**A Quick Introduction to Supervised vs. Unsupervised Learning**

The field of machine learning contains a massive set of algorithms that can be used for understanding data. These algorithms can be classified into one of two categories:

**1. Supervised Learning Algorithms:** Involves building a model to estimate or predict an output based on one or more inputs.

**2. Unsupervised Learning Algorithms:** Involves finding structure and relationships from inputs. There is no “supervising” output.

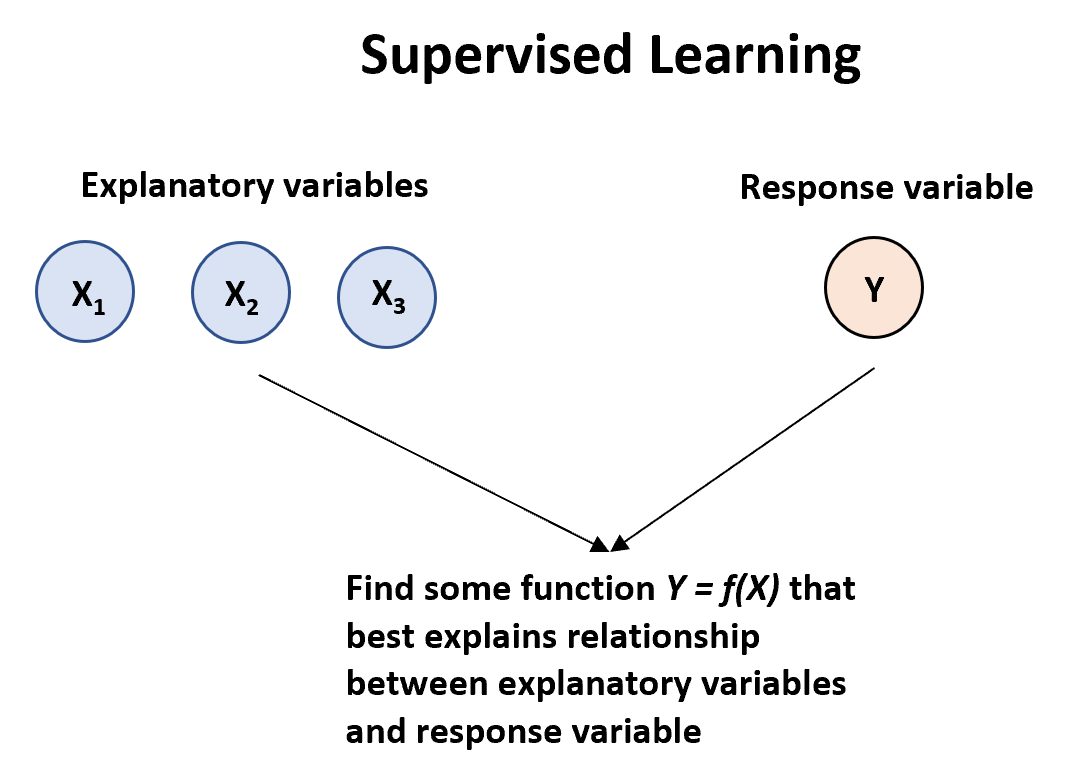
This tutorial explains the difference between these two types of algorithms along with several examples of each.

**Supervised Learning Algorithms**

A **supervised learning algorithm** can be used when we have one or more explanatory variables (X1, X2, X3, …, Xp) and a [response variable](https://www.statology.org/explanatory-response-variables/) (Y) and we would like to find some function that describes the relationship between the explanatory variables and the response variable:

**Y = *f*(X) + ε**

where *f* represents systematic information that X provides about Y and where ε is a random error term independent of X with a mean of zero.



There are two main types of supervised learning algorithms:

**1. Regression:** The output variable is continuous (e.g. weight, height, time, etc.)

To understand the association between an employee’s age and education, as well as the calendar year, on his wage.

The Wage data involves predicting a continuous or quantitative output value.

**2. Classification:** The output variable is categorical (e.g. male or female, pass or fail, benign or malignant, etc.) non-numerical

Predicting whether a given day’s stock market performance will fall into the Up bucket or the Down bucket.

The inputs go by different names, such as *predictors*, *independent variables*, *features*, or sometimes just *variables*.

The output variable in previous example, sales is variable often called the *response* or *dependent variable.*

There are two main reasons that we use supervised learning algorithms:

1. **Prediction:** We often use a set of explanatory variables to predict the value of some response variable

* *square footage* and *number of bedrooms* to predict *house price*

For instance, consider a company that is interested in conducting a direct-marketing campaign.

* The goal is to identify individuals who will respond positively to a mailing, based on observations of demographic variables measured on each individual.

In this case, the demographic variables serve as predictors, and response to the marketing campaign (either positive or negative) serves as the outcome.

