INTRODUCTION TO MACHINE LEARNING

Amit Kumar Manjhi

AGENDA

- **What is Machine Learning?**
- **Key Concepts in Machine Learning**
- **How would you apply Machine Learning to solve your problem?**
- **Python Tools For Machine Learning**
- **An Example of Machine Learning Problem**

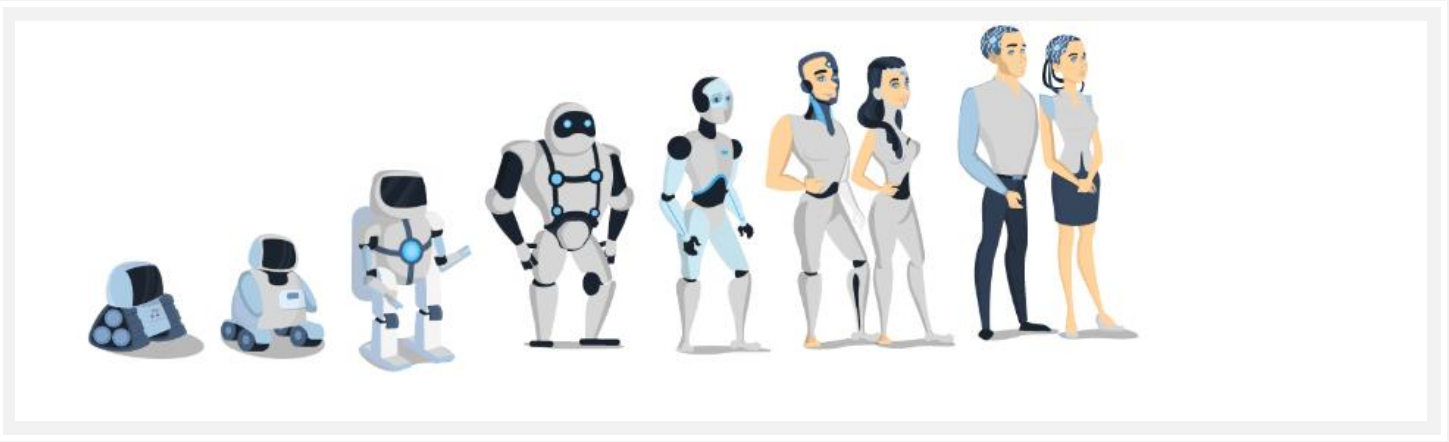
What Is Machine Learning?

- Machine Learning is the science (and art) of programming computers so they can *learn from data*.

- General Definition:

- Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed. [Arthur Samuel, *1959]*

- **Applied Machine Learning:** Here applied means how to interpret the machine learning algorithms at a high level.



Why Machine Learning?

- Consider how you would write a spam filter using traditional programming techniques

