Introduction to Machine Learning

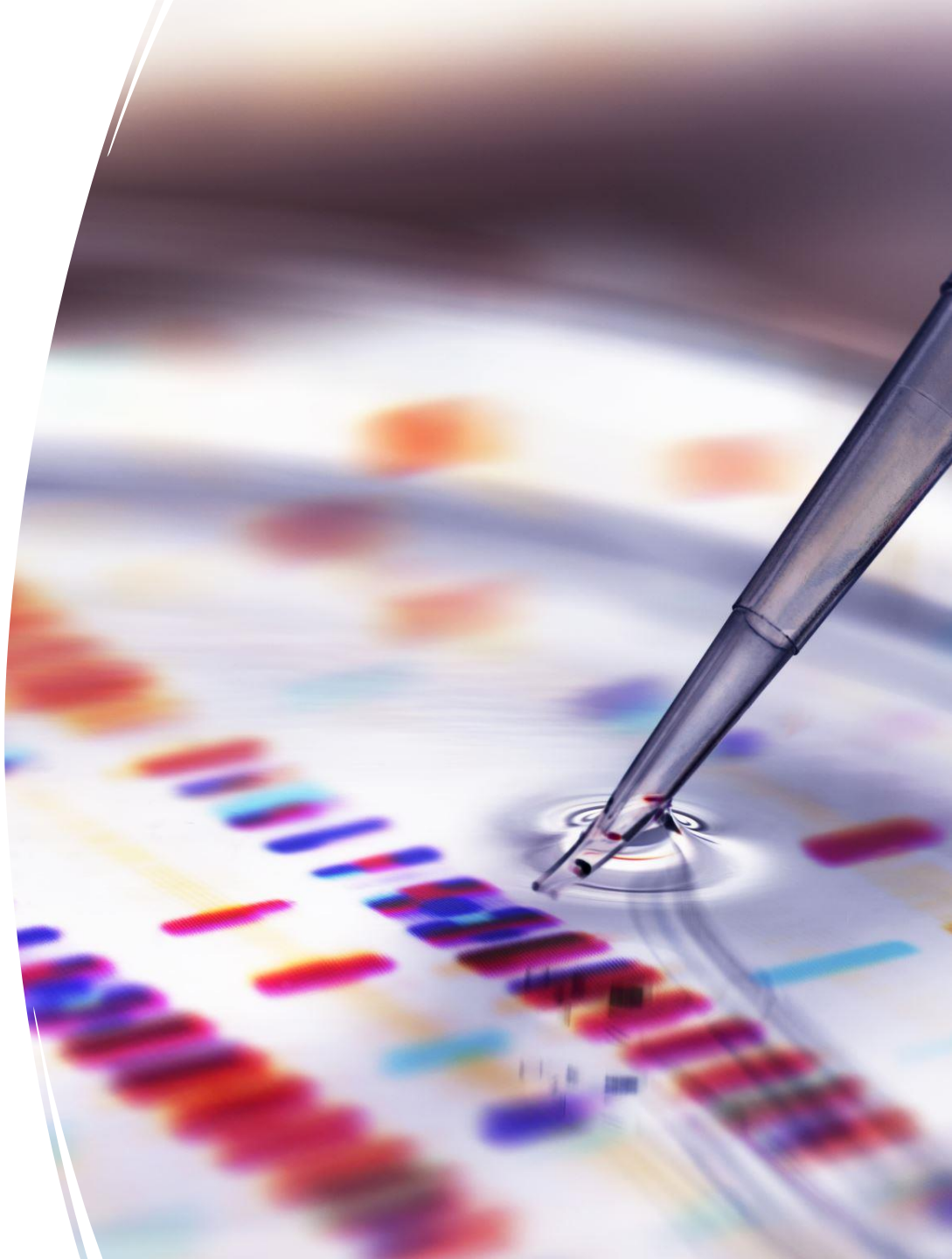
- Dr.A. BAZILA BANU
- PROFESSOR & Head
- AIML

Machine learning

- Machine Learning is the science of teaching machines how to learn by themselves.

