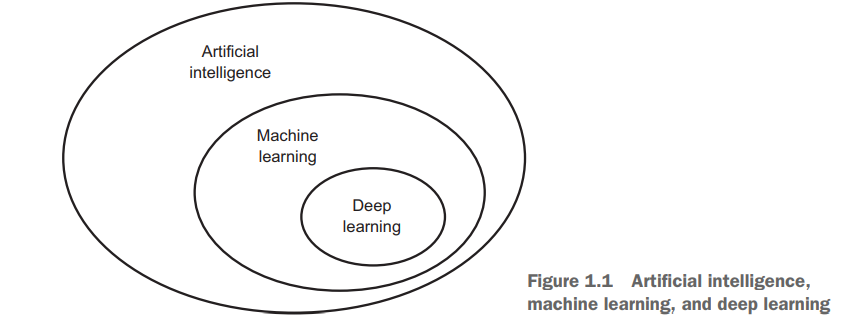
**Chapter 1 Summary:**

# **This chapter covers**

* High-level definitions of fundamental concepts
* Timeline of the development of machine learning
* Key factors behind deep learning’s rising popularity and future potential



**Artificial Intelligence** since 1950s

Human learning on the basics of experiment is called intelligence. If we put this intelligence in machine it is called Artificial Intelligence.

Artificial Intelligence is Ability of a computer program or a machine to think and learn itself. In One Line (Simply A Program that can sense, Reason, act & adapt.)

**Explore** **Symbolic AI:** Using If-else, password checking. Problem (Tagging, perception problem) , Classical Programing , Handcrafted Rules.

**Explore**: Turning Test

**Machine Learning** …1990s

Machine Learning is a subset of AI Algorithms whose performance improves as they are exposed to more data over time.

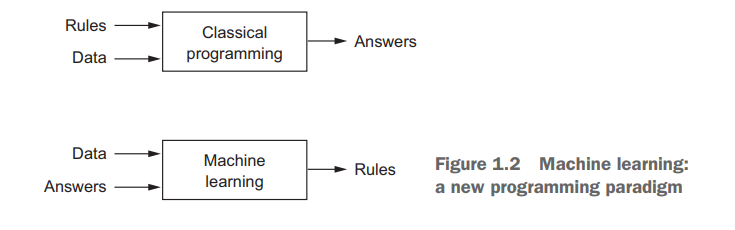
In simple a machine that can learn from data, not from handcrafted rules.

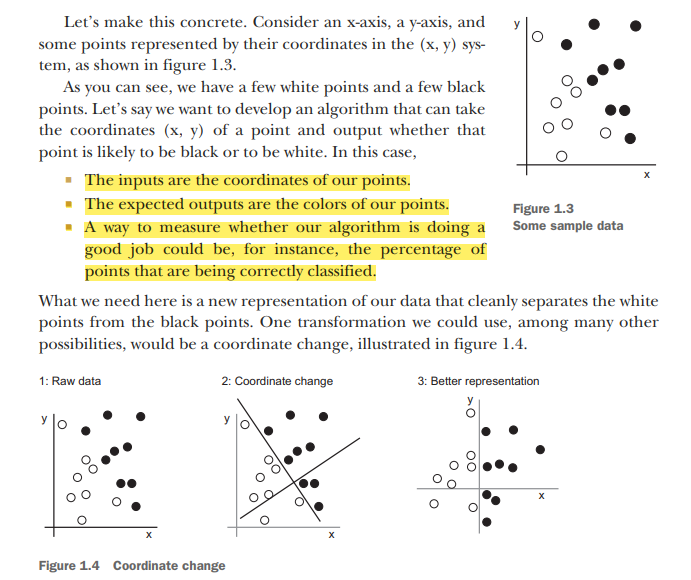
**Two Points**

* Machine have the ability to learn from data.
* Not learn from Hand crafted rules.
* A machine-learning system is trained rather than explicitly programmed.

Three Things we need to remember in ML

**Input + Programs = Output (Classical Programing)**

**Input + Output = Program (ML)**

**Example:**

**This new representation basically solves the classification problem.**

**Bayesian:**

Bayesian machine learning is a particular set of approaches to probabilistic machine learning (for other probabilistic models, see Supervised Learning).

