PH451, PH551

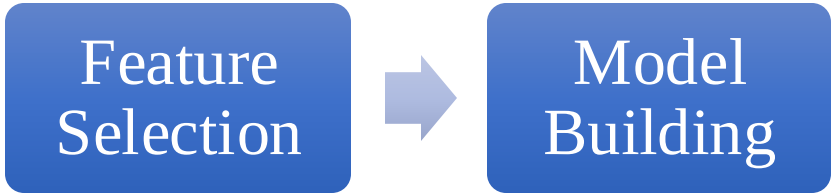
February 18, 2025

Announcements

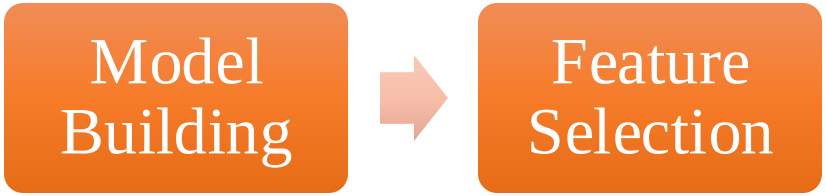
- **Hands-on #4 – due Thu**
- **Hackathon #1 – due Fri**
- **Hands-on #5 – due next Thu**

Recap: Feature Selection

Filters



Wrappers



Embedded-Hybrid

Recap: Wrapper Methods

Selection tied to a model:

- More accurate
- Assess feature interactions
- Search for optimal subset of features

Model Building

Types:

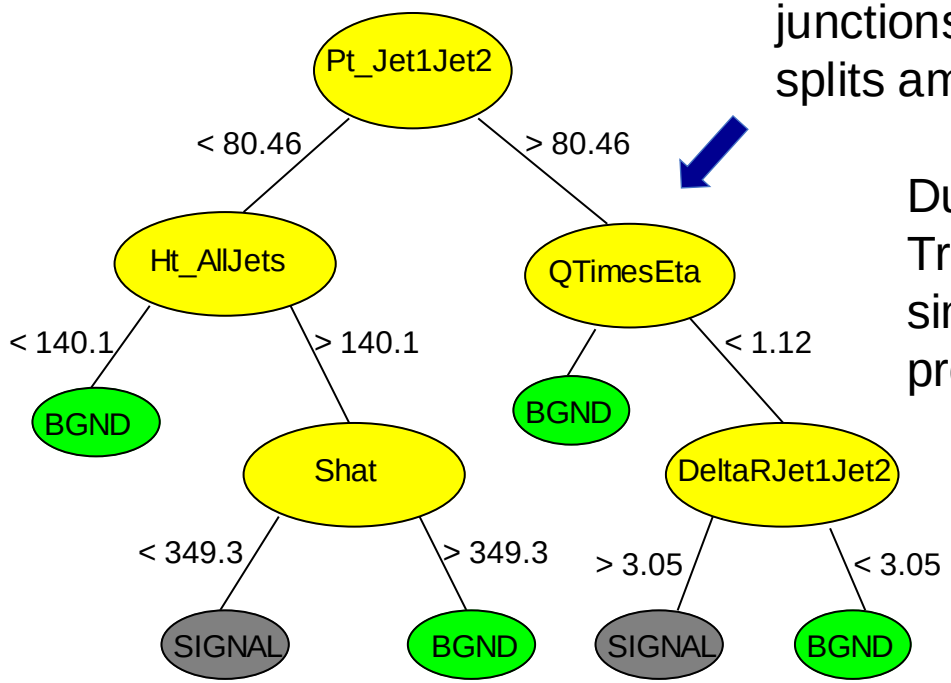
- **Methodical**
- **Probabilistic**
  - random hill-climbing
- **Heuristic**
  - forward backward elimination

Feature Selection

Recap: Regularized Trees

Inspired by J. Friedman and Popescu, 2008 work on rules regularization

Decision Tree:



Votes taken at decision junctions on possible splits among the features

During voting Regularized Trees **penalize** features similar to those used in previous decisions