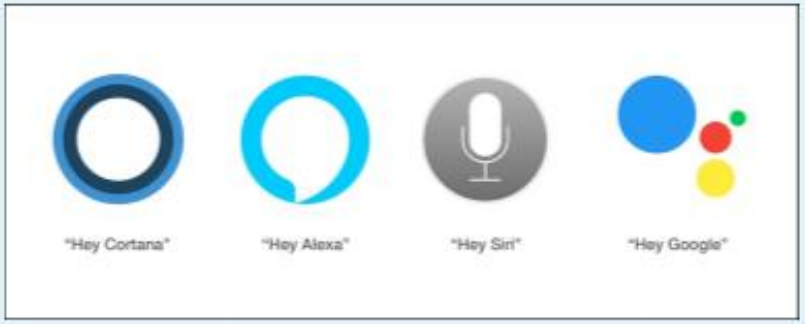
- Automatically learn and improve from experience without being explicitly programmed.

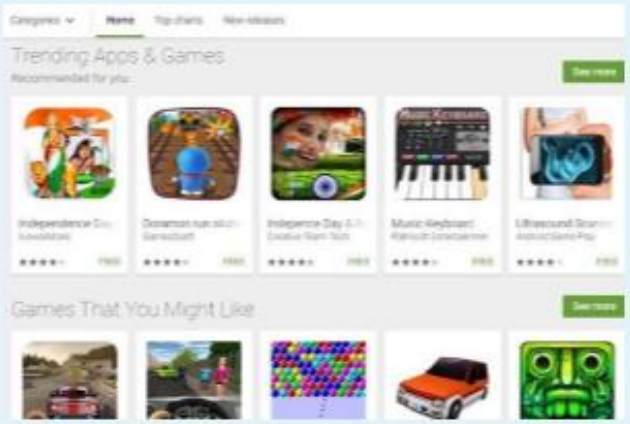
Applications of ML



Machine Learning Use Case in Smart Phones

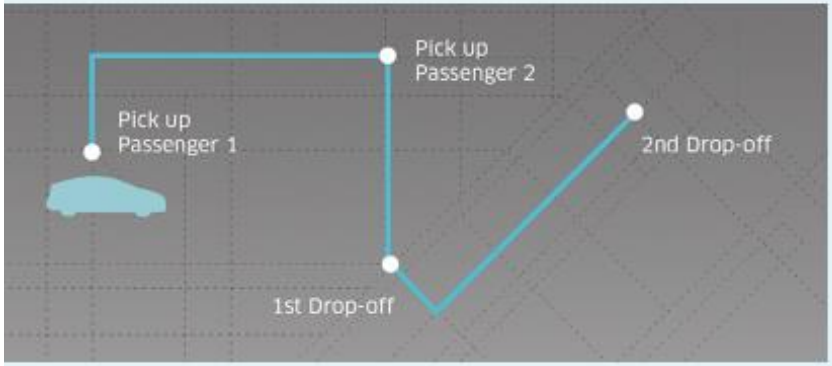
- Voice Assistants
- Smartphone Cameras
- App Store and Play Store Recommendations
- Face Unlock