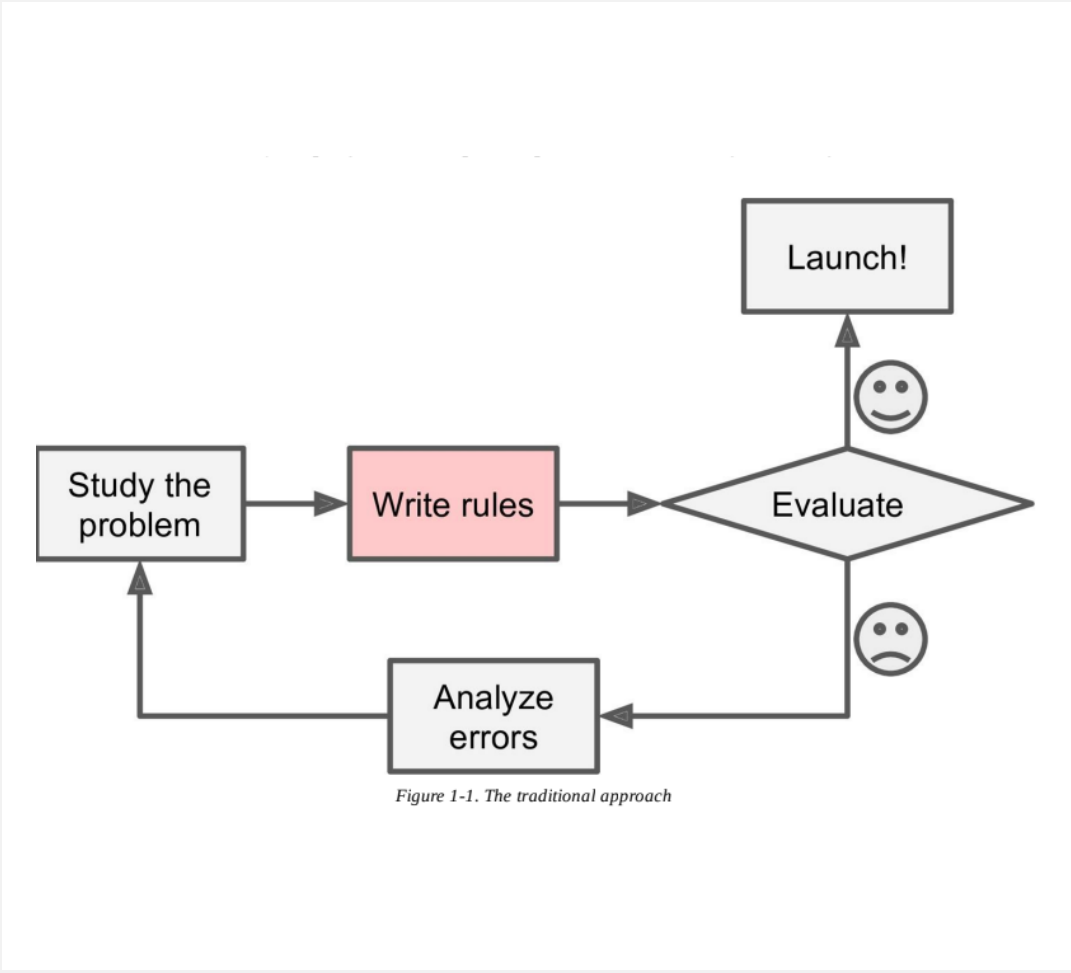


Figure 1-1. The traditional approach

MACHINE LEARNING APPROACH

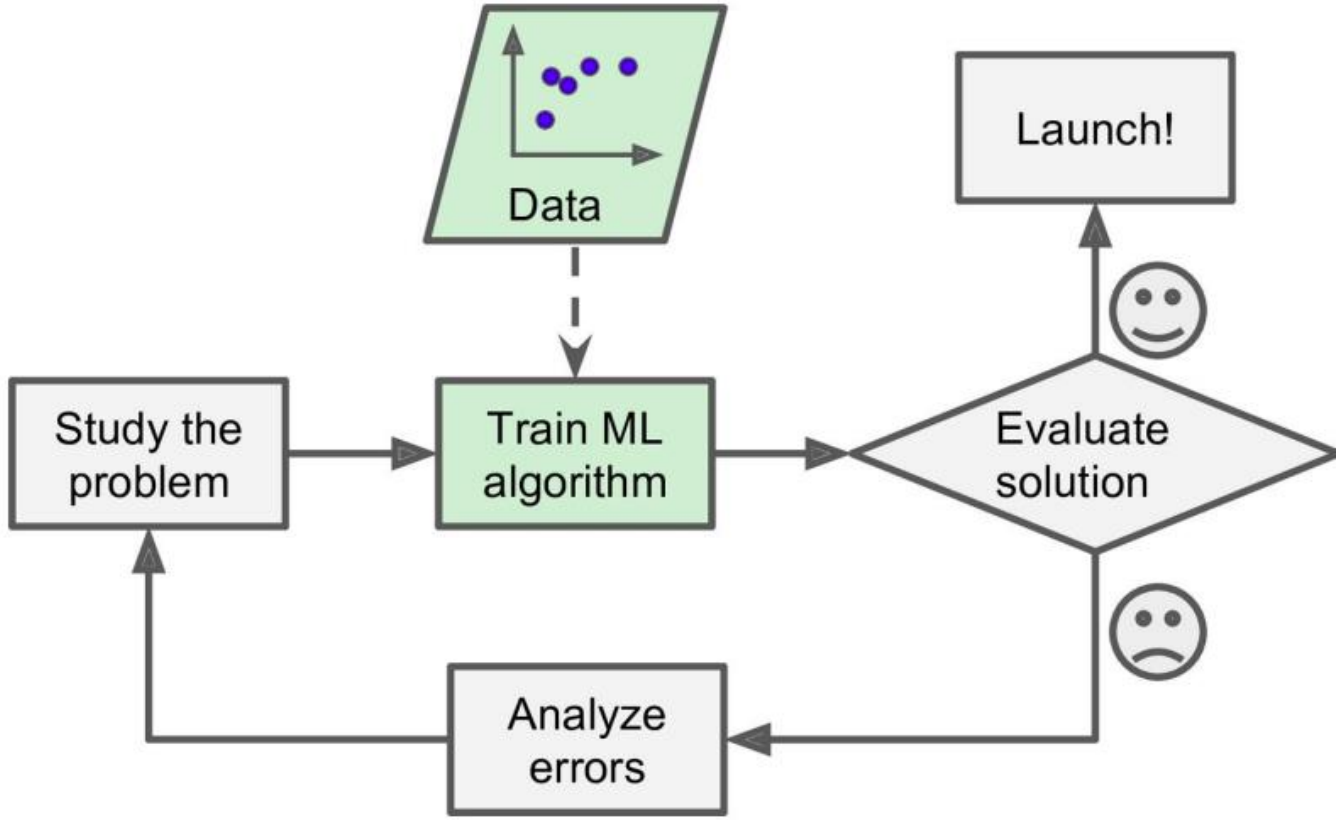


Figure 1-2. Machine Learning approach

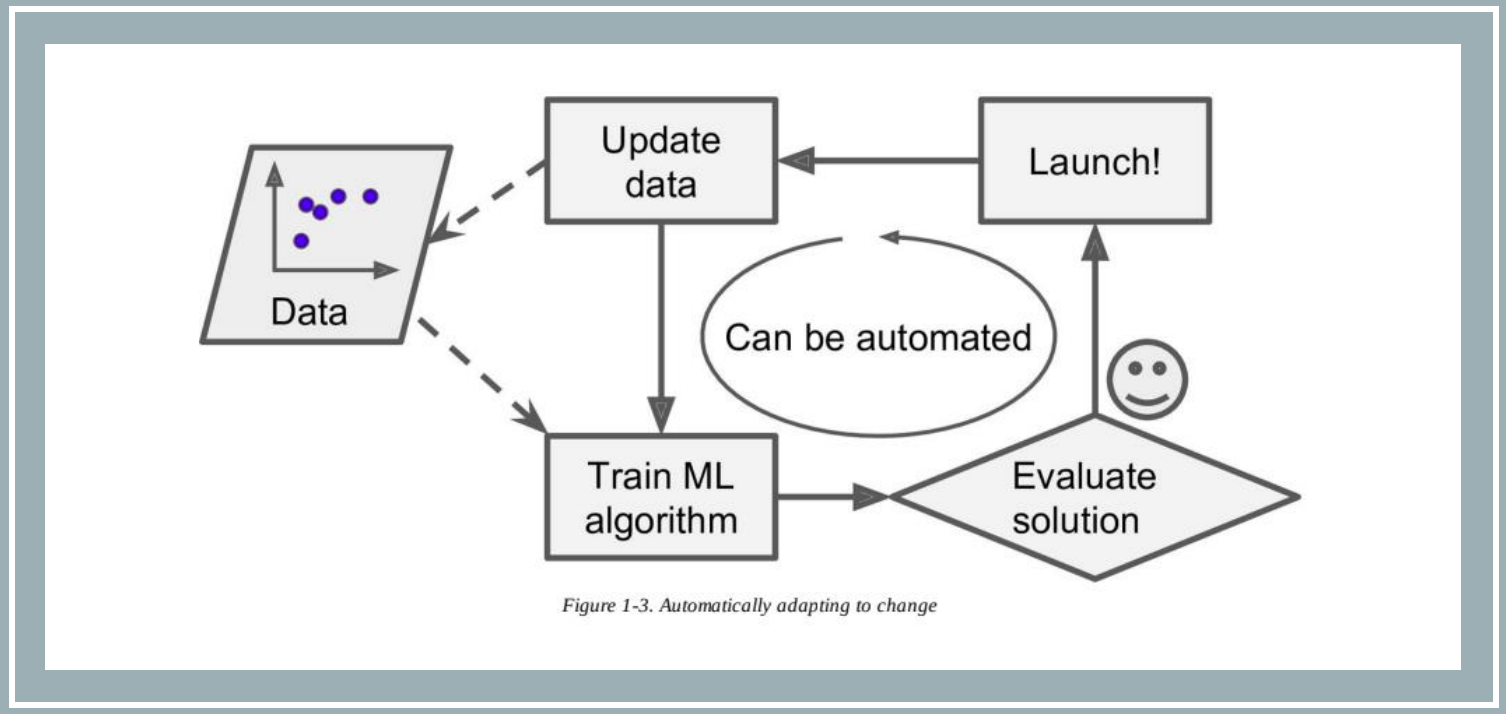


Figure 1-3. Automatically adapting to change

Machine Learning Approach

To Summarize, Machine Learning is Great For:

- Problems for which existing solutions require a lot of hand-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better.