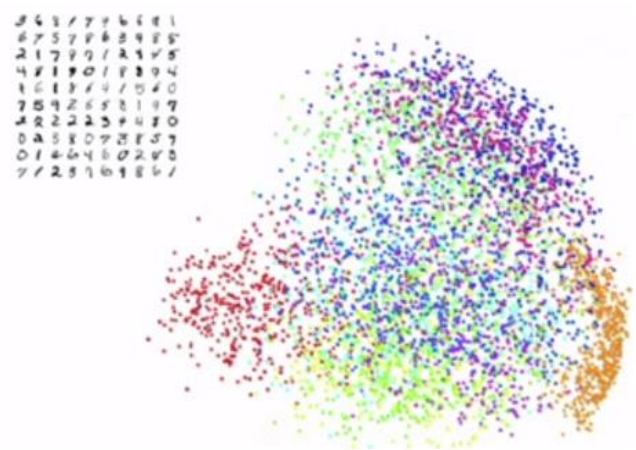
Recap: PCA

Principal Components

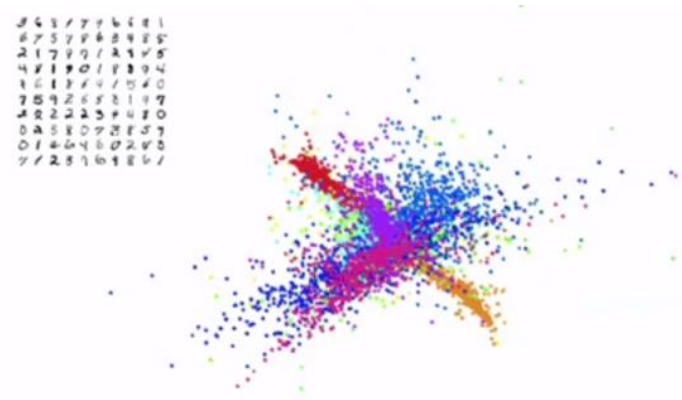
- Linear Method, Pearson/Hotelling 1901/1933
- Find the **hyperplane closest** to data and project into it
  - Minimize the **squared distance** between original data and projection
  - Orthogonal axes that **maximize remaining variance** (principal components)
  - Find with Singular Value Decomposition (SVD)
  - Ignore components of lesser significance

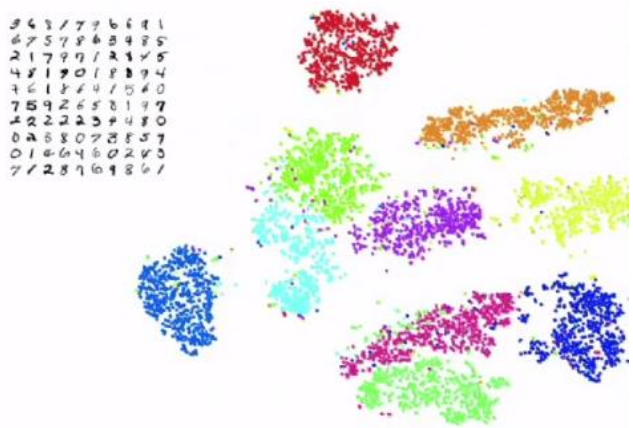
MNIST

PCA



LLE





t-SNE

Unsupervised Learning

Why?

- **Supervised learning:**
  - **Class labels**
- **Unsupervised**
  - **No labels**
    - **Part of human learning experience**

Applications

- **Dimensionality Reduction and Visualization**
  - **All methods we met so far were unsupervised (PCA/LLE/t-SNE)!**
- **Anomaly Detection**
- **Recommendations**
- **Clustering**

