**Machine Learning – Chapter 1**

**Why do Machine Learning?**

* Understand and improve efficiency of human learning  
  Eg: Using methods for teaching and tutoring people, as done in CAI -- Computer-aided instruction
* Discover new things or structure that is unknown to humans  
  Example: Data mining

**Machine Learning:**

**Definition:**

Computational methods using experience to improve performance

E.g., to make accurate predictions

Experience: data-driven task, statistics, probability.

Example: use height and weight to predict gender.

Computer science: need to design efficient and accurate algorithms, analysis of complexity, theoretical guarantees.

**Machine Learning:**

Study of algorithms that

• improve their performance P

• at some task T

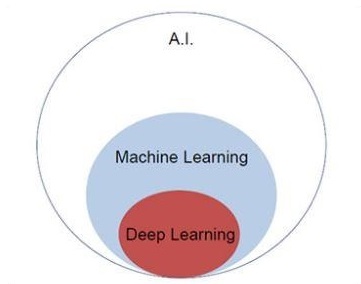
• with experience E

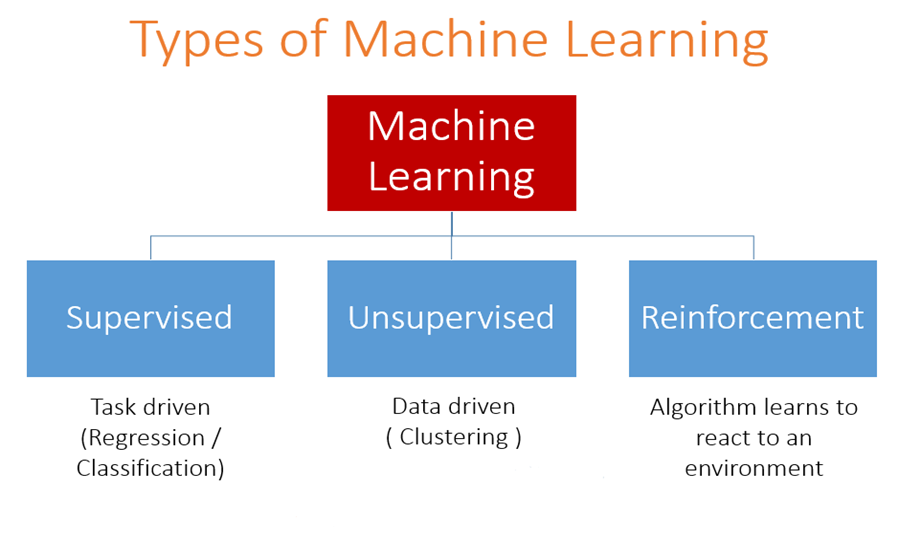
well-defined learning task: <P,T,E>

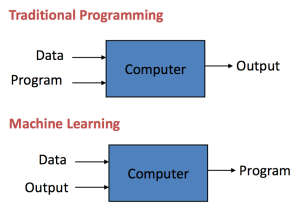
**Examples of Learning Tasks**

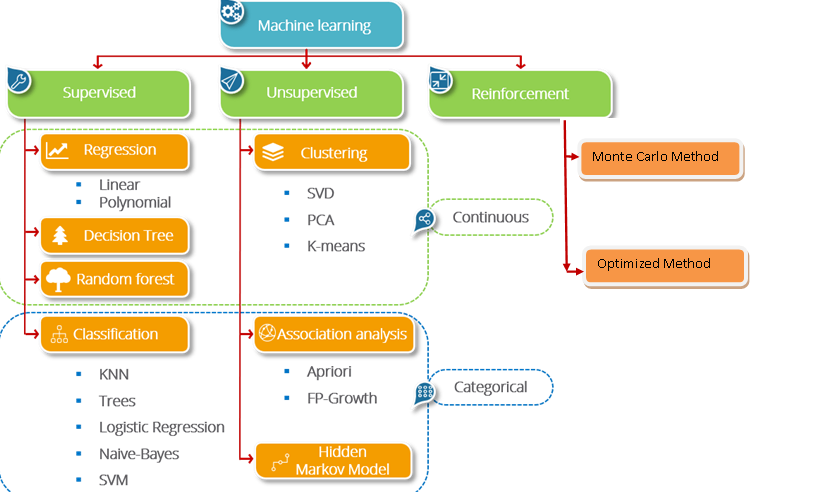
1. Optical character recognition.
2. Text or document classification
3. Spam detection
4. Forecasting (e.g., stock market trends)
5. Pattern recognition
6. Robotics
7. Morphological analysis
8. Part-of-speech tagging
9. Statistical parsing
10. Speech recognition, speech synthesis, speaker verification.
11. Image recognition, face recognition
12. Fraud detection (credit card, telephone), network intrusion.
13. Games (chess, backgammon)
14. Unassisted control of a vehicle (robots, navigation).
15. Medical diagnosis.
16. Recommendation systems, search engines, information extraction systems.

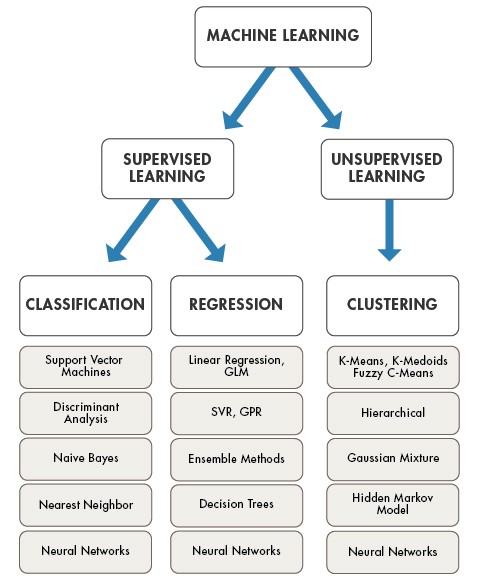
**Types of Machine Learning:**

* **Supervised learning -** labeled training data.
  + Supervised learning deals with learning a function from available training data.
  + A supervised learning algorithm analyzes the training data and produces a function, which can be used for mapping new examples
  + In supervised learning, we define the possible outcomes.
  + Example: A day can be - hot, cold, or moderate.
  + Since we defined the options are – hot, cold, and moderate
  + The algorithm will spit out predictions within that framework.
  + In Supervised learning: We provide the algorithm with X and Y, and try to figure out the relationships between the two
  + **Examples:** 
    - Classifying e-mails as spam,
    - Labeling webpages based on their content, and
    - Voice recognition.
* **Unsupervised learning -** no labeled data .
  + Unsupervised learning makes sense of unlabeled data without having any predefined dataset for its training.
  + Unsupervised learning is an extremely powerful tool for analyzing available data and look for patterns and trends.
  + The data we have and tries to figure out itself what the different potential groupings
  + An unsupervised learning algorithm might find that the groups with the most distinct features are days that are very hot, moderate, and cloudy.
  + Instead of deciding the groups of outcomes in advance and trying to map relationships to them, we let the algorithm find the ones it feels are the most natural.
  + In unsupervised learning: there is no Y – we’re just trying to understand the underlying organization of the data in of itself.
  + Let the computer learn how to do something, and use this to determine structure and patterns in data
  + **Examples:** 
    - k-means
    - self-organizing maps, and
    - hierarchical clustering
* **Reinforcement learning [Machine learns to act on its own to achieve the goal]**
  + Finds a sequence of actions (policy) that reaches a goal.
  + No supervised output but delayed reward.
  + Eg: playing chess or a computer game, robot in a maze.
  + Machine learning works with large amounts of data. It is useful for small amounts of data too.
  + Deep learning works **efficiently** if the amount of data increases rapidly.
  + Deep learning algorithms are **designed to heavily depend on high-end machines** 









