

# Artificial Intelligence and Machine Learning

## + Other Statistical Methods

Grinnell College

Spring 2025

# Artificial Intelligence

**Artificial intelligence** (AI) refers to the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making.

It is a field of research in computer science that develops and studies methods and software that enable machines to [...] use learning and intelligence to take actions that maximize their chances of achieving defined goals. <sup>1</sup>

## Examples

- ▶ Checkers / Chess / Go
- ▶ Search engines
- ▶ Streaming recommendations
- ▶ Image labeling
- ▶ Fraud detection (finance)

---

<sup>1</sup>Sindhu V, Nivedha S, Prakash M (February 2020). "An Empirical Science Research on Bioinformatics in Machine Learning". Journal of Mechanics of Continua and Mathematical Sciences (7). doi:10.26782/jmcms.spl.7/2020.02.00006

# Machine Learning

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions. <sup>2</sup>

Many of the methods used in machine learning employ heavy use of statistics and also mathematical optimization (specifying a function and trying to find values that maximize or minimize it)

## Examples

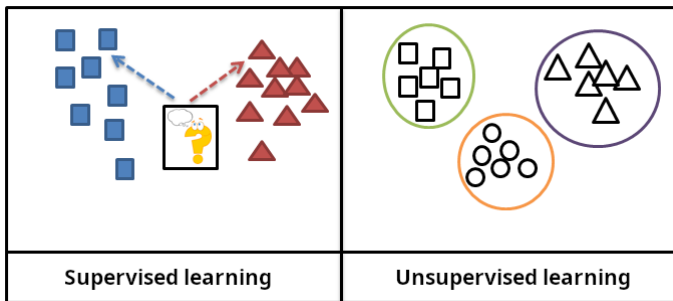
- ▶ estimating parameter values (Maximum Likelihood Estimation)
- ▶ finding variables to make predictions (minimize prediction error)

---

<sup>2</sup>Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming". Artificial Intelligence in Design '96. Dordrecht, Netherlands: Springer Netherlands. pp. 151–170.

# Types of Machine Learning

In ML there is usually a distinction made between Supervised and Unsupervised learning. This pretty much boils down to whether or not the machine knows what the actual answers should be



## Common uses (not exhaustive)

- ▶ Supervised: prediction and classification
- ▶ Unsupervised: clustering and grouping

# Types of Machine Learning

## **Supervised Learning**

Model is trained on data with expected answers given. The challenge is then finding out how to get to those answers. We might use a big dataset with vectors of variables and values (Explanatory and Response)

In supervised learning, the algorithm “learns” from the training data set by iteratively making predictions on the data and adjusting for the correct answer.

## **Example**

Supervised learning model can predict how long your commute will be based on the time of day, weather conditions and so on. But first, you must train it to know that rainy weather extends the driving time.

# Types of Machine Learning

## Unsupervised Learning

Using machine learning algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns in data without the need for human intervention.

**Note:** These still require some human intervention for validating output variables

### Example

An unsupervised learning model can identify that online shoppers often purchase groups of products at the same time. However, a data analyst would need to validate that it makes sense for a recommendation engine to group baby clothes with an order of diapers, applesauce, and sippy cups.

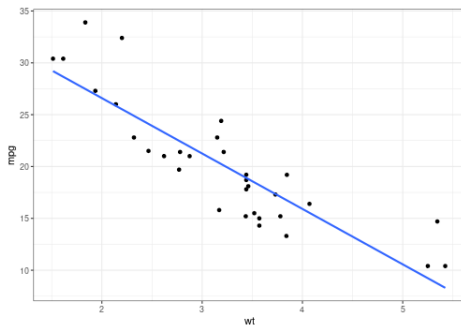
# Supervised – Simple Linear Regression

Regression was actually a form of Machine Learning!