Clustering

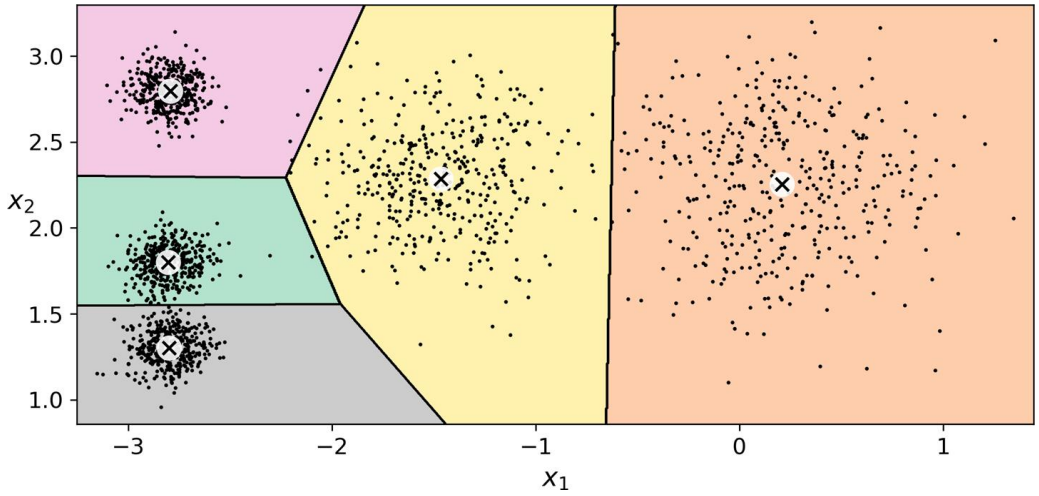
- **Find groups of instances similar to each other and different from instances in other groups**
- **Key questions:**
  - **How many groups?**
  - **How to partition the data?**

K-Means

Lloyd, 1957

- Popular algorithm
- Relatively quick
- Find cluster centers (centroids)
- Assign instances to closest cluster
- Note: have to specify # of clusters (k)

K-Means Algorithm



- **Randomly** pick k instances as centroids
- **Label**, **update** centroids, repeat…

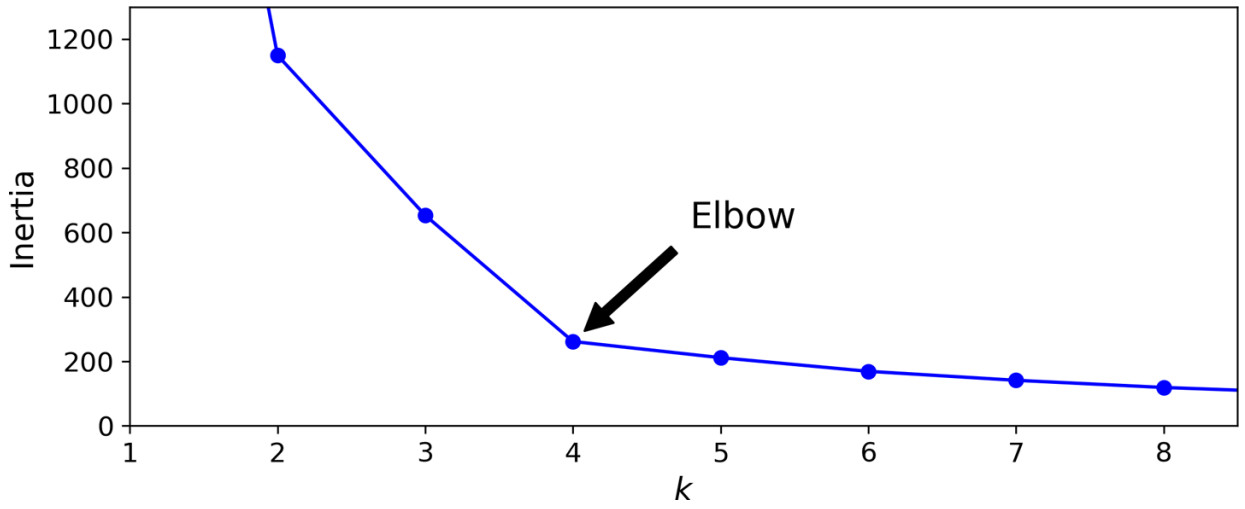
- **Issues:** may converge to local minimum instead of global, strong dependence on initialization
- **Improvements**: K-Means++ (2006) start with centroids that are distant from each other, mini-batches