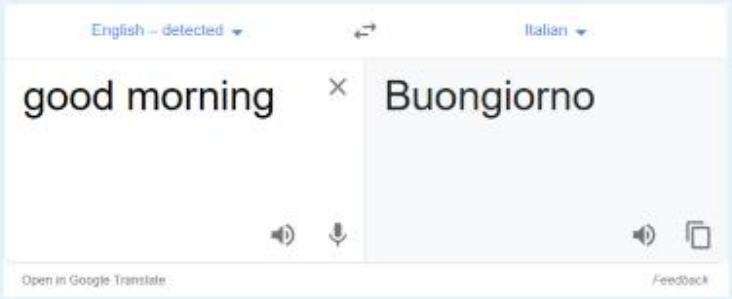
Applications in Transportation

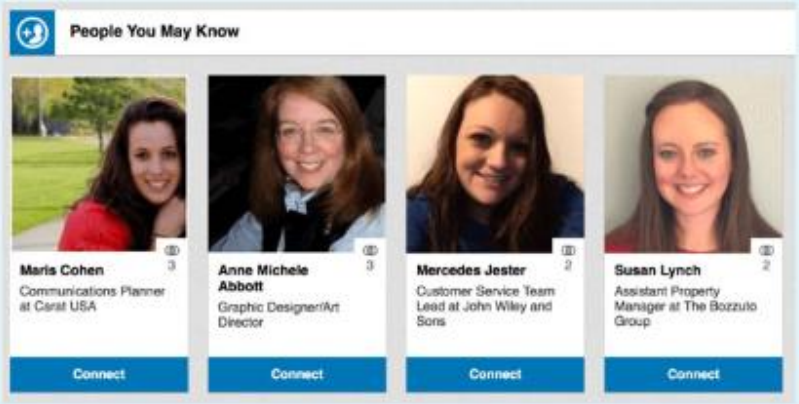
- Dynamic Pricing in Travel
- Transporting and Commuting - Uber
- Google Maps



Applications in Web Services

- Email Filtering
- Google Search
- Google Translate
- Facebook and LinkedIn Recommendations





Machine Learning Use Cases in Sales and Marketing

- Recommendation Engine
- Personalized Marketing
- Customer Support (Chatbots)

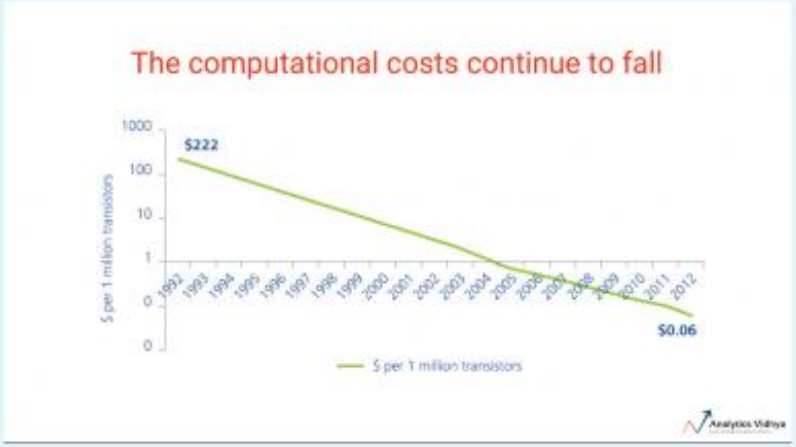
Machine Learning Use Cases in Financial Domain

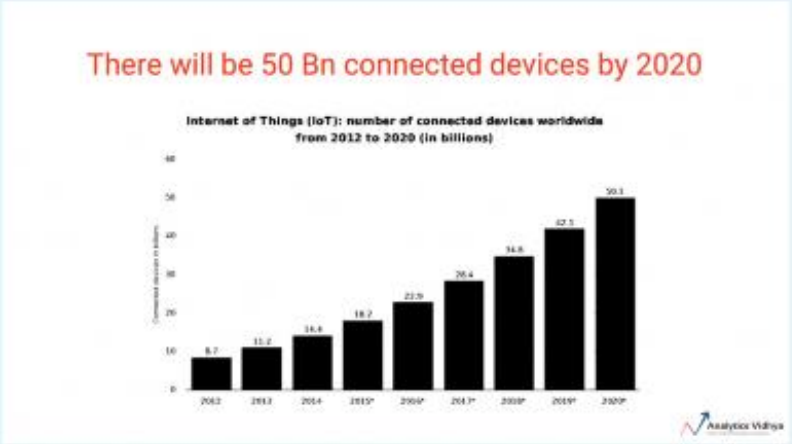
- Fraud Detection
- Personalized Banking

Why is Machine Learning Getting Attention Recently?