- ▶ We told R we wanted to predict some variables using others, told it we wanted a line, and it found out which line to use

**Example:** Predicting vehicle MPG using weight (tons)

$$\hat{y} = 37.285 - 5.34 \times \text{Weight}$$

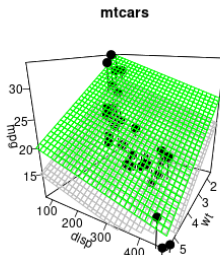
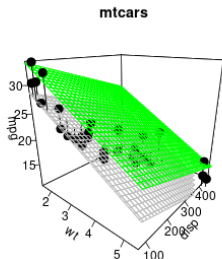


# Supervised – Multiple Regression

We can actually use any number of quantitative or categorical variables in a regression equation

- ▶ idea: use more variables, get better predictions!
- ▶ cons: gets really tough to work with and interpret

$$\hat{y} = 34.67 - 3.27 \times \text{weight} - 0.018 \times \text{displacement} + 0.17 \times \mathbb{1}_{\text{Manual}}$$



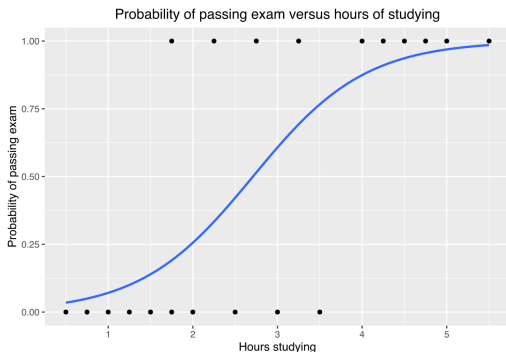


# Supervised – Logistic Regression

We can modify regression to make predictions to predict the probability an observation belongs to 1 of 2 groups

- ▶ regression equation predicts a probability
- ▶ whichever ever groups is more likely  $\rightarrow$  predict that one

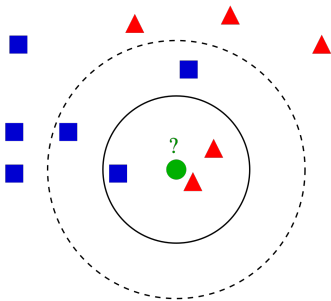
$$\hat{p}_i = \frac{1}{1 + e^{-b_0 - b_1 X_1 - \dots - b_m X_m}}$$



# Supervised: k-Nearest-Neighbors

Another supervised classification (kNN) classifies a new data point according to what similar points are in the data set.

- ▶ k-nearest-neighbors → use the nearest k data points

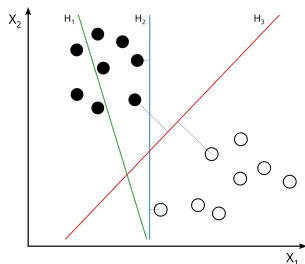


## Number of neighbors affects classification

- ▶  $k = 3$  (solid black line) → red triangle
- ▶  $k = 5$  (dashed black line) → blue square

# Supervised: Support Vector Machine

SVMs are a method of classification that tries to find a linear separation between groups, then use that to classify data points



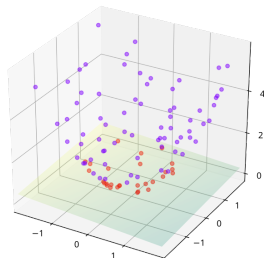
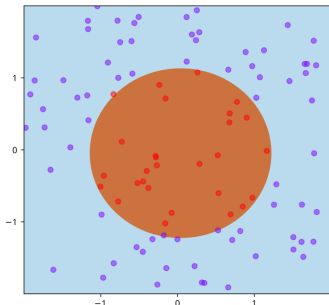
Basic idea: find the line that separates the groups that gives us the widest 'margin' between the groups

- ▶ can be generalized to more than 2 dimensions (lots of variables)

# Supervised: Support Vector Machine

Sometimes groups are not 'linearly' separable. A solution might be to map the data to a higher dimensional representation, where the points can be separated (kernel trick)

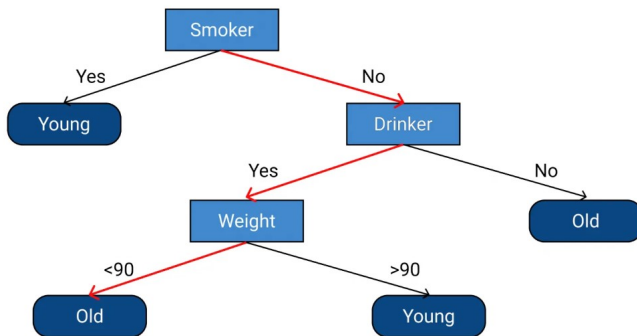
- ▶ this gives us a separating 'plane' instead of a line



# Supervised: Decision Trees

Decision Trees are classification methods that give us a tree of choices with separations based on explanatory variables.

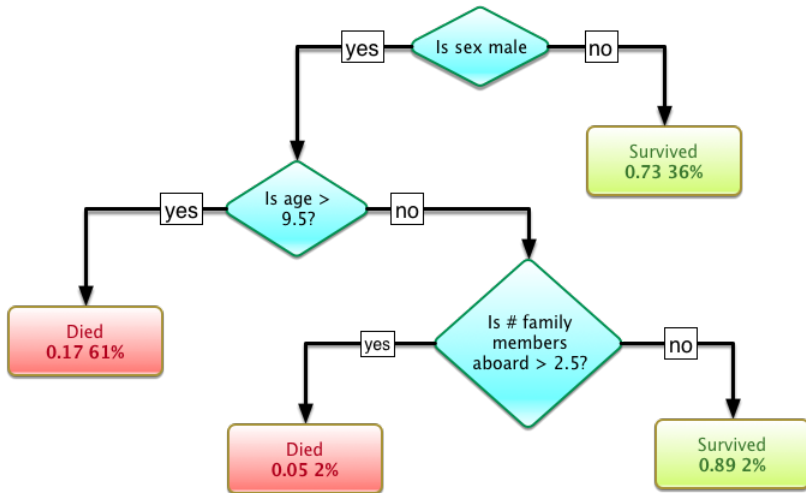
**Example**<sup>3</sup> Smoking, drinking, and weight could be decent predictors of whether someone dies young. Define "Young" as age of death less than 70



<sup>3</sup><https://www.upgrad.com/blog/decision-tree-in-machine-learning/>

# Another Decision Tree Example

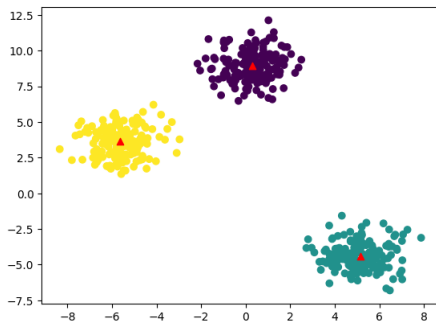
## Titanic Survival



# Unsupervised: k-Means Clustering

Suppose we have a bunch of points scattered around and we wish to find ways to separate them. We can start by telling the machine how many groups we want.

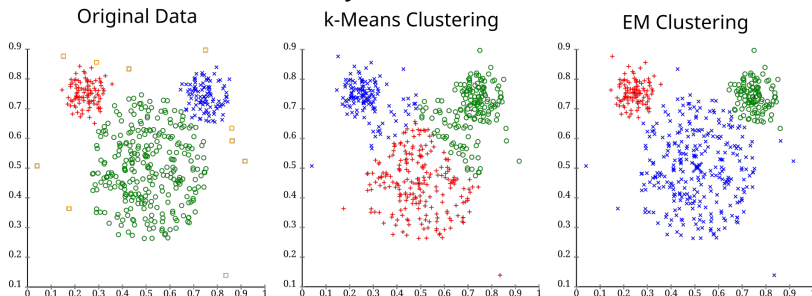
- ▶ specify the number of groups
- ▶ method: pick  $k$  points that minimize the sum of squared (Euclidean) distances of points from their nearest group



# Unsupervised: k-Means Clustering

- ▶ algorithm actually picks  $k$  points at random within the range of data and iterates trying to find improvements
- ▶ tends toward equally sized clusters → not always helpful

Different cluster analysis results on "mouse" data set:

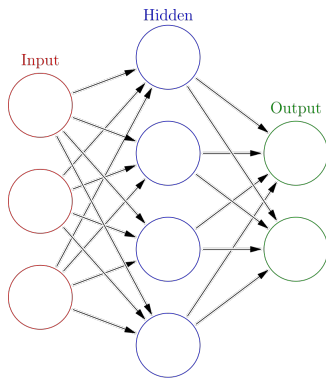




# Neural Network

An (artificial) neural network (ANN) is a machine learning program that makes decisions in a manner similar to the human brain, by using processes that mimic the way biological neurons work together to identify phenomena, weigh options and arrive at conclusions.