The company is not interested in obtaining a deep understanding of the relationships between each individual predictor and the response; instead, the company simply wants an accurate model to predict the response using the predictors. This is an example of modeling for prediction.

1. **Inference:** We may be interested in understanding the way that a response variable is affected as the value of the explanatory variables change.

* how much does home price increase, on average, when the number of bedrooms increases by one

We instead want to understand the relationship between X and Y , or more specifically, to understand how Y changes as a function of X1, . . .,Xp.

* F(X) cannot be treated as a black box, because we need to know its exact form.
  + Which predictors are associated with the response?
  + What is the relationship between the response and each predictor?
  + Can the relationship between Y and each predictor be adequately summarized using a linear equation, or is the relationship more complicated?
* Consider the Advertising data, One may be interested in answering questions such as:

– Which media contribute to sales?

– Which media generate the biggest boost in sales? or

– How much increase in sales is associated with a given increase in TV advertising?

* The brand of a product that a customer might purchase based on variables such as price, store location, discount levels, competition price, and so forth.

– In this situation one is interested in how each of the individual variables affects the probability of purchase.

– For instance, what effect will changing the price of a product have on sales?

Finally, some modeling could be conducted both for prediction and inference. For example, in a real estate setting, one may seek to relate values of homes to inputs such as crime rate, zoning, distance from a river, air quality, schools, income level of community, size of houses, and so forth.

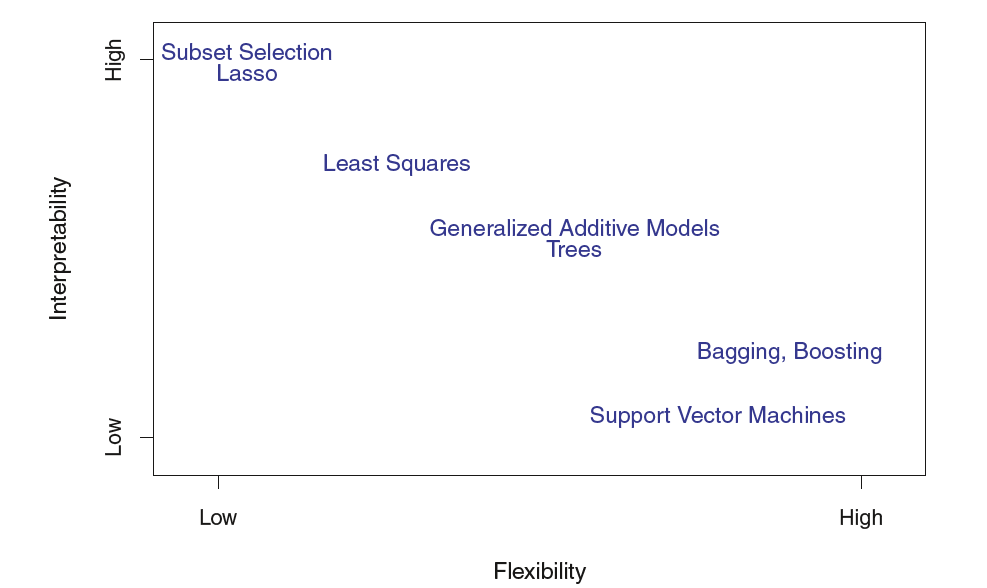
* In this case one might be interested in how the individual input variables affect the prices—that is, how much extra will a house be worth if it has a view of the river? This is an inference problem.
* Alternatively, one may simply be interested in predicting the value of a home given its characteristics: is this house under- or over-valued? This is a prediction problem.

Depending on whether our goal is inference or prediction (or a mix of both), we may use different methods for estimating the function *f*. For example, linear models offer easier interpretation but non-linear models that are difficult to interpret may offer more accurate prediction.

Here is a list of the most commonly used supervised learning algorithms:

* Linear regression
* Logistic regression
* Linear discriminant analysis
* Quadratic discriminant analysis
* Decision trees
* Naive bayes
* Support vector machines
* Neural networks

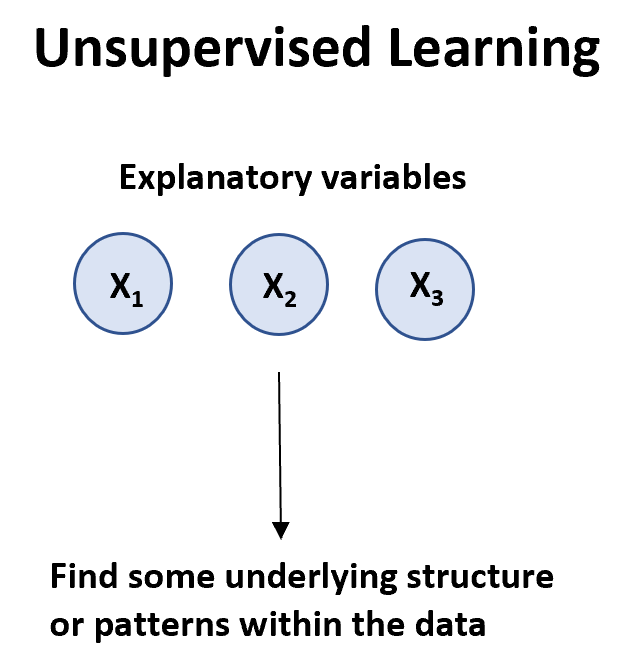
The Trade-Off Between Prediction Accuracy and Model Interpretability