- Complex problems for which there is no good solution at all using a traditional approach: the best Machine Learning techniques can find a solution.
- Fluctuating environments: a Machine Learning system can adapt to new data.

Key Concept In Machine Learning

- Machine Learning systems can be classified according to the amount and type of supervision they get during training.
- **Supervised learning (**Train Me!) : Once the model gets trained it can start making a prediction or decision when new data is given to it.
- Supervised Algorithms:
- k-Nearest Neighbors, Linear Regression ,Logistic Regression ,Support Vector Machines (SVMs) ,Decision Trees and Random Forests

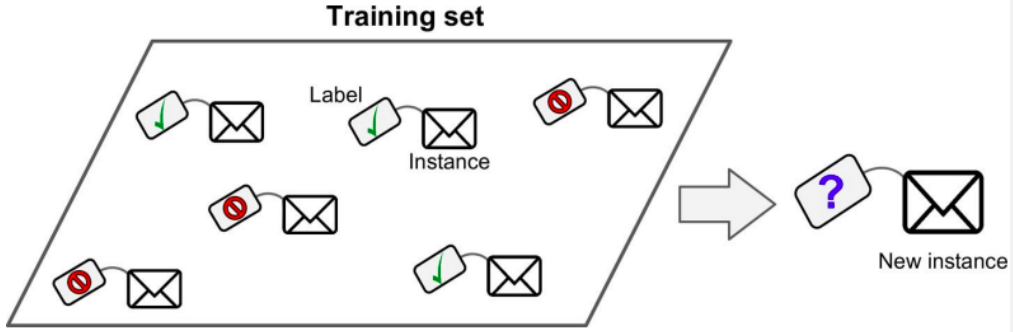


Figure 1-5. A labeled training set for supervised learning (e.g., spam classification)

Unsupervised Learning

- **Unsupervised Learning** – I am self sufficient in learning.

- Unsupervised Algorithms:

Clustering k-Means ,Hierarchical Cluster Analysis (HCA) ,Principal Component Analysis (PCA)