Bayesian learning treats model parameters as random variables — in Bayesian learning, parameter estimation amounts to computing posterior distributions for these random variables based on the observed data.

**P(h/D) = P(D/h)P(h) / P(D)**

**Deep learning** (Multi Stage Learning)

A machine that learns from Neural networks models or artificial neural network inspired by the human brain& in which feature engineering is Easy as Compared to Machine learning.

**Two Points**

* We use neural network (Neurons)
* Feature Engineering is Auto

**Work Y = WX + B**

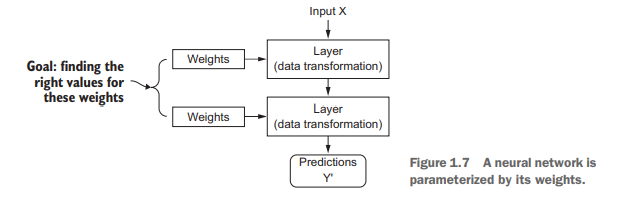
**Here X = Input**

**Y= Output**

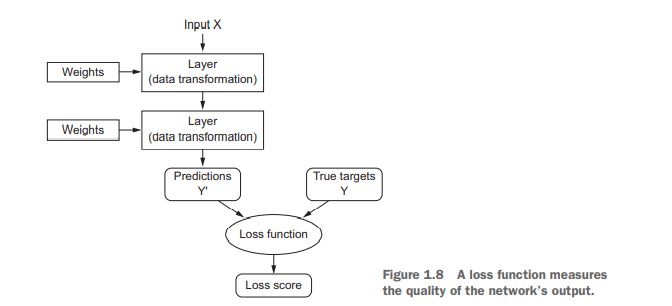
**W= Coefficient**

**B= Interception**

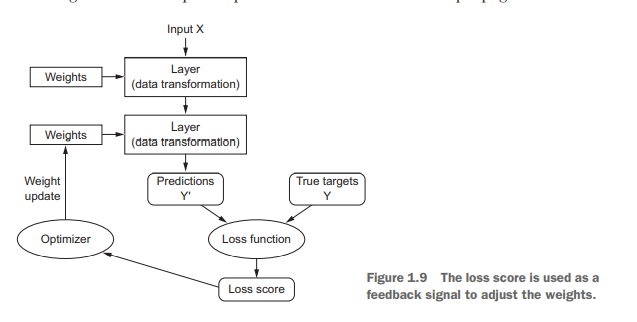
**Neural network:**

****

**Loss function:**

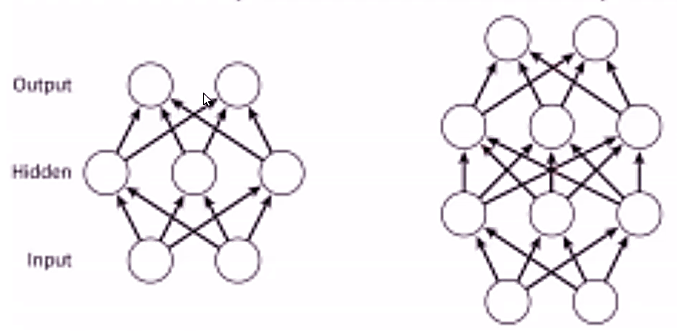
****

**Backpropagation algorithm:**

****

**Shallow Learning:**

The features extraction in Shallow Machine Learning is a manual process that requires domain knowledge of the data that we are learning from

****In other words, "Shallow Learning" is a type of machine learning where we learn from data described by pre-defined features.

While in "Deep Learning" the feature extraction is algorithmically computed without manual human intervention

**Deep learning:** Deep learning isn’t always the right tool for the job—sometimes there isn’t enough data for deep learning to be applicable, and sometimes the problem is better solved by a different algorithm. If deep learning is your first contact with machine learning, then you may find yourself in a situation where all you have is the deep-learning hammer, and every machine-learning problem starts to look like a nail. The only way not to fall into this trap is to be familiar with other approaches and practice them when appropriate.

