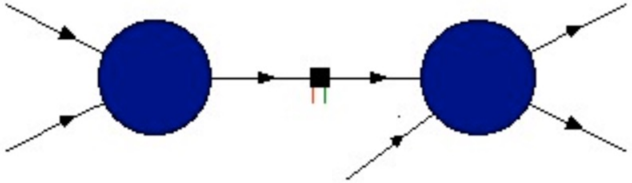
Learning Rate

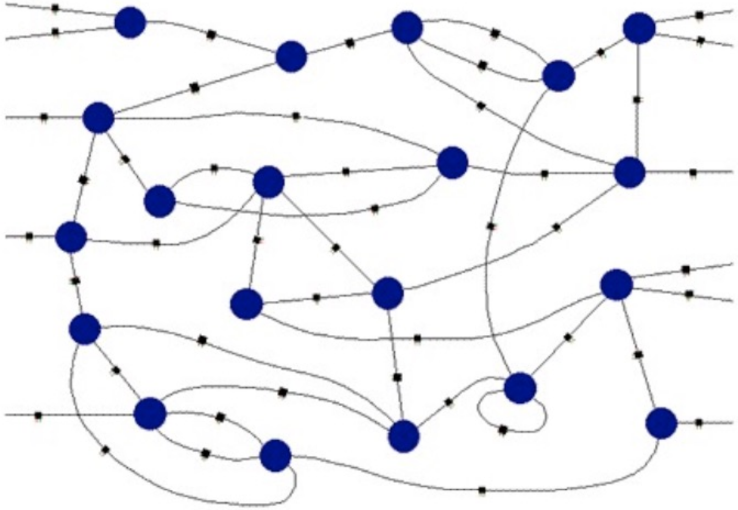
Interaction between neurons i and j

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Connectionist Machines

- Multiple connectionist paradigms proposed
  - Alan Turing's Connectionist model (1948):
  - Parallel Distributed Processing (1986)
    - Rumelhart, Hinton, McClelland
  - Requirements of a connectionist system
    - Bechtel and Abrahamson (1991)
- Main properties
  - Network of processing elements
  - All world knowledge is stored in the connections between the elements

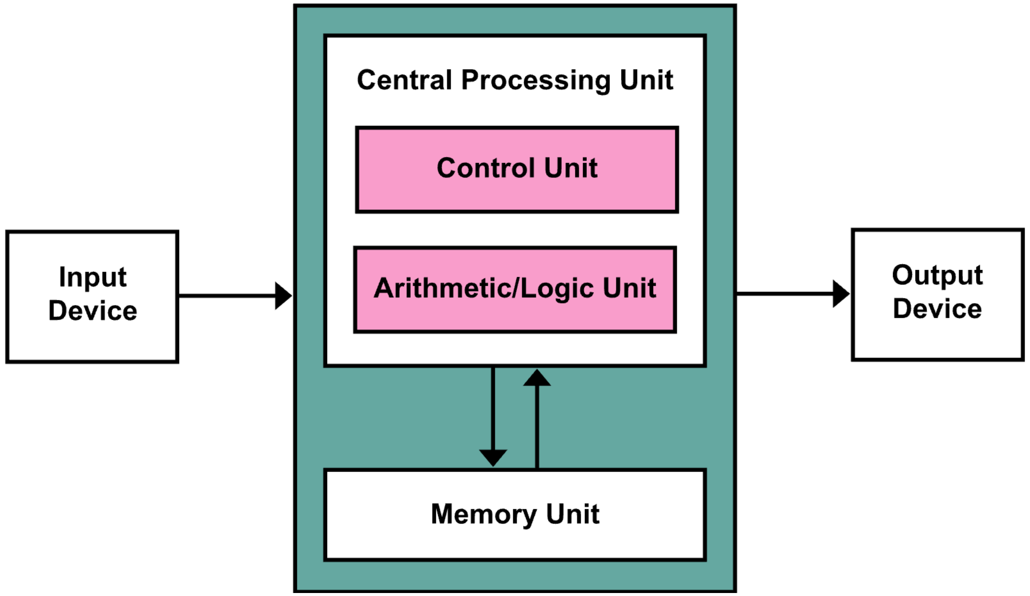




Source: Carnegie Mellon Univeristy, Bhiksha Raj and Rita Singh, 11-785 Introduction to Deep Learning