**How do Machines Learn?**

Machine Learning Process divided into three parts

1. Data Input – Past Data / Information is used for future decision making
2. Abstraction – Input data represents through algorithms
3. Generalization – Abstracted representation is generalized in form of framework for making decisions

**Abstraction**: [Abstract the knowledge which comes as input data in the form of a model]

In Machine learning process: Knowledge is fed in the form of input data

Map / model in machine learning is represented as raw data

**Model can be represented as,**

* Computational blocks if/else
* Mathematical equations
* Data structures tree/graph
* Logical grouping of similar observations

**Choose of the model can solve specific problem,**

1. Type of the problem – forecast / prediction / analysis of trends / understanding different group of objects
2. Nature of input data – Data types / no value in many fields
3. Domain of the problem – High rate of data input and need immediate inference Eg: Fraud detection

* Once the model is chosen
* Fit the model based on the input data
* If Mathematical equation: y = c1 +c2x [Linear regression]
* We need to find the value of c1 and c2
* Fitting the model is – finding the value of unknown co-efficient / constants
* Process of fitting the model based on input data is **training data**

**Generalization**

Abstract the knowledge which comes as input data in the form of a model

Abstraction – is training the model – for future decision – Generalization

Model is trained base on the finite set of data – but we need to apply the model to make decision on unknown data – called as **test data**

Risk: May not take correct decision

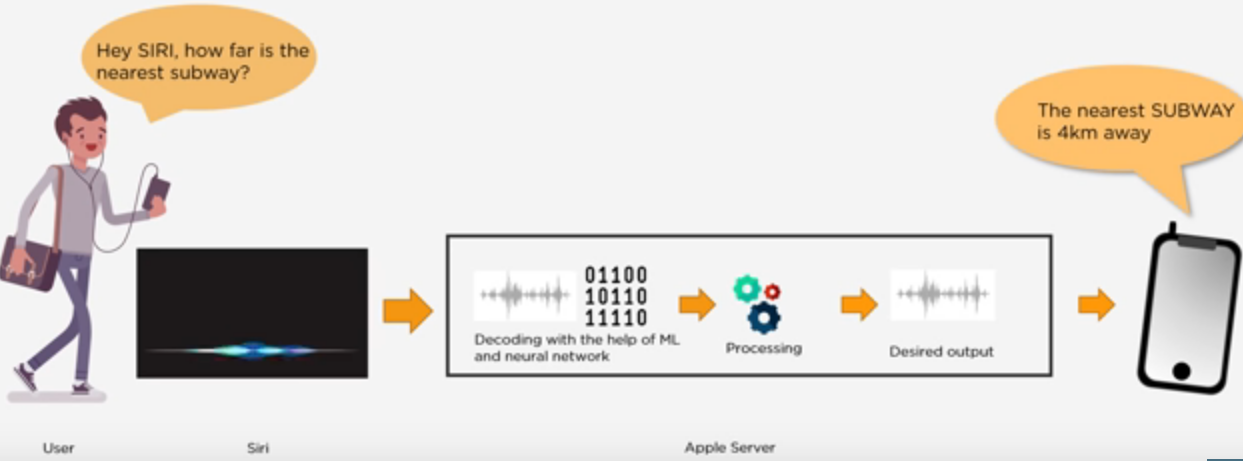
**Well-posed learning problem:**

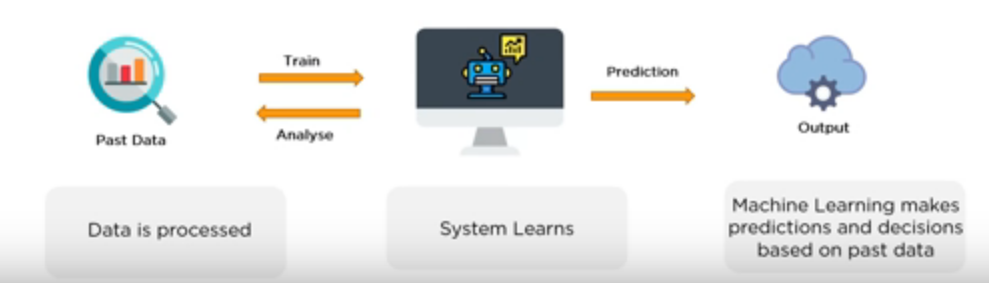
Framework should answer:

1. What is the problem? – describe problem informally and formally, list assumptions and similar problems
2. Why problem need to be solved? List motivation of solving the problem, benefits of solutions and how solutions can be used.
3. How to solve the problem?
   1. Data collection, data preparation, program design to solve the problem,

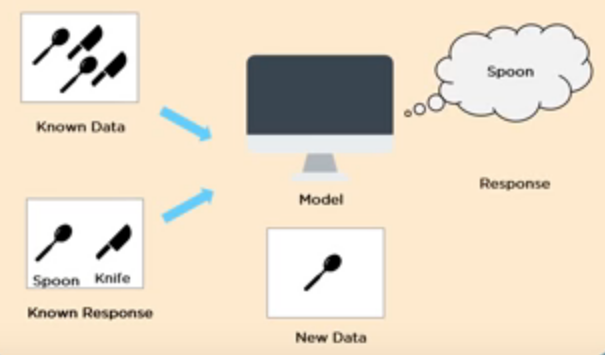
**Types of Machine Learning:**

1. Supervised learning
   * Predictive learning, predicts the class of unknown objects based on prior class information / labeled dataset
2. Unsupervised learning
   * Descriptive learning, machine finds pattern in unknown objects by grouping similar objects
3. Reinforcement learning
   * Machine learns to act on its own to achieve the goal





**Supervised Learning: [Algorithm is trained used data that is labeled]**

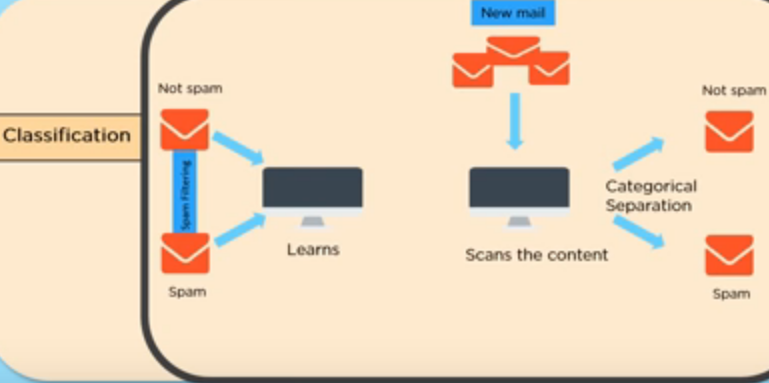
****

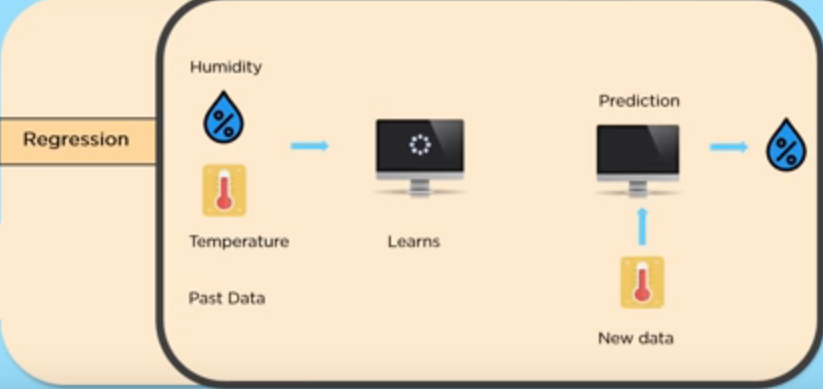
**Supervised Learning Types:**

1. Classification: When the output variable is categorical, with two / more classes (yes / no, true / false), we make use of classification
2. Regression: Relationship between two / more variables, where a change in one variable is associated with a change in other variable.