Performance Metric

Inertia

- Mean squared (L2) distance between each instance and closest centroid

"Elbow" method



Predictive Clustering

Example of a **multi-function regression** model based on trees or rules

- **Decision trees** are equated to clustering trees by P. Langley in 1996, first noted by Fisher in 1993
- **Cluster "hierarchy"** each tree node corresponds to a cluster. Root node contains full dataset partitioned recursively into sub-clusters

Predictive Clustering Trees

Predictive clustering implementation

- **Decision tree and rule induction** system
- Designed for **multi-task learning** and multi-label classification
- Well-suited for both **classification** and **regression** problems

Clustering Concept

Use **decision tree induction** to obtain clusters with:

- **minimal intra-cluster distance**
  - between examples from the same cluster
- **maximal inter-cluster distance**
  - between examples from different clusters
  - In classification trees distance metric is class entropy

Illustrative Example

14 input variables {a, b, c, d…}

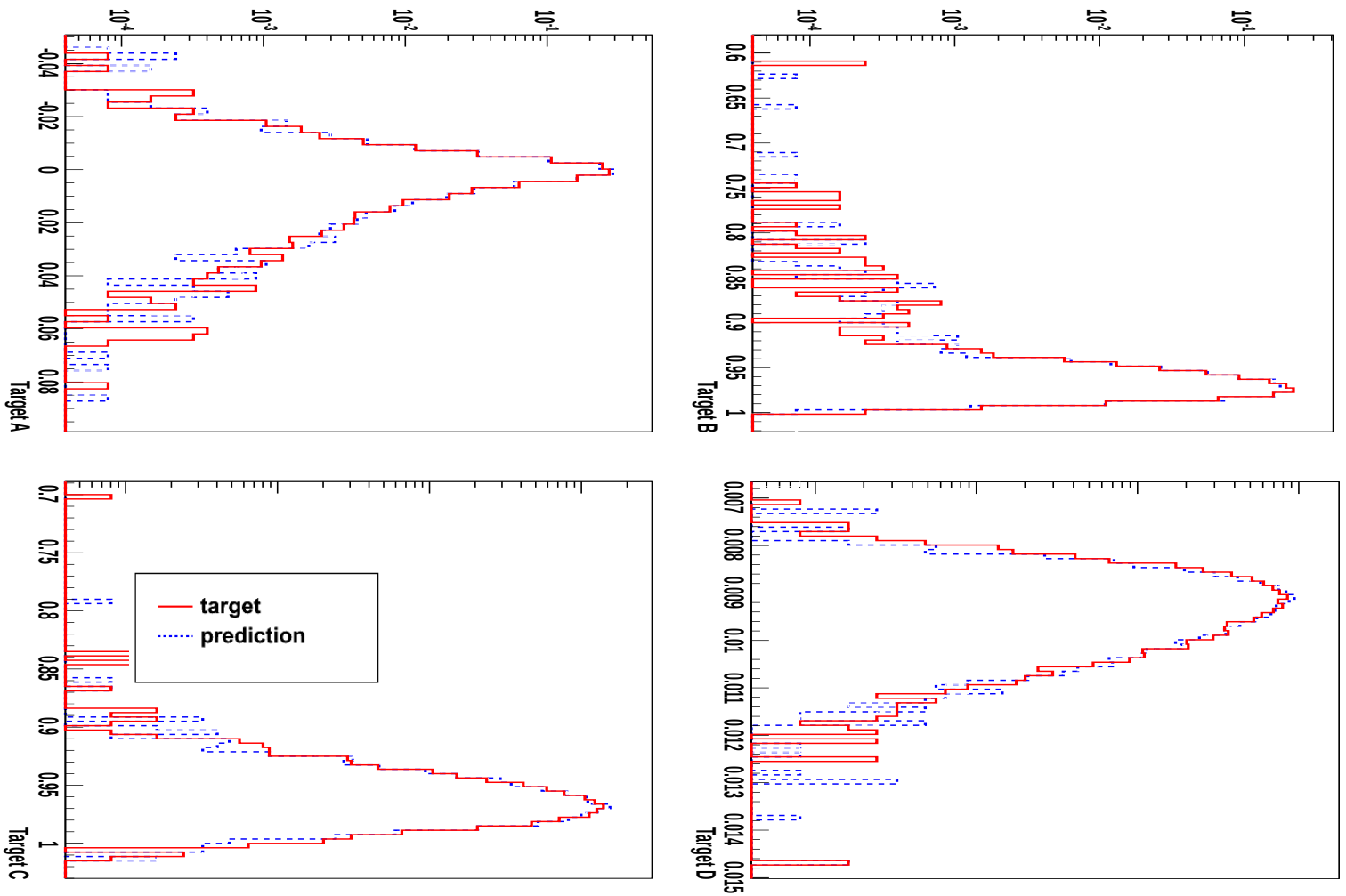
- 4 of them strongly correlated

14 target outputs to estimate {A, B, C, D…}

- 4 of them strongly correlated

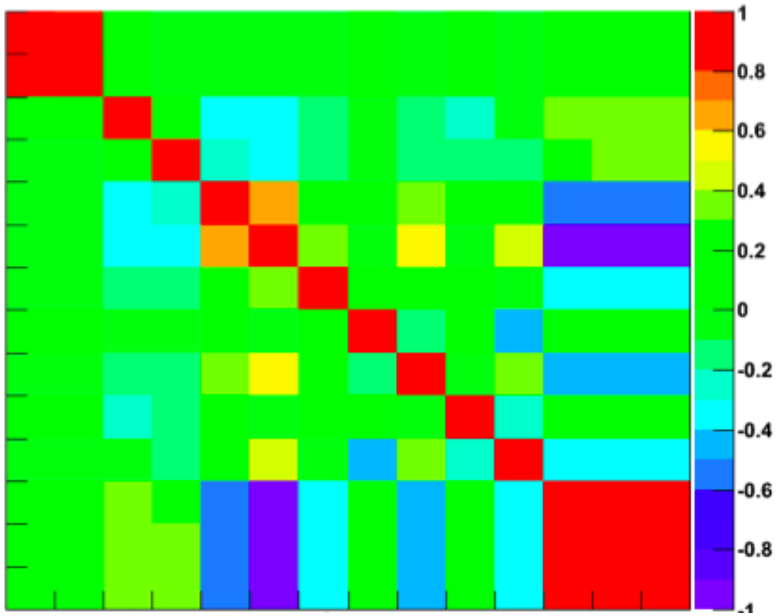
Challenge: build a predictive model to describe simultaneously all the outputs {A,B,C,D…}, provided a corresponding set of inputs.

