Machine Learning

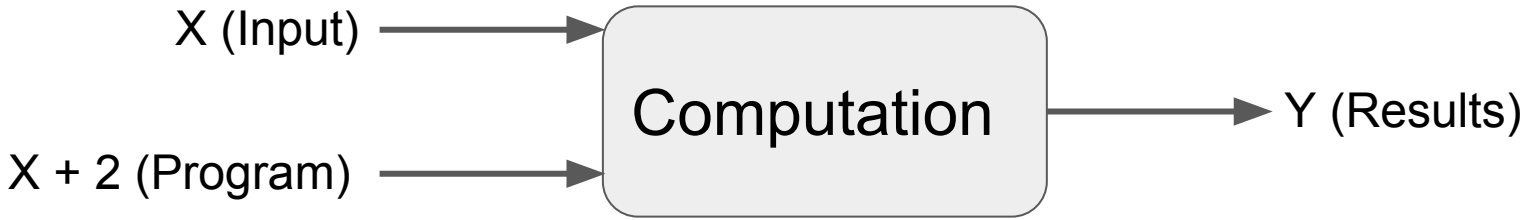
With

Scikit Learn

Part - I

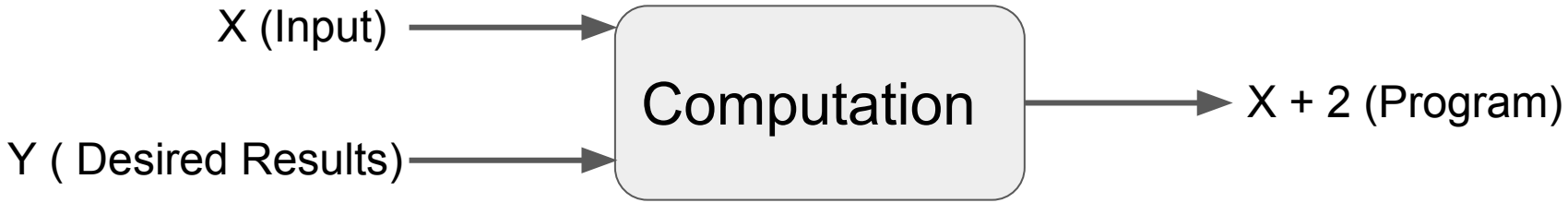
Traditional Programming

Traditional Programming is the art of automating a task by writing rules for a computer to follow.



Machine Learning

Gives “computer the ability to learn without being explicitly programmed”



Machine Learning Overview

Supervised Learning Learning the mapping function from input to output.	**Unsupervised Learning** *Learning the underlying structure in the data*	**Reinforcement Learning** The learner is not told which action to take, but instead must discover which action will yield the maximum reward.
Classification Learning the mapping function from input to ***discrete*** output.	**Clustering** Learning the ***inherent groupings*** in the data	
Regression Learning the mapping function from input to ***continues*** output.	**Dimensionality Reduction** Reducing the ***dimensionality*** of data while maintaining its structure and usefulness.	

Quiz 1

Q. A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measure by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. In this setting, what is T?

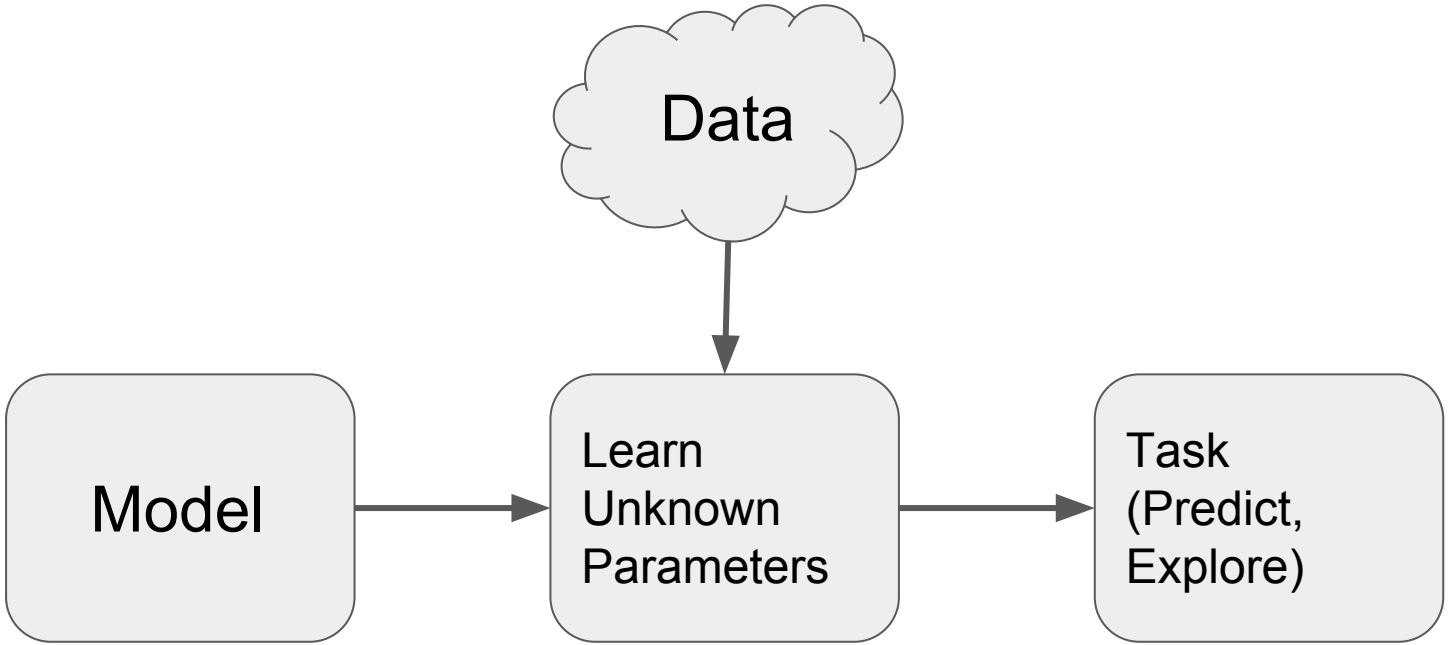
1. The process of the algorithm examining a large amount of historical weather data
2. None of these.
3. The weather prediction task.
4. The probability of it correctly predicting a future date’s weather.

Quiz 2

Q. Some of the problems below are best addressed using a supervised learning algorithm or unsupervised learning algorithm. Which of the following would you apply?

1. Given genetic (DNA) data from a person, predict the odds of him/her developing diabetes over the next 10 years.
2. Given data on how 1000 medical patients respond to drug, discover whether there are different categories of patients in terms of how they respond to the drug.
3. Given a large dataset of medical records from patients suffering from heart disease, try to learn whether there might be different clusters of such patients for which we might tailor separate treatments.
4. Have a computer examine an audio clip of a piece of music, and classify whether or not there are vocals in the audio clip.

Machine Learning Process



K Nearest Neighbour (KNN)

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure

Model Parameters and Hyper parameters

- Parameters
  - A model parameter is a configuration variable that is internal to the model and whose value can be learned from data.

- Hyper parameters
  - A model hyperparameter is a configuration that is external to the model and whose value cannot be estimated from data.

Dataset

- Training Set
  - Dataset to learn the parameters of the model.

- Validation Set
  - Dataset to tune the hyper parameters of the model.

- Test Set
  - Dataset to evaluate the performance of the model.

Scikit Learn

A python package that provides solid implementation of wide range of machine learning algorithms.