**All Machine Learning Models Explained in 6 Minutes**

**Intuitive explanations of the most popular machine learning models.**

<https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a>

[[](https://towardsdatascience.com/@terenceshin?source=post_page-----9fe30ff6776a----------------------)](https://towardsdatascience.com/@terenceshin?source=post_page-----9fe30ff6776a----------------------)

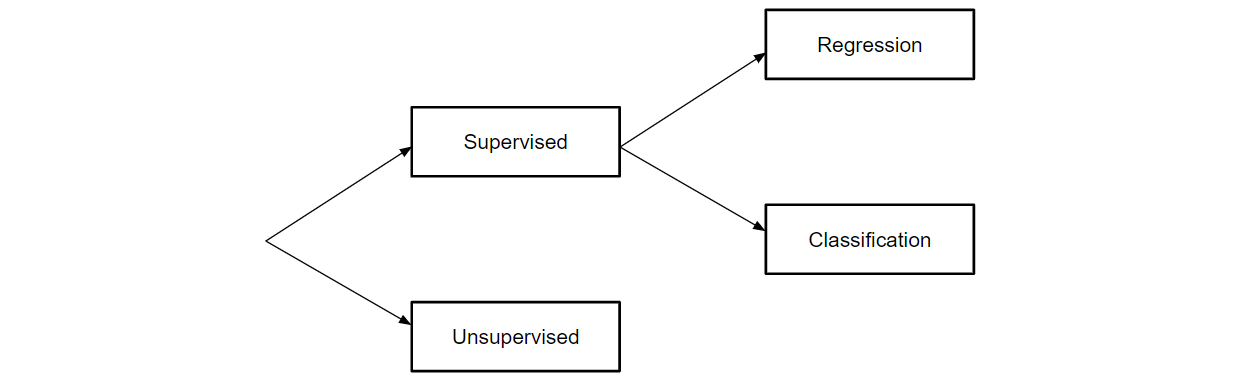
[Terence Shin](https://towardsdatascience.com/@terenceshin?source=post_page-----9fe30ff6776a----------------------)

[Follow](https://medium.com/m/signin?operation=register&redirect=https%3A%2F%2Ftowardsdatascience.com%2Fall-machine-learning-models-explained-in-6-minutes-9fe30ff6776a&source=-360a9d4d19ab-------------------------follow_byline-)

[Jan 5](https://towardsdatascience.com/all-machine-learning-models-explained-in-6-minutes-9fe30ff6776a?source=post_page-----9fe30ff6776a----------------------) · 7 min read



[In my previous article](https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4), I explained what **regression** was and showed how it could be used in application. This week, I’m going to go over the majority of common machine learning models used in practice, so that I can spend more time building and improving models rather than explaining the theory behind it. Let’s dive into it.



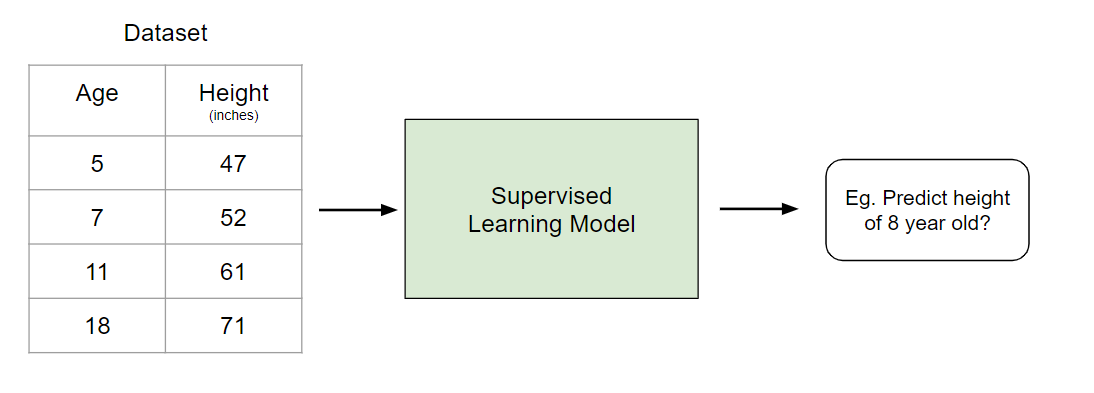
Fundamental Segmentation of Machine Learning Models

All machine learning models are categorized as either **supervised** or **unsupervised**. If the model is a supervised model, it’s then sub-categorized as either a **regression** or **classification** model. We’ll go over what these terms mean and the corresponding models that fall into each category below.

**Supervised Learning**

**Supervised learning** involves learning a function that maps an input to an output based on example input-output pairs [1].

