MACHINE LEARNING ALGORITHMS

- In 1959, Arthur Samuel defined **machine learning** as a "Field of study that gives computers the ability to learn without being explicitly programmed“
- **Machine learning** explores the study and construction of **algorithms** that can learn from and make predictions on data

Hypothesis testing

A **hypothesis test** is a statistical **test** that is used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population

Elements of a hypothesis test:

Null hypothesis - Statement regarding the value(s) of unknown parameter(s). Typically will imply no association between explanatory and response variables in our applications (will always contain an equality)

Alternative hypothesis - Statement contradictory to the null hypothesis (will always contain an inequality)

Test statistic - Quantity based on sample data and null hypothesis used to test between null and alternative hypotheses

Rejection region - Values of the test statistic for which we reject the null in favor of the alternative hypothesis

Example - Efficacy Test for New drug

- Drug company has new drug, wishes to compare it with current standard treatment
- Federal regulators tell company that they must demonstrate that new drug is better than current treatment to receive approval
- Firm runs clinical trial where some patients receive new drug, and others receive standard treatment
- Numeric response of therapeutic effect is obtained (higher scores are better).
- Parameter of interest: $m_{New} - m_{Std}$

Example - Efficacy Test for New drug

Null hypothesis - New drug is no better than standard trt

$H0: \mu new - \mu std \leq 0 \quad (\mu new - \mu std = 0)$

Alternative hypothesis - New drug is better than standard trt

$HA : \mu new - \mu std > 0$

Types of Machine Learning

Machine learning tasks are typically classified into three broad categories, depending on the nature of the learning "signal" or "feedback" available to a learning system. These are:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

- **Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs**
- **Supervised learning** is the machine **learning** task of inferring a function from labeled training data.

- **Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end.**
- **Unsupervised learning** is the machine learning task of inferring a function to describe hidden structure from unlabeled data

f

Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle), without a teacher explicitly telling it whether it has come close to its goal or not. Another example is learning to play a game by playing against an opponent.

Between supervised and unsupervised learning is semi-supervised learning, where the teacher gives an incomplete training signal: a training set with some (often many) of the target outputs missing. Transduction is a special case of this principle where the entire set of problem instances is known at learning time, except that part of the targets is missing.

•

- A support vector machine is a classifier that divides its input space into two regions, separated by a linear boundary. Here, it has learned to distinguish black and white circles.
- Among other categories of machine learning problems, learning to learn learns its own inductive bias based on previous experience.