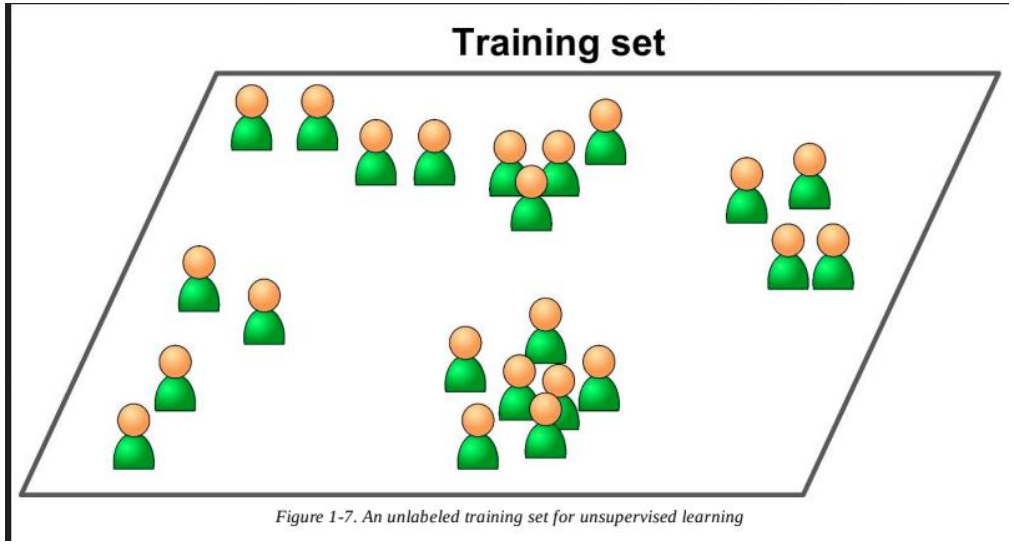


Figure 1-7. An unlabeled training set for unsupervised learning

Semisupervised Learning

- Some algorithms can deal with partially labeled training data, usually a lot of unlabeled data and a little bit of labeled data. This is called *semisupervised learning*

- Most semisupervised learning algorithms are combinations of unsupervised and supervised algorithms. For example, *deep belief networks* (DBNs)

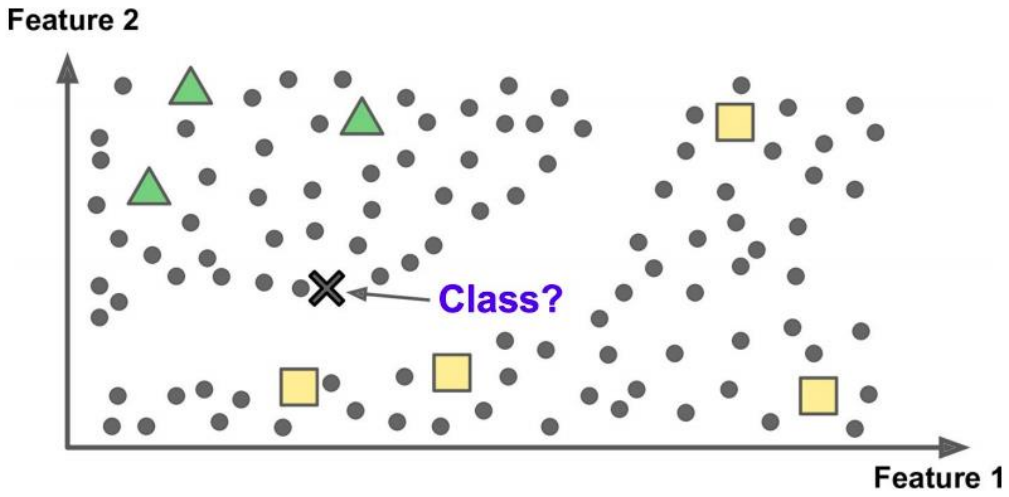


Figure 1-11. Semisupervised learning

Reinforcement learning – my life my rules! (Hit & trial)

- The learning system, called an *agent* in this context, can observe the environment, select and perform actions, and get *rewards* in return.
- For example, many robots implement Reinforcement Learning algorithms to learn how to walk. DeepMind's AlphaGo program is also a good example of Reinforcement Learning