For example, if I had a dataset with two variables, age (input) and height (output), I could implement a supervised learning model to predict the height of a person based on their age.



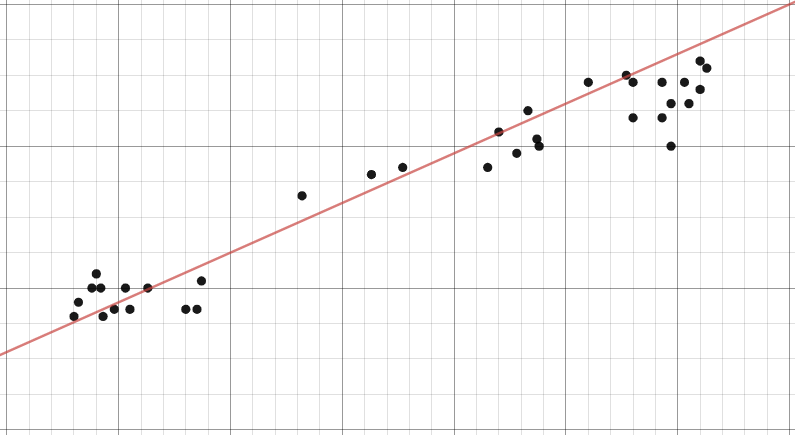
Example of Supervised Learning

To re-iterate, within supervised learning, there are two sub-categories: regression and classification.

**Regression**

In **regression** models, the output is continuous. Below are some of the most common types of regression models.

**Linear Regression**



Example of Linear Regression

The idea of linear regression is simply finding a line that best fits the data. Extensions of linear regression include multiple linear regression (eg. finding a plane of best fit) and polynomial regression (eg. finding a curve of best fit). You can learn more about linear regression in my [previous article](https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4).

**Decision Tree**

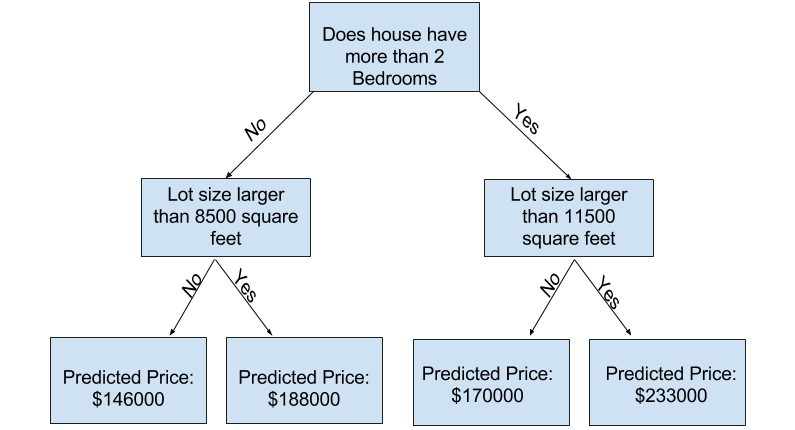
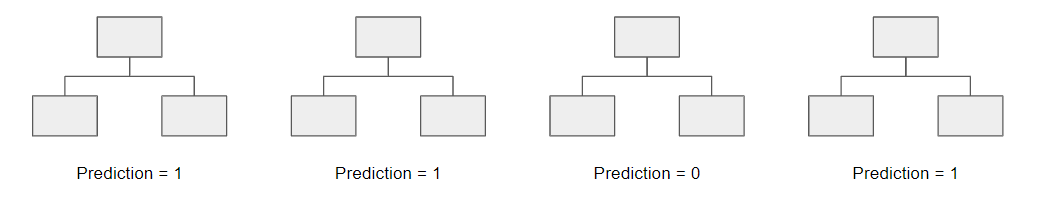


Image taken from Kaggle

**Decision trees** are a popular model, used in operations research, strategic planning, and machine learning. Each square above is called a **node**, and the more nodes you have, the more accurate your decision tree will be (generally). The last nodes of the decision tree, where a decision is made, are called the **leaves** of the tree. Decision trees are intuitive and easy to build but fall short when it comes to accuracy.