- ▶ composed of layers of nodes (neurons): inputs, hidden connections, and an output
- ▶ nodes are connected to each other (like neurons in brain)
- ▶ if output of a node is above some threshold → node activates and sends data to nodes further in the chain

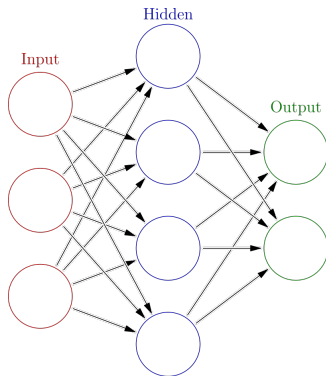


# Neural Networks

## How do they work?

We can treat each node in the middle as its own regression model. The input nodes all factor into whether or not this node fires.

- ▶ For each node connection to the previous layer, assign weights to the signals being sent through, add them all up
  - ▶ weights determine how important each previous node is in determining whether the current one activates
- ▶ resulting sum is passed through an 'activation function' that determines whether the node continues firing to other nodes



## How do they work?

It basically boils down to a big, complicated linear regression!

- ▶ have to choose inputs
- ▶ for classification we will have just one output: firing or not firing will tell us the classification
- ▶ each node has a number of weights (slopes) according to the number of previous nodes
- ▶ determining values for weights basically is like finding lots of slopes in many regressions, and so requires A LOT of data

# Neural Networks

## Activation Functions

There are different activation functions that are used to determine whether a node fires.

- ▶ logistic / sigmoid (like in logistic regression)
- ▶ linear (above a certain value)
- ▶ hyperbolic tangent
- ▶ Gaussian (Normal)

## Training

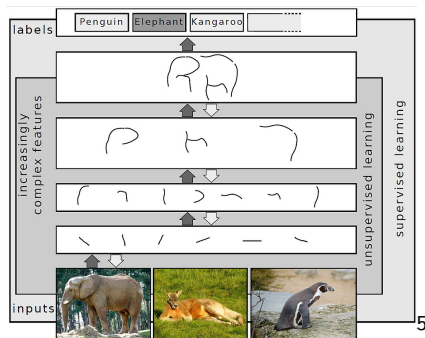
"As we start to think about more practical use cases for neural networks, like image recognition or classification, we'll leverage supervised learning, or labeled datasets, to train the algorithm. As we train the model, we'll want to evaluate its accuracy using a cost function."

- ▶ we will use the mean-squared-error (avg. of errors<sup>2</sup>)
- ▶ repeatedly run data through, adjust weights and activation thresholds so that we get less error (at a certain point we stop improving)

# Deep Learning

**Deep Learning** is the application of neural networks that utilize multiple layers of neurons instead of just 1 hidden layer. Harder to train, but we are able to get more flexibility and the nodes 'learn' to pick up certain features which help prediction/classification

- ▶ in theory: more nodes  $\rightarrow$  find more features



<sup>5</sup>Source: <https://commons.wikimedia.org/w/index.php?curid=82466022>

# Large Language Models

"LLMs are designed to understand and generate text like a human, in addition to other forms of content, based on the vast amount of data used to train them. They have the ability to infer from context, generate coherent and contextually relevant responses, translate to languages other than English, summarize text, answer questions (general conversation and FAQs) and even assist in creative writing or code generation tasks."

"They are able to do this thanks to billions of parameters that enable them to capture intricate patterns in language and perform a wide array of language-related tasks."

# Large Language Models

"LLMs operate by leveraging deep learning techniques and vast amounts of textual data. [...] During the training process, these models learn to predict the next word in a sentence based on the context provided by the preceding words. The model does this through attributing a probability score to the recurrence of words [...]."

"To ensure accuracy, this process involves training the LLM on a massive corpora of text (in the billions of pages), allowing it to learn grammar, semantics and conceptual relationships through [...] self-supervised learning. Once trained on this training data, LLMs can generate text by autonomously predicting the next word based on the input they receive, and drawing on the patterns and knowledge they've acquired."

# Generative Pre-Trained Transformer

**Generative Pre-Trained Transformers** are a type of LLM. They are made out of something called "transformers," which converts text to number representations (tokens), and then are run through long chains of multi-layered neural networks that prioritize words from the input, and try to associate context to them based training from the material used for the model.



## Other Sources:

### **Supervised vs. Unsupervised Learning**

<https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning>

### **k-Means Clustering**

<https://www.geeksforgeeks.org/k-means-clustering-introduction/>

**Neural Networks** <https://www.ibm.com/think/topics/neural-networks>