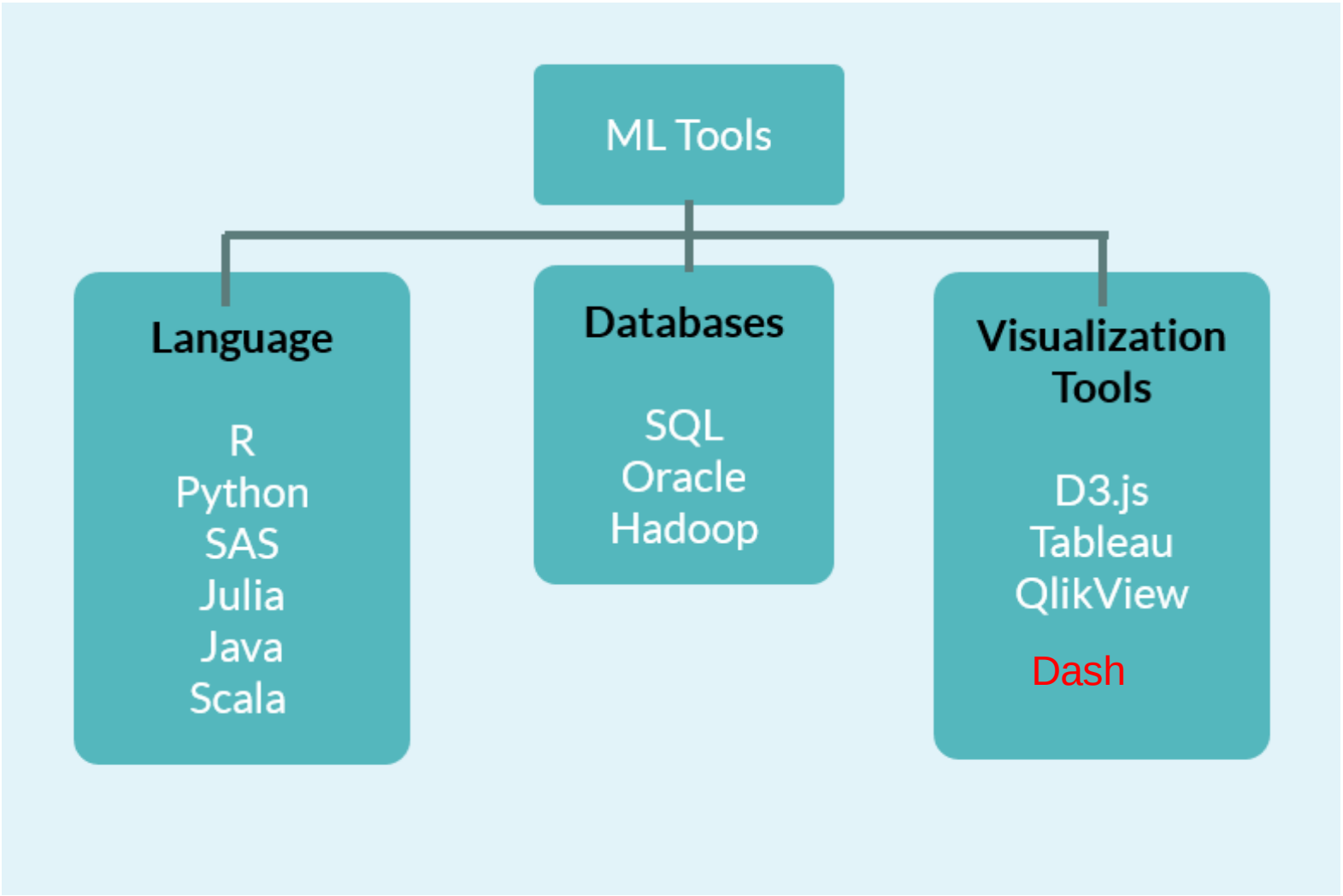
This development is driven by a few underlying forces:

- The amount of data generation is increasing significantly with reduction in the cost of sensors
- The cost of storing this data has reduced significantly
- The cost of computing has come down significantly
- Cloud has democratised Compute for the masses