Illustrative Example

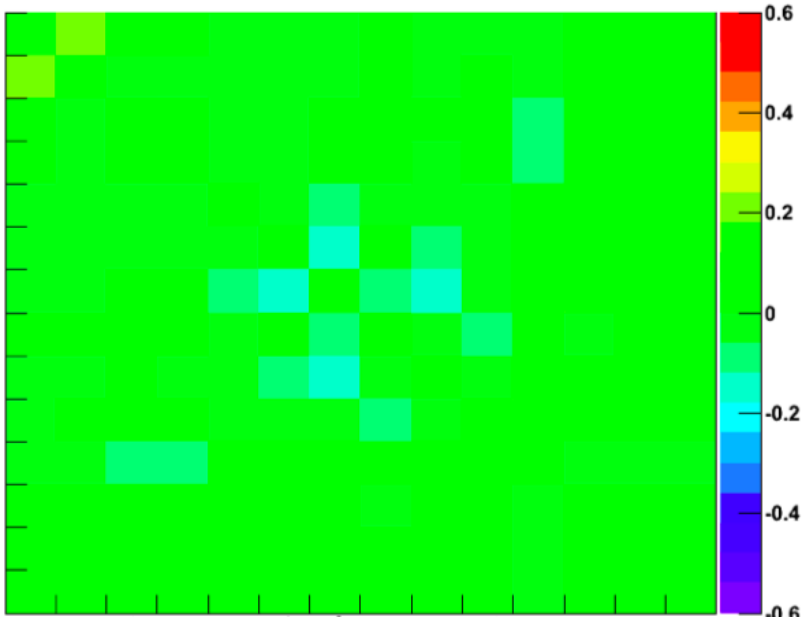


Illustrative Example

Target Correlations



Prediction-Target Difference

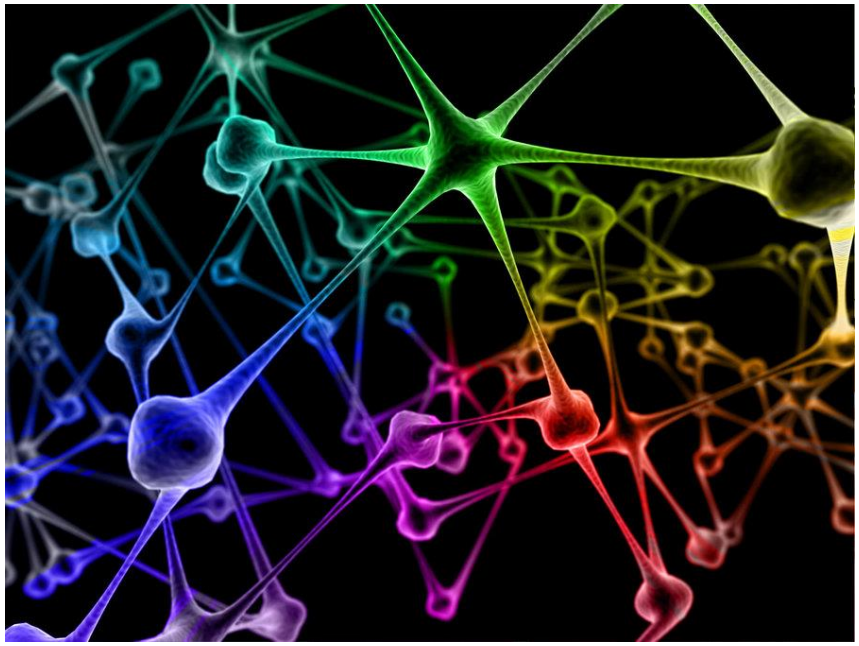


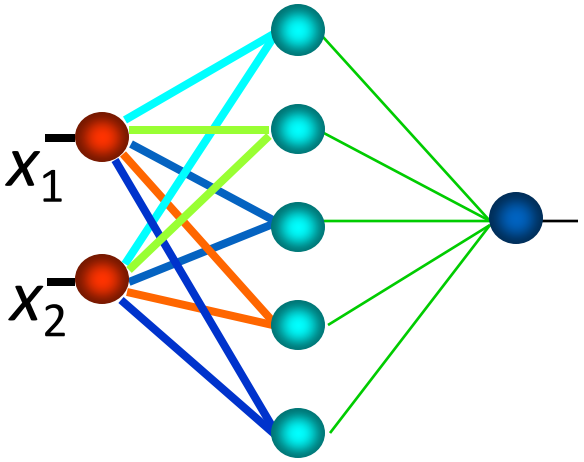
Very close to Zero

Outline

- **Neural Networks**
- **Shallow Learning**

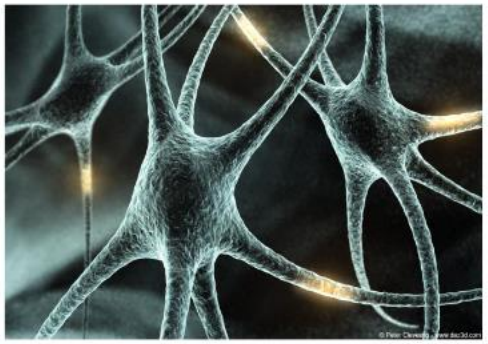