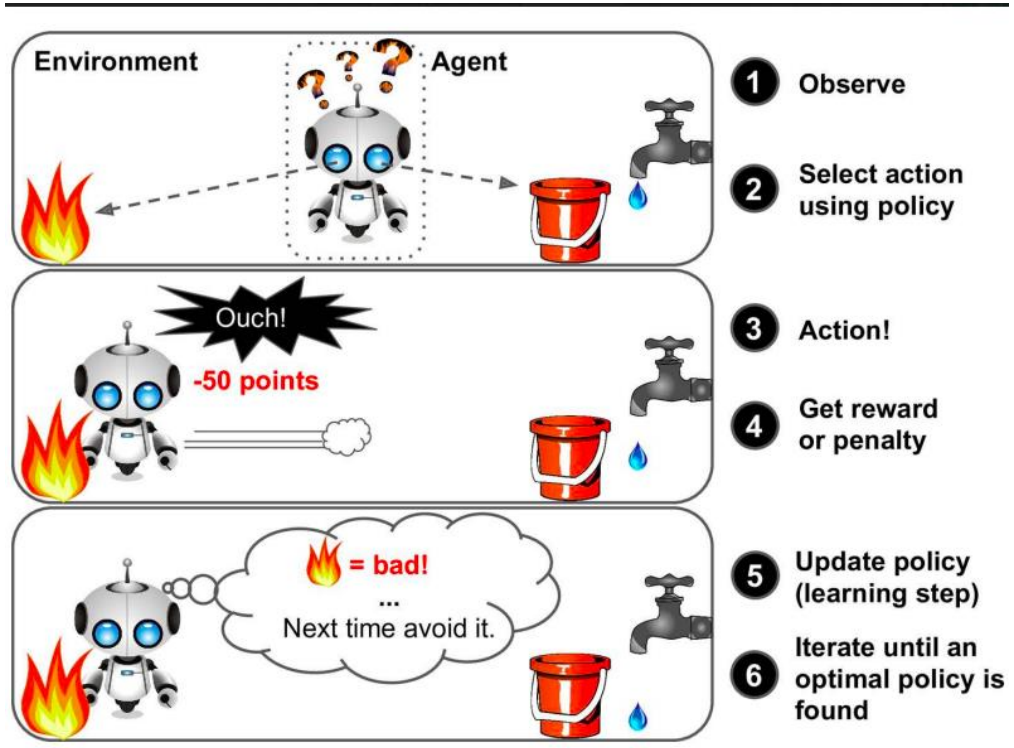


Figure 1-12. Reinforcement Learning

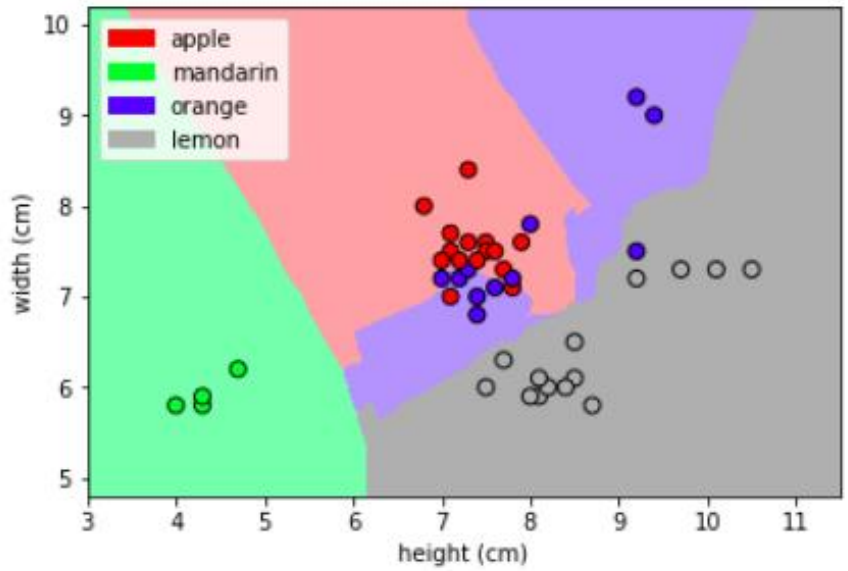
How would you apply machine learning to solve your problem?

- There are three basic steps:
- **Representation:** How to represent learning problems in terms of something that a computer can understand. In this step, you have to also decide what kind of algorithm to apply to selected data.
- **Evaluation:** In this step, we can calculate the quality and accuracy score for the prediction coming out from the machine learning algorithm typically called classifier.
- **Optimization:** In this step, we need to search for an optimal classifier that gives the best outcome for the selected problem.

Python Tools For Python

- **scikit-learn:** scikit learn is the most widely used Python library for machine learning. It is an open-source project
- **SciPy:** sciPy is a python library that supports data manipulation and commonly used in scientific computing which includes statistical distribution, optimization of functions, linear algebra, and variety of specialized mathematical functions.
- **Numpy:** Numpy is a scientific computing python library that contains the support for fundamentals data structure used by scikit-learn.such as multidimensional-array.
- **Pandas:** Pandas is a python library for data manipulation and analysis. It supports data structures like DataFrame.
- **matplotlib:** matplotlib is widely used python 2D plotting library that produces high-quality figures in a variety of formats

AN EXAMPLE OF MACHINE LEARNING PROBLEM



References

- https://www.edureka.co/blog/what-is-machine-learning/
- https://www.coursera.org/learn/python-machine-learning
- https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/