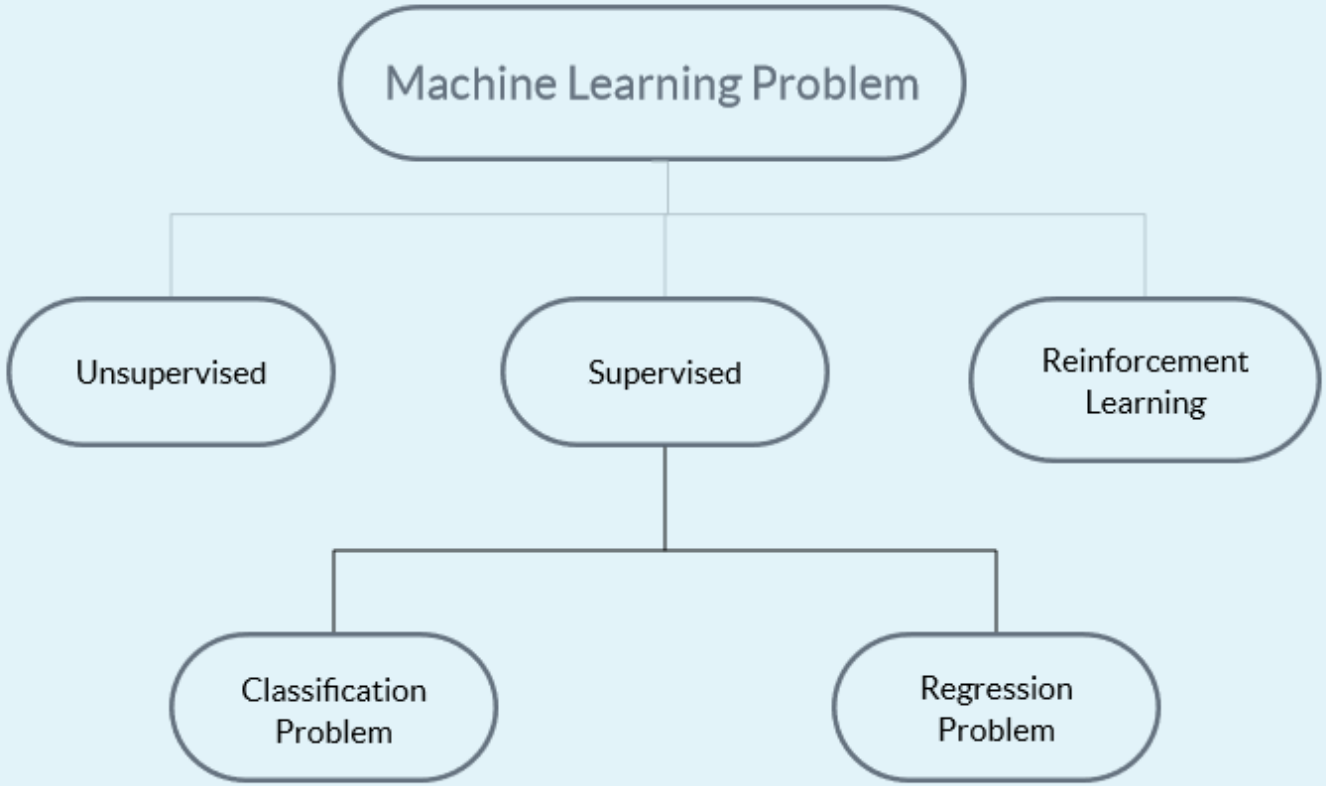


Tools Used in Machine Learning



Problems Solved Using Machine Learning

Machine Learning problems can be divided into three broad classes:



Machine Learning -Types

- **Supervised learning**
  - discover patterns in the data that relate data attributes with a target (class) attribute.
  - These patterns are then utilized to predict the values of the target attribute in future data instances.
  - Supervised learning assumes that training examples are classified (labeled by class labels)
- **Unsupervised learning**
  - The data have no target attribute.
  - We want to explore the data to find some intrinsic structures in them.
  - Unsupervised learning concerns the analysis of unclassified examples.

Supervised learning-Example

- you get bunch of photos **with information what is on them** and you train a model to recognize new photos
- you have bunch of molecules and **information which are drugs** and you train a model to answer whether new molecule is also a drug