Neural Networks





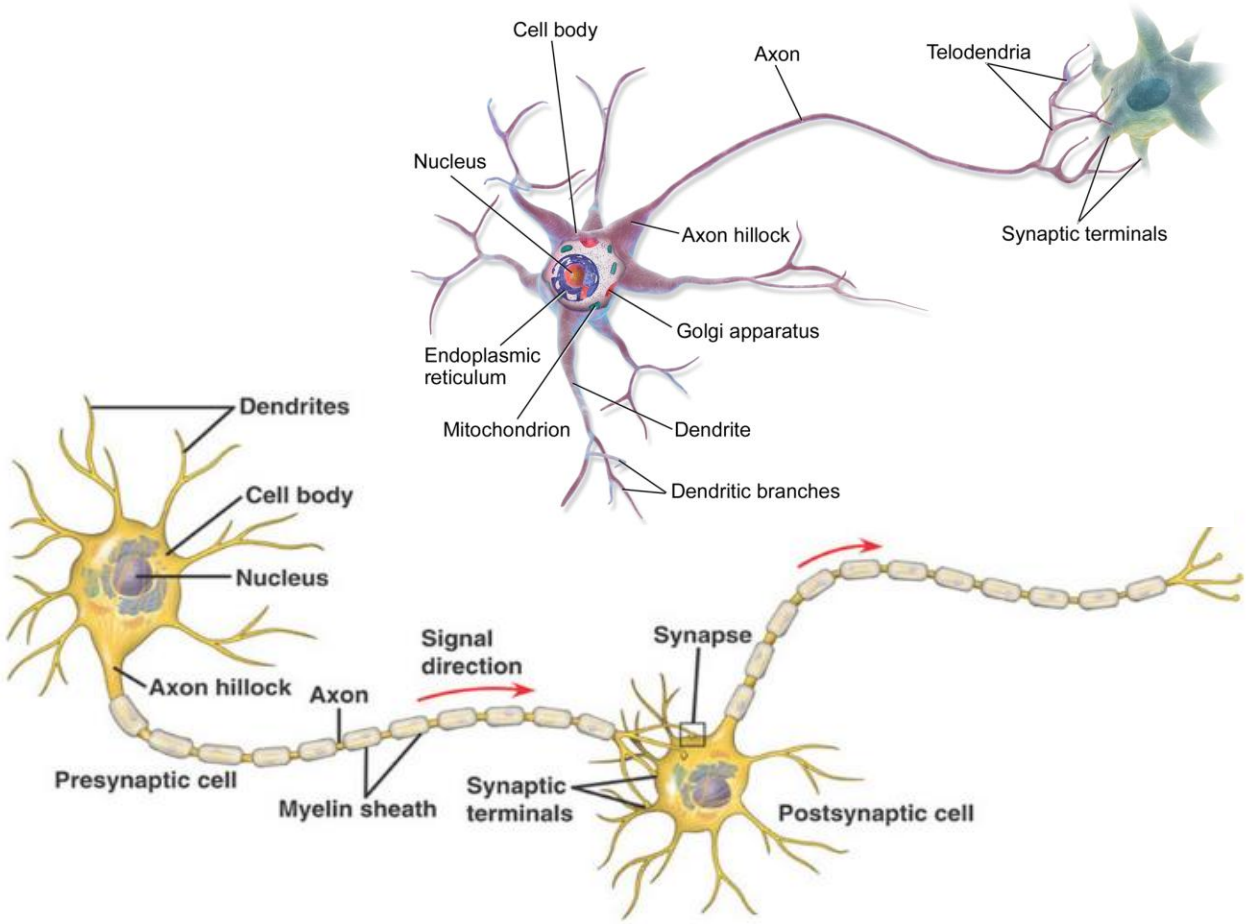
Neural Networks

Biological systems built of very complex webs of interconnected neurons



- Highly connected to other neurons
- Performs computations by combining signals from other neurons
- Outputs of these computations may be transmitted to one or more other neurons

