Outline

• Biological inspiration
• A mathematical definition of learning
• Supervised, unsupervised, semi-supervised
• Example: Decision tree
• Example: K-nearest neighbors
The Giant Squid Axon

• 1909: Williams describes the giant squid axon (III: 1mm thick)
• 1939: Young describes the synapse.
• 1952: Hodgkin & Huxley publish an electrical current model for the generation of binary action potentials from real-valued inputs.

Image released to the public domain by Ikkisan, 2007.
In 1943, McCulloch & Pitts proposed that biological neurons have a nonlinear activation function (a step function) whose input is a weighted linear combination of the currents generated by other neurons.

They showed lots of examples of mathematical and logical functions that could be computed using networks of simple neurons like this.
Biological Inspiration: Hodgkin & Huxley

Hodgkin & Huxley won the Nobel prize for their model of cell membranes, which provided lots more detail about how the McCulloch-Pitts model works in nature. Their nonlinear model has two step functions:

- $I < \text{threshold1}$: $V = -75\,mV$
- threshold1 < $I < \text{threshold2}$: $V$ has a spike, then returns to rest.
- threshold 2 < $I$: $V$ spikes periodically

![Hodgkin & Huxley Circuit Model of a Neuron Membrane](https://commons.wikimedia.org/w/index.php?curid=21725464)

Membrane voltage versus time. As current passes 0mA, spike appears. As current passes 10mA, spike train appears.

![Membrane voltage versus time](https://commons.wikimedia.org/w/index.php?curid=30310965)
Biological inspiration: Hebbian learning

“Neurons that fire together, wire together.

... The general idea is an old one, that any two cells or systems of cells that are repeatedly active at the same time will tend to become `associated’ so that activity in one facilitates activity in the other.”

- D.O. Hebb, 1949
Biological inspiration: Long-term potentiation

Figures this page are public domain, by Thomas W. Sulcer, 2011

1. A synapse is repeatedly stimulated

2. More dendritic receptors

3. More neurotransmitters

4. A stronger link between neurons
Kohonen’s computational model of Hebbian learning: the self-organizing map

- Blue blob = distribution of training data
- White disc = training sample drawn from that distribution
- Black grid = positions coded by 25 neurons
- Yellow highlight = neuron closest to the training sample is updated to move closer to the sample
- After many data presentations, the neural map matches the training distribution.

CC-SA-3.0, MCLD, 2010
Biological Inspiration: Simple, Complex, and Hypercomplex Cells in the Visual Cortex

D. Hubel and T. Wiesel (1959, 1962, Nobel Prize 1981) found that the human visual cortex consists of a hierarchy of *simple*, *complex*, and *hypercomplex* cells.

- Simple cells (in visual area 1, called V1) fire when you see a simple pattern of colors in a particular orientation (figure (b), at right)

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
Biological Inspiration: Simple, Complex, and Hypercomplex Cells in the Visual Cortex

D. Hubel and T. Wiesel (1959, 1962, Nobel Prize 1981) found that the human visual cortex consists of a hierarchy of simple, complex, and hypercomplex cells.

• Complex cells are sensitive to moving stimuli of a particular orientation traveling in a particular direction (figure (d) at right).

• Complex cells can be modeled as linear combinations of simple cells!
Biological Inspiration: Simple, Complex, and Hypercomplex Cells in the Visual Cortex

D. Hubel and T. Wiesel (1959, 1962, Nobel Prize 1981) found that the human visual cortex consists of a hierarchy of simple, complex, and hypercomplex cells.

• Hypercomplex cells are sensitive to moving stimuli of a particular orientation traveling in a particular direction, and they also stop firing if the stimulus gets too long.

• Hypercomplex cells can be modeled as linear combinations of complex cells!
Outline

• Biological inspiration
• A mathematical definition of learning
• Supervised, unsupervised, semi-supervised
• Example: Decision tree
• Example: K-nearest neighbors
Learning: learn a function $\hat{y} = f(x)$, where $x$=features, $y$=true label, $\hat{y}$=estimated label.

Features ($x$)

Class label ($y$)

- Zebra
- Giraffe
- Hippopotamus
A mathematical definition of learning

- **Environment**: there are two random variables, $x \sim X$ and $y \sim Y$, that are jointly distributed according to
  \[ P(X = x, Y = y) \]

- **Data**: $P(X, Y)$ is unknown, but we have a sample of training data
  \[ D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \]

- **Objective**: We would like a function $f$ such that $f(x) \approx y$

- **Definition of learning**: Learning is the task of estimating the function $f$, given knowledge of nothing other than $D$. 
Outline

• Biological inspiration
• A mathematical definition of learning
• Supervised, unsupervised, semi-supervised
• Example: Decision tree
• Example: K-nearest neighbors
Supervised, Unsupervised & Semi-Supervised learning

• **Supervised learning**: \( \mathcal{D} = \{(x_1, y_1), ..., (x_n, y_n)\} \). Learn \( f \).

• **Unsupervised learning**: \( \mathcal{D} = \{x_1, ..., x_n\} \). Learn \( P(X = x) \).