**Examples:** Spam Mail [Classification]

Weather Prediction [Regression]

****

****

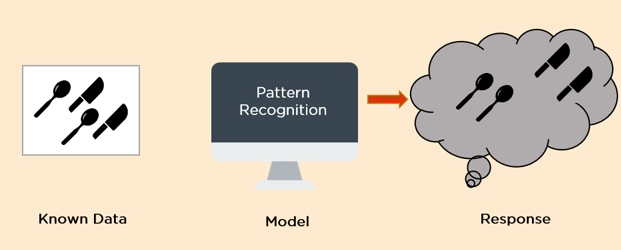
**Areas where supervised learning is used:**

1. Prediction the result of the game
2. Predicting whether tumor is malignant/benign
3. Predicting price of real estate, stocks etc
4. Risk assessment
5. Fraud detection
6. Image classification
7. Visual recognition, Handwriting recognition

**Supervised algorithms:**

1. Naïve Bayes
2. K-nearest neighbor (kNN)
3. Decision tree
4. Linear regression
5. Logistic regression
6. Support Vector Machine (SVM)

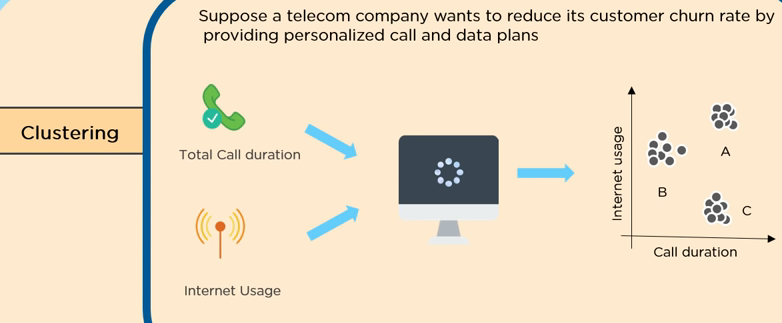
**Un-Supervised Learning: [Algorithm is trained used data that is unlabeled]**

****

**Un-Supervised Learning Types: [Descriptive model / pattern discovery / knowledge discovery]**

1. **Clustering:** The method of dividing objects into clusters which are similar between them and dissimilar belongs to other cluster
2. **Association**: Discovering the probability of co-occurrence of items in a collection

**Clustering Example:**

****

**Association Example:**

****

**Areas where un-supervised learning is used:**

1. Market Basket analysis
2. Delivery store optimization
3. Semantic clustering
4. Identifying accident prone areas

**Un-supervised algorithms**

1. K-means
2. Principal Component Analysis (PCA)
3. Self-organizing map (SOM)
4. Apriori algorithm
5. DBSCAN

**Reinforcement learning:**

These types of learning is used when there is no idea about class / label of particular data

This model has to do classification – it will be get rewarded if the classification is correct, else get punished

Most complex to learn and apply

**Applications:**

1. Self driving cars
2. Intelligent robot
3. AlphaGo Zero

**Reinforcement algorithms:**

1. Q-learning
2. Sarsa

**Problems not to be solved using Machine Learning**

1. Where the task needs human intervention frequently
   1. Example: Traffic control where human involvement is needed
2. Price calculator
3. Dispute tracking application
4. If task is already optimized
5. Insufficient training data & Bad data 🡪 Quality of the prediction may not be good

**Tools for Machine Learning**

1. Python
2. R
3. Matlab
4. SAS [Statistical Analysis System]
5. SPSS [Statistical Package for the Social Science]
6. Julia – Open source has Matlab, Python, R