**Probabilistic modeling (**text classification)

Posterior probability p(x/y) vs Bay’s rule p(x/y) = (p(x/y) p(y))/p(x)

**Naive Bayes algorithm: (Used for classification)**

Naive Bayes is suitable for solving multi-class prediction problems. If its assumption of the independence of features holds true, it can perform better than other models and requires much less training data. Naive Bayes is better suited for categorical input variables than numerical variables. Predict the probability of different class based on various attributes. This algorithm is mostly **used** in text classification and with problems having multiple classes.

**logistic regression (Hello World):** (Classifier)

Logistic Regression, also known as Logit Regression or Logit Model, is a mathematical model used in statistics to estimate (guess) the probability of an event occurring having been given some previous data. Logistic Regression works with binary data, where either the event happens (1) or the event does not happen (0).

categorical variable means (yes/no, true, false)

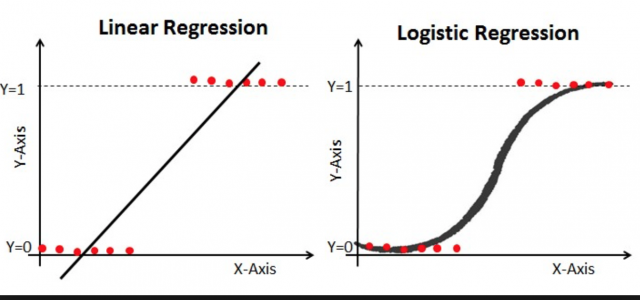
**Sigmoid Function y = 1 / (1+e-x ) between 0 to 1 ,0.5 center (Unclassifiable), Data Set Should Be free from Missing Values**

**Data Point**

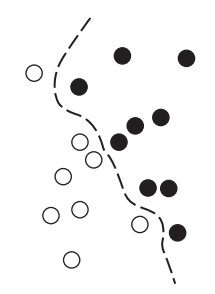
**30-50**

**60- 100 (Binary)**

**Sklearn**

**Used in : Fraud Detection, Diseases Detection, Email Spam or no Spam**

**Kernel methods:**

A kernel function is a computationally tractable operation that maps any two points in your initial space to the distance between these points in your target representation space, completely bypassing the explicit computation of the new representation. Kernel functions are typically crafted by hand rather than learned from data—in the case of an SVM, only the separation hyperplane is learned.

SVMs aim at solving classification problems by finding good decision boundaries between two sets of points belonging to two different categories. A decision boundary can be thought of as a line(hyper line ) or surface separating your training data into two spaces corresponding to two categories. To classify new data points, you just need to check which side of the decision boundary they fall on.

But SVMs proved hard to scale to large datasets and didn’t provide good results for perceptual problems such as image classification. Because an SVM is a shallow method, applying an SVM to perceptual problems requires first extracting useful representations manually (a step called **feature engineering**), which is difficult and brittle.’

**Decision Tree :**

Most used ML algorithm (Hierarchical model ) in which every node splits into branches against a rule. They are able to capture Linear and Non-linear Relations in data. Another name of decision tree is CART (Classification and regression tree).

Random Forest : (CART : Classification and regression tree)

Group of Decision trees algorithm to vote class to measure it.(by combining lots of tree)

Boosting Machines.

We give data to it to perform learning when it make mistake first time then it is ot make the mistake again. When convert weak learner to strong learner. (Not Used now a days after neural networks)

**Algorithms used in ML now a days are :**

* **Random Forest**
* **SVM**
* **Probalistic Model**

**Explore Kaggle**

**Important Note:**

Gradient boosting machines, for shallow Learning problems; and deep learning, for perceptual problems. In technical terms, this means you’ll need to be familiar with XGBoost and Keras—the two libraries that currently dominate Kaggle competitions.

**Why deep learning? Why now?**

why did deep learning only take off after 2012? What changed in these two decades? In general, three technical forces are driving advances in machine learning:

* **Hardware**
* **Datasets and benchmarks**
* **Algorithmic advances**

**Pending:**

**Explore:**

**Adam**

**One hot Encodind**

**Precision vs accuracy vs recall**