• **Semi-supervised**: \( \mathcal{D} = \{(x_1, y_1), ..., (x_m, y_m), x_{m+1}, ..., x_n\} \), where \( m \ll n \). Learn \( f \) as well as possible.
Supervised learning examples

• Visual object detection: $X=$pixels of an image, $Y=$object label
• Automatic speech recognition: $X=$audio signal, $Y=$word sequence
• Sentiment detection: $X=$text of a tweet, $Y=$emotion label
• Electrocardiography: $X=$electrocardiogram, $Y=$detection of irregularity
• Self-driving car: $X=$visual+LIDAR, $Y=$positions and velocities of every obstacle
Unsupervised learning examples

• Distribution learning: estimate $P(X = x)$
• Support learning: estimate the domain, i.e., the set of $x$ for which $P(X = x) \neq 0$
• Clustering: Group the data into clusters of objects that are similar, according to some similarity criterion
• Manifold estimation: Given a high-dimensional feature vector $x$ (e.g., 100,000 dimensions), find a low-dimensional vector (e.g., 300 dimensions) that captures most of the differences among tokens
Example: clustering

DBSCAN density-based clustering. RBG = three clusters, gray = tokens that could not be assigned to a cluster. CA-SA 3.0, Chire 2011

OPTICS is a DBSCAN variant, improving handling of different density clusters. CA-SA 3.0, Chire 2011
Semi-supervised learning examples

• Distribution learning:
  • Learn \( P(X = x) \) from unlabeled data
  • Learn \( P(X = x|Y = y) \) from labeled data
  • Combine them to compute \( P(Y = y|X = x) \)

• Clustering:
  • Learn the clusters from unlabeled data
  • Use the labeled data to label at least one example per cluster
  • Combine them to estimate \( f(X) \)
Outline

• Biological inspiration
• A mathematical definition of learning
• Supervised, unsupervised, semi-supervised, self-supervised
  • Example: Decision tree
  • Example: K-nearest neighbors
Decision tree learning: An example

• The Titanic sank.
• You were rescued.
• You want to know if your friend was also rescued.
• You can’t find them.
• Can you use machine learning methods to estimate the probability that your friend survived?
Survival of the Titanic: A machine learning approach

1. Gather data about as many of the passengers as you can.
   • \( X \) = variables that describe the passenger, e.g., age, gender, number of siblings on board.
   • \( Y = 1 \) if the person is known to have survived
2. Learn a function, \( f(X) \), that matches the known data as well as possible
3. Apply \( f(x) \) to your friend’s facts, to estimate their probability of survival
Survival of the Titanic: A machine learning approach

Decision-tree learning:
- 1\textsuperscript{st} branch = variable that best distinguishes between groups with higher vs. lower survival rates (e.g., gender)
- 2\textsuperscript{nd} branch = variable that best subdivides the remaining group
- Quit when all people in a group have the same outcome, or when the group is too small to be reliably subdivided.
Survival of the Titanic: A machine learning approach

In each leaf node of this tree:

• Number on the left = probability of survival
• Number on the right = percentage of all known cases that are explained by this node

CC-SA 4.0, Gilgoldm, 2020
Another example with more detail: Complications of spinal surgery

- $X = (\text{age}, \text{starting vertebra number})$
- $Y = 1$ (red) if complications, $Y = 0$ (green) if not
- Figure on the left shows the tree.
- Figures on the right: the tree divides up the 2D input space ($X$) into regions of lower and higher probability of complications
Outline

• Biological inspiration
• A mathematical definition of learning
• Supervised, unsupervised, semi-supervised, self-supervised
• Example: Decision tree
• Example: K-nearest neighbors
Classifier example: dogs versus cats
Can you write a program that can tell which ones are dogs, and which ones are cats?
Nearest Neighbors Classifier

• Given n different training images. Each one has a known class label.

• Input to the classifier: a test image \( x \) whose correct label is unknown.

• Classification function:
  1. Find the training token, \( x_i \), that is most similar to the test token.
  2. Find out the corresponding class label, \( y_i = \text{correct\_label}(x_i) \).
  3. Output \( y_i \) as the best guess for the label of test token \( x \).
Example of Nearest-Neighbor Classification

Test Token: Maltese

This is the most similar training token...

Therefore the Maltese is classified as a dog.
K-Nearest Neighbors (KNN) Classifier

The nearest-neighbors classifier sometimes fails if one of the training tokens is unusual. In that case, a test token that is similar to the weird training token might get misclassified. Solution: K-Nearest Neighbors.

Test token:

Most similar training token:
K-Nearest Neighbors Classification Function

1. Find the K training tokens, \( x_i \), that are most similar to the test token (K is a number chosen in advance by the system designer, e.g., \( K = 3 \)).
2. Find out the corresponding class labels, \( y_i = \text{correct_label}(x_i) \).
3. Vote! Find the class label that is most frequent among the K-nearest neighbors, and output that as the label of the test token.

Test token: 3 most similar training tokens:
Outline

• Biological inspiration: “Neurons that fire together, wire together.”
• A mathematical definition of learning
  Given a training dataset $\mathcal{D} = \{(x_1, y_1), ..., (x_n, y_n)\}$, estimate a function $f(x)$ such that, for all $(x, y) \sim (X, Y)$, $f(x) \approx y$
• Supervised, unsupervised, semi-supervised learning
• Example: Decision tree
• Example: K-nearest neighbors