**Random Forest**

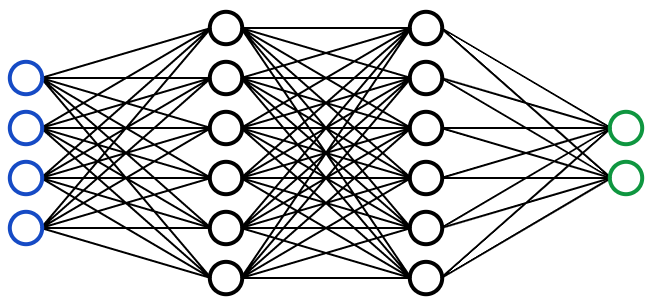
**Random forests** are an [ensemble learning](https://en.wikipedia.org/wiki/Ensemble_learning) technique that builds off of decision trees. Random forests involve creating multiple decision trees using [bootstrapped datasets](https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/) of the original data and randomly selecting a subset of variables at each step of the decision tree. The model then selects the mode of all of the predictions of each decision tree. What’s the point of this? By relying on a “majority wins” model, it reduces the risk of error from an individual tree.



For example, if we created one decision tree, the third one, it would predict 0. But if we relied on the mode of all 4 decision trees, the predicted value would be 1. This is the power of random forests.

StatQuest does an amazing job walking through this in greater detail. See [here](https://www.youtube.com/watch?v=J4Wdy0Wc_xQ&vl=en).

**Neural Network**



Visual Representation of a Neural Network

A **neural network** is a multi-layered model inspired by the human brain. Like the neurons in our brain, the circles above represent a node. The blue circles represent the **input layer,** the black circles represent the **hidden layers,** and the green circles represent the **output layer.** Each node in the hidden layers represents a function that the inputs go through, ultimately leading to an output in the green circles.

Neural networks are actually very complex and very mathematical, so I won’t get into the details of it but…

Tony Yiu’s article gives an intuitive explanation of the process behind neural networks (see [here](https://towardsdatascience.com/understanding-neural-networks-19020b758230)).

If you want to take it a step further and understand the math behind neural networks, check out this free online book [here](http://neuralnetworksanddeeplearning.com/index.html).

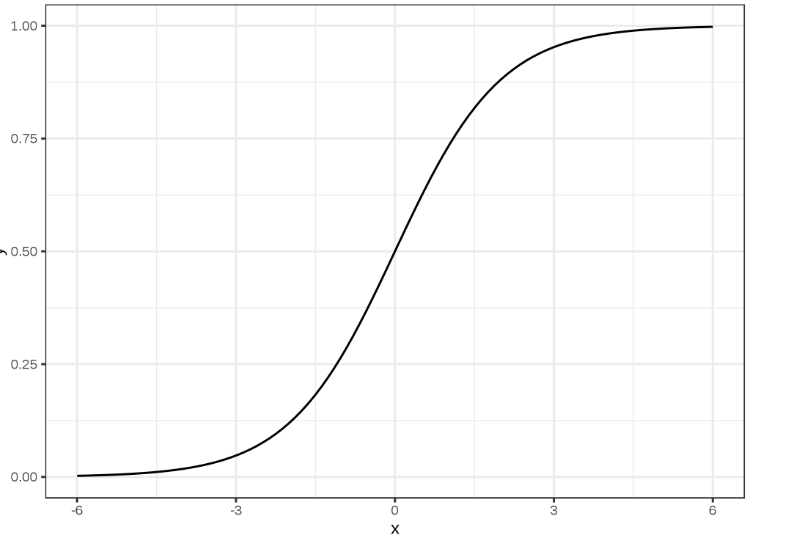
If you’re a visual/audio learner, 3Blue1Brown has an amazing series on neural networks and deep learning on YouTube [here](https://www.youtube.com/watch?v=aircAruvnKk).

**Classification**

In classification models, the output is discrete. Below are some of the most common types of classification models.

**Logistic Regression**

Logistic regression is similar to linear regression but is used to model the probability of a finite number of outcomes, typically two. There are a number of reasons why logistic regression is used over linear regression when modeling probabilities of outcomes (see [here](https://stackoverflow.com/questions/12146914/what-is-the-difference-between-linear-regression-and-logistic-regression)). In essence, a logistic equation is created in such a way that the output values can only be between 0 and 1 (see below).



**Support Vector Machine**

A **Support Vector Machine** is a supervised classification technique that can actually get pretty complicated but is pretty intuitive at the most fundamental level.

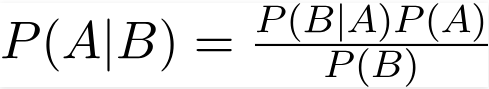
Let’s assume that there are two classes of data. A support vector machine will find a **hyperplane** or a boundary between the two classes of data that maximizes the margin between the two classes (see below). There are many planes that can separate the two classes, but only one plane can maximize the margin or distance between the classes.



If you want to get into greater detail, Savan wrote a great article on Support Vector Machines [here](https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72).

**Naive Bayes**

Naive Bayes is another popular classifier used in Data Science. The idea behind it is driven by Bayes Theorem:



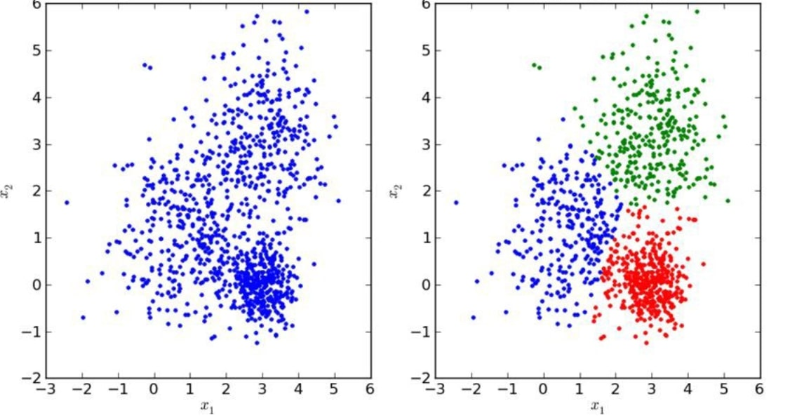
While there are a number of unrealistic assumptions made in regards to Naive Bayes (hence why it’s called ‘Naive’), it has proven to perform quite most of the time and it is also relatively fast to build.

See [here](https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c) if you want to learn more about them.

**Decision Tree, Random Forest, Neural Network**

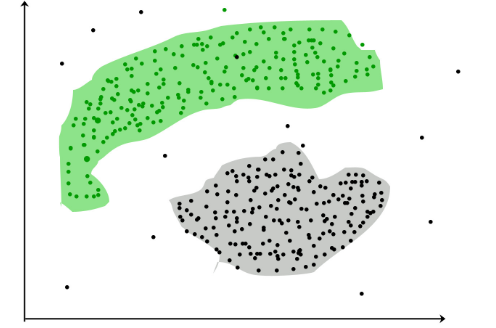
These models follow the same logic as previously explained. The only difference is that that output is discrete rather than continuous.

**Unsupervised Learning**



Unlike supervised learning, **unsupervised learning** is used to draw inferences and find patterns from input data without references to labeled outcomes. Two main methods used in unsupervised learning include clustering and dimensionality reduction.

**Clustering**



Taken from GeeksforGeeks

Clustering is an unsupervised technique that involves the grouping, or **clustering**, of data points. It’s frequently used for customer segmentation, fraud detection, and document classification.

Common clustering techniques include **k-means** clustering, **hierarchical** clustering, **mean shift** clustering, and **density-based** clustering. While each technique has a different method in finding clusters, they all aim to achieve the same thing.

**Dimensionality Reduction**

Dimensionality reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables [2]. In simpler terms, its the process of reducing the dimension of your feature set (in even simpler terms, reducing the number of features). Most dimensionality reduction techniques can be categorized as either **feature elimination** or **feature extraction.**

A popular method of dimensionality reduction is called **principal component analysis.**

**Principal Component Analysis (PCA)**

In the simplest sense, **PCA** involves project higher dimensional data (eg. 3 dimensions) to a smaller space (eg. 2 dimensions). This results in a lower dimension of data, (2 dimensions instead of 3 dimensions) while keeping all original variables in the model.

There is quite a bit of math involved with this. If you want to learn more about it…

Check out this awesome article on PCA [here](https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c).

If you’d rather watch a video, StatQuest explains PCA in 5 minutes [here](https://www.youtube.com/watch?v=HMOI_lkzW08&vl=en).

**Conclusion**

Obviously, there is a ton of complexity if you dive into any particular model, but this should give you a fundamental understanding of how each machine learning algorithm works!

*Check out this* [***link***](https://towardsdatascience.com/basic-statistics-you-need-to-know-for-data-science-1fdd290f59b5) *if you want to learn* ***all basic statistics for data science.***

*Check out this*[***link***](https://towardsdatascience.com/an-extensive-guide-to-exploratory-data-analysis-ddd99a03199e) *if you want to learn a* ***step by step process for exploratory data analysis (EDA).***

**References**

[1] Stuart J. Russell, Peter Norvig, Artificial Intelligence: A Modern Approach (2010), Prentice Hall

[2] Roweis, S. T., Saul, L. K., Nonlinear Dimensionality Reduction by Locally Linear Embedding (2000), *Science*

**Thanks for the read!**

If you like my work and want to support me, sign up on my email list [here](https://forms.gle/UGdTom9G6aFGHzPD9) to be the first to hear about new and exclusive content! :)