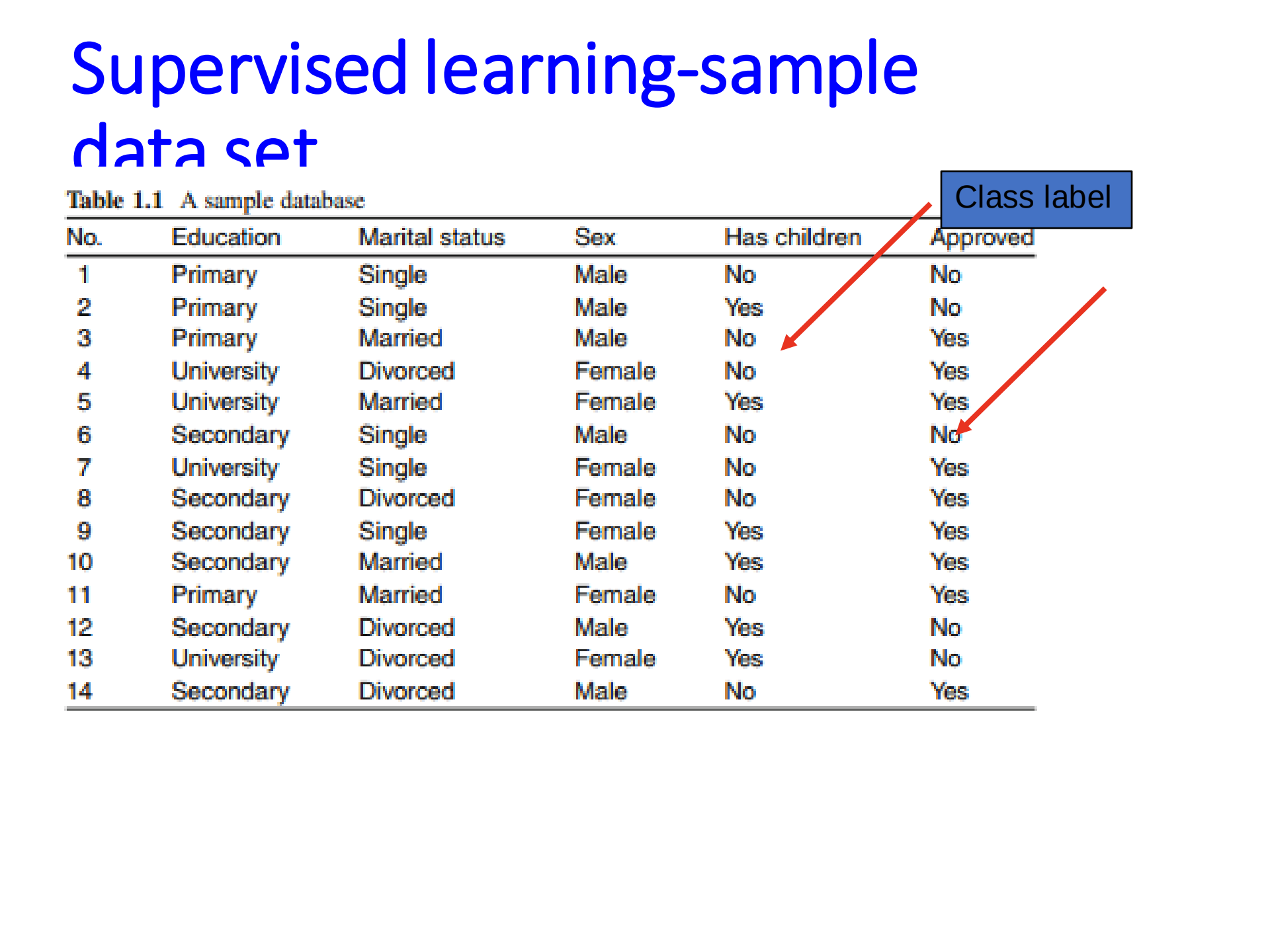
Unsupervised learning-Example

- you have bunch of photos of 6 people but **without information** and want to **divide** this dataset into 6 piles, each with photos of one individual

Supervised learning-sample data set

Table 1.1 A sample database

No.	Education	Marital status	Sex	Has children	Approved
1	Primary	Single	Male	No	No
2	Primary	Single	Male	Yes	No
3	Primary	Married	Male	No	Yes
4	University	Divorced	Female	No	Yes
5	University	Married	Female	Yes	Yes
6	Secondary	Single	Male	No	No
7	University	Single	Female	No	Yes
8	Secondary	Divorced	Female	No	Yes
9	Secondary	Single	Female	Yes	Yes
10	Secondary	Married	Male	Yes	Yes
11	Primary	Married	Female	No	Yes
12	Secondary	Divorced	Male	Yes	No
13	University	Divorced	Female	Yes	No
14	Secondary	Divorced	Male	No	Yes



Supervised learning- Techniques

- **Prediction (Estimation):**
  - Binary Classification
  - Multilevel Classification
  - Interval Prediction
  - Regression
- **Classification:**
  - *k*-Nearest Neighbors
  - Naïve Bayes
  - Bayes Network
  - Decision Trees
  - Discriminant Analysis
  - --Logistic Regression
  - --Neural Nets

UnSupervised learning-Techniques

- Cluster Analysis
- Principle Components
- Association Rules
- Collaborative Filtering

Supervised learning-(Prediction)

Binary Classification

- special case in which there are only two classes

Multilevel Classification

-More than two classes

Interval Prediction

-Estimate of an interval in which future observations will fall, with a certain probability, given what has already been observed. Prediction intervals are often used in regression analysis.