Real Neurons



Neural Network

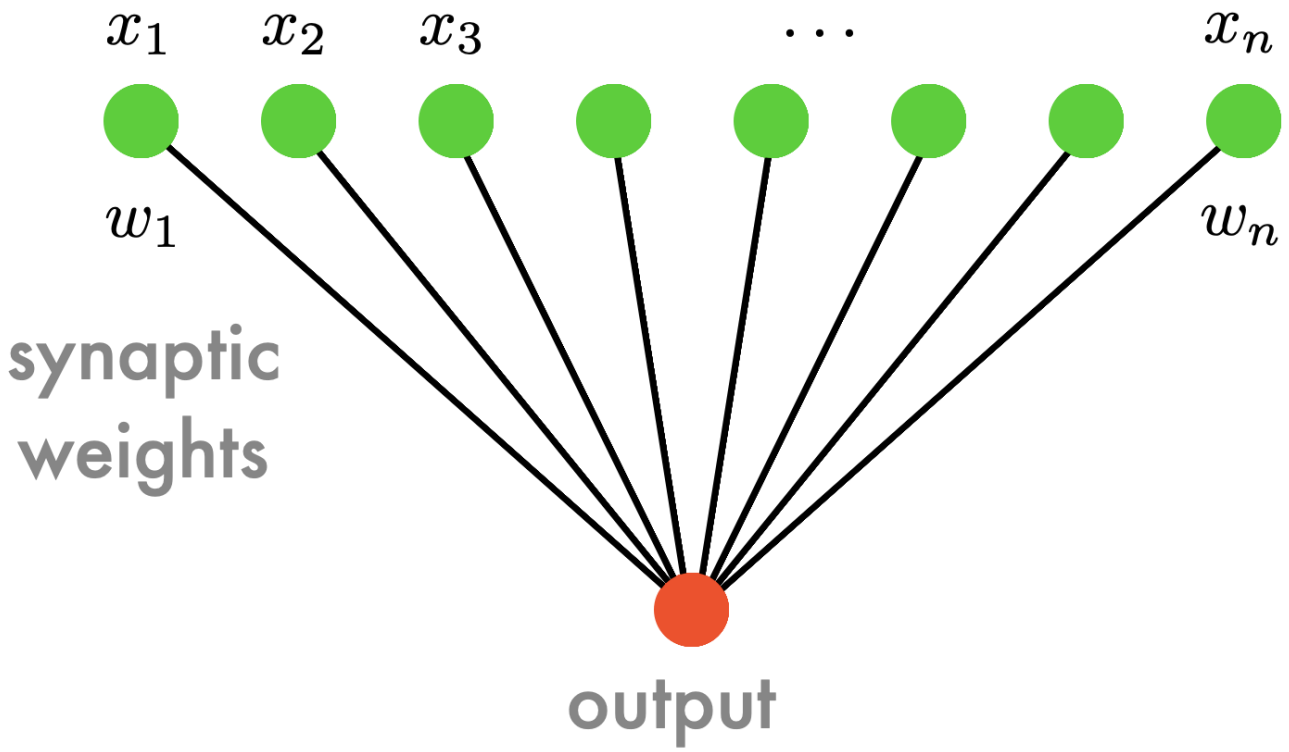
Humans:

- Neuron switching time ~0.001s
- Number of neurons ~ 10^{10}
- Connections per neuron ~10^{4}
- Scene recognition time ~ 0.1s

ANNs:

- Many **threshold switching** units
- Many **weighted interconnections**
- Highly **parallel, distributed** processes

Perceptron



Frank Rosenblatt, 1957

Perceptron Learning

Perceptrons

- **Threshold Logic Unit/Step Function**
  - Linear combination of inputs
  - Classify above threshold
- **Hebbian Learning Rule:**
  - “Fire together, wire together”
- **Linear decision boundary**
  - XoR Classification Problem Minsky and Papert 1969
- **Stack into MultiLayer Perceptrons (MLPs)**