* A representation of the tradeoff between flexibility and interpretability, using different statistical learning methods.
* In general, as the flexibility of a method increases, its interpretability decreases.
* when inference is the goal, there are clear advantages to using simple and relatively inflexible statistical learning methods.
* When interested in prediction, and the interpretability of the predictive model is simply not of interest.

**Unsupervised Learning Algorithms**

An **unsupervised learning algorithm** can be used when we have a list of variables (X1, X2, X3, …, Xp) and we would simply like to find underlying structure or patterns within the data.



There are two main types of unsupervised learning algorithms:

**1. Clustering:** Using these types of algorithms, we attempt to find “clusters” of [observations](https://www.statology.org/observation-in-statistics/) in a dataset that are similar to each other. This is often used in retail when a company would like to identify clusters of customers who have similar shopping habits so that they can create specific marketing strategies that target certain clusters of customers.

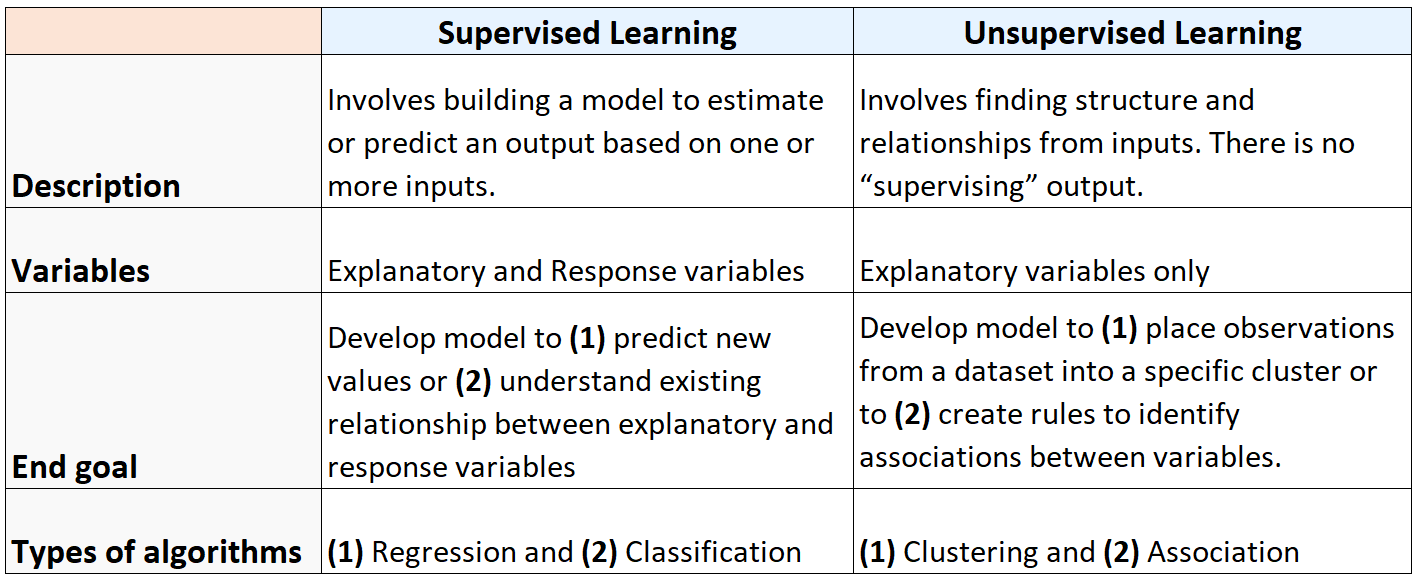
**2. Association:** Using these types of algorithms, we attempt to find “rules” that can be used to draw associations. For example, retailers may develop an association algorithm that says “if a customer buys product X they are highly likely to also buy product Y.”

Here is a list of the most commonly used unsupervised learning algorithms:

* Principal component analysis
* K-means clustering
* K-medoids clustering
* Hierarchical clustering
* Apriori algorithm

**Summary: Supervised vs. Unsupervised Learning**

The following table summarizes the differences between supervised and unsupervised learning algorithms:



And the following diagram summarizes the types of machine learning algorithms:

