## **Concepts of Machine Learning**

### Why do we need to care about machine learning?

A breakthrough in machine learning would be worth ten Microsofts.

— [Bill Gates](https://en.wikipedia.org/wiki/Bill_Gates" \t "https://smartbridge.teachable.com/courses/846395/lectures/_blank), Former Chairman, Microsoft

Machine Learning is getting computers to program themselves. If programming is automation, then machine learning is automating the process of automation.

Writing software is the bottleneck, we don’t have enough good developers. Let the data do the work instead of people. Machine learning is the way to make programming scalable.

* ****Traditional Programming****: Data and program is run on the computer to produce the output.
* ****Machine Learning****: Data and output is run on the computer to create a program. This program can be used in traditional programming.

Machine learning is like farming or gardening. Seeds is the algorithms, nutrients is the data, the gardner is you and plants is the programs.

### Key Elements of Machine Learning

There are tens of thousands of machine learning algorithms and hundreds of new algorithms are developed every year.

Every machine learning algorithm has three components:

* ****Representation****: how to represent knowledge. Examples include decision trees, sets of rules, instances, graphical models, neural networks, support vector machines, model ensembles and others.
* ****Evaluation****: the way to evaluate candidate programs (hypotheses). Examples include accuracy, prediction and recall, squared error, likelihood, posterior probability, cost, margin, entropy k-L divergence and others.
* ****Optimization****: the way candidate programs are generated known as the search process. For example combinatorial optimization, convex optimization, constrained optimization.

All machine learning algorithms are combinations of these three components. A framework for understanding all algorithms.

### Types of Learning

There are four types of machine learning:

* ****Supervised learning****: (also called inductive learning) Training data includes desired outputs. This is spam this is not, learning is supervised.
* ****Unsupervised learning****: Training data does not include desired outputs. Example is clustering. It is hard to tell what is good learning and what is not.
* ****Semi-supervised learning****: Training data includes a few desired outputs.
* ****Reinforcement learning****: Rewards from a sequence of actions. AI types like it, it is the most ambitious type of learning.

Supervised learning is the most mature, the most studied and the type of learning used by most machine learning algorithms. Learning with supervision is much easier than learning without supervision.

Inductive Learning is where we are given examples of a function in the form of data (x) and the output of the function (f(x)). The goal of inductive learning is to learn the function for new data (x).

* ****Classification****: when the function being learned is discrete.
* ****Regression****: when the function being learned is continuous.
* ****Probability Estimation****: when the output of the function is a probability.

### Machine Learning in Practice

Machine learning algorithms are only a very small part of using machine learning in practice as a data analyst or data scientist. In practice, the process often looks like:

1. Start Loop
2. ****Understand the domain, prior knowledge and goals****. Talk to domain experts. Often the goals are very unclear. You often have more things to try then you can possibly implement.
3. ****Data integration, selection, cleaning and pre-processing****. This is often the most time consuming part. It is important to have high quality data. The more data you have, the more it sucks because the data is dirty. Garbage in, garbage out.
4. ****Learning models****. The fun part. This part is very mature. The tools are general.
5. ****Interpreting results****. Sometimes it does not matter how the model works as long it delivers results. Other domains require that the model is understandable. You will be challenged by human experts.
6. ****Consolidating and deploying discovered knowledge****. The majority of projects that are successful in the lab are not used in practice. It is very hard to get something used.
7. End Loop

It is not a one-shot process, it is a cycle. You need to run the loop until you get a result that you can use in practice. Also, the data can change, requiring a new loop.