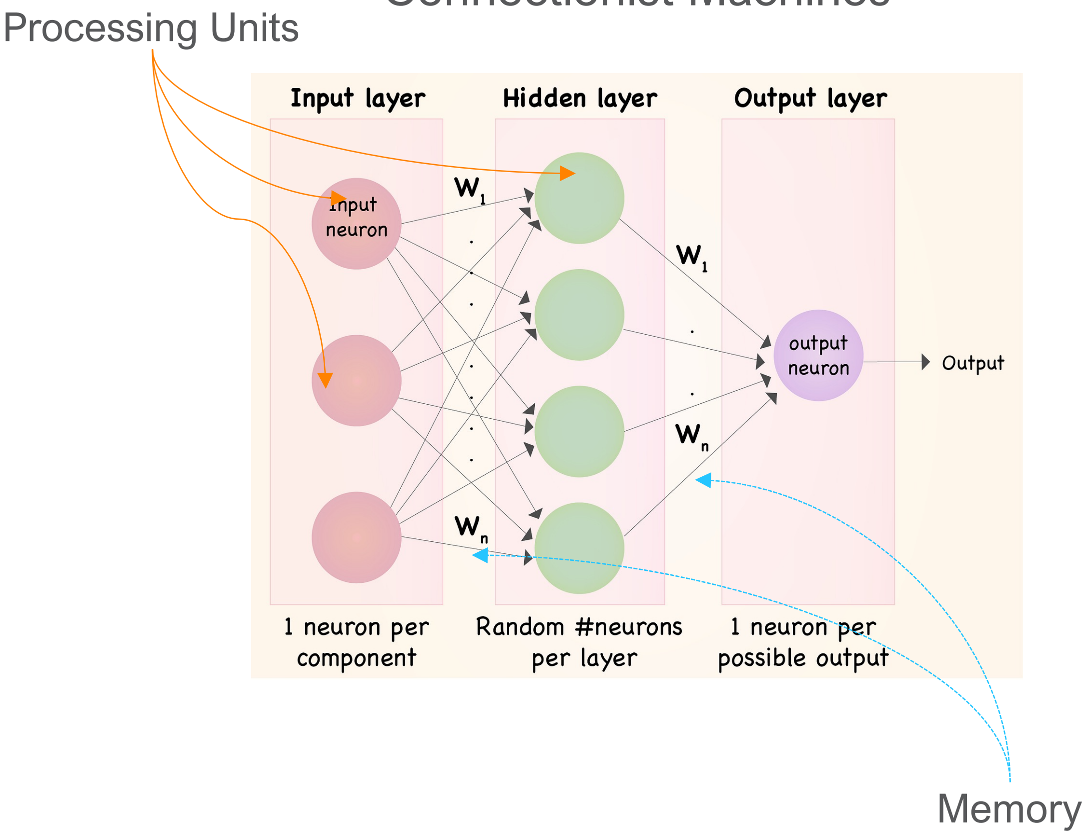
Von Neumann vs Connectionist Machines

Von Neumann Machines



The typical modern computer!

Connectionist Machines



Wikipedia is the source of the Von Neumann Machine diagram.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Pop Quiz

A neural network is a **Von Neumann Machine** because it is a network of processing elements, and all world knowledge is stored in the connections between the elements.

A. True

B. False

Review

- Hierarchical Clustering
  - There is no need to know the number of clusters beforehand
  - Find inner cluster patterns
  - Explains the relationship between samples
- ANN
  - Connectionism machines
    - Network of processing units
    - Memory is in the connections
  - Key developments
    - Backpropagation
    - ReLU simple activation function
    - Deep Convolutional Networks

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Next Lecture

- Deep Neural Networks

Helper Slides

https://ttpoll.com/p/306471