Regression

-statistical process for estimating the relationships among variables.

Linear Regression

- Regression analysis generates an equation to describe the statistical relationship between one or more predictor variables and the response variable.

It consists of 3 stages:

1) analyzing the correlation and directionality of the data.

2) Estimating the model, i.e., fitting the line

3) Evaluating the validity and usefulness of the model.

Applications of Regression Analysis

- There are three major uses for Regression Analysis:

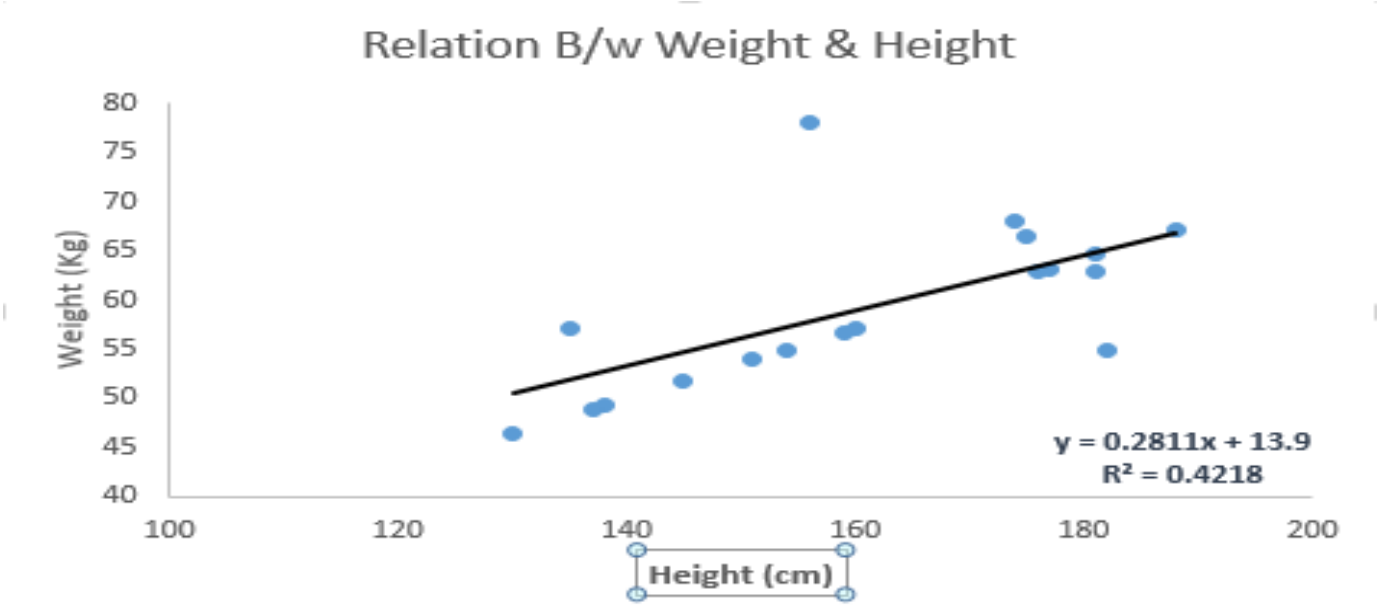
  1) causal analysis,

  2) forecasting an effect

  3) trend forecasting.

Linear Regression

Linear Regression establishes a relationship between **dependent variable (Y)** and one or more **independent variables (X)** using a **best fit straight line** (also known as regression line)



It is represented by an equation **Y=a+b*X + e**, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

ANOVA

- Analysis of variance (**ANOVA**) is a collection of statistical models used to analyze the differences among group means and their associated procedures (such as "variation" among and between groups), developed by statistician and evolutionary biologist Ronald Fisher.
-

Linear Regression Interpretation

- **P-value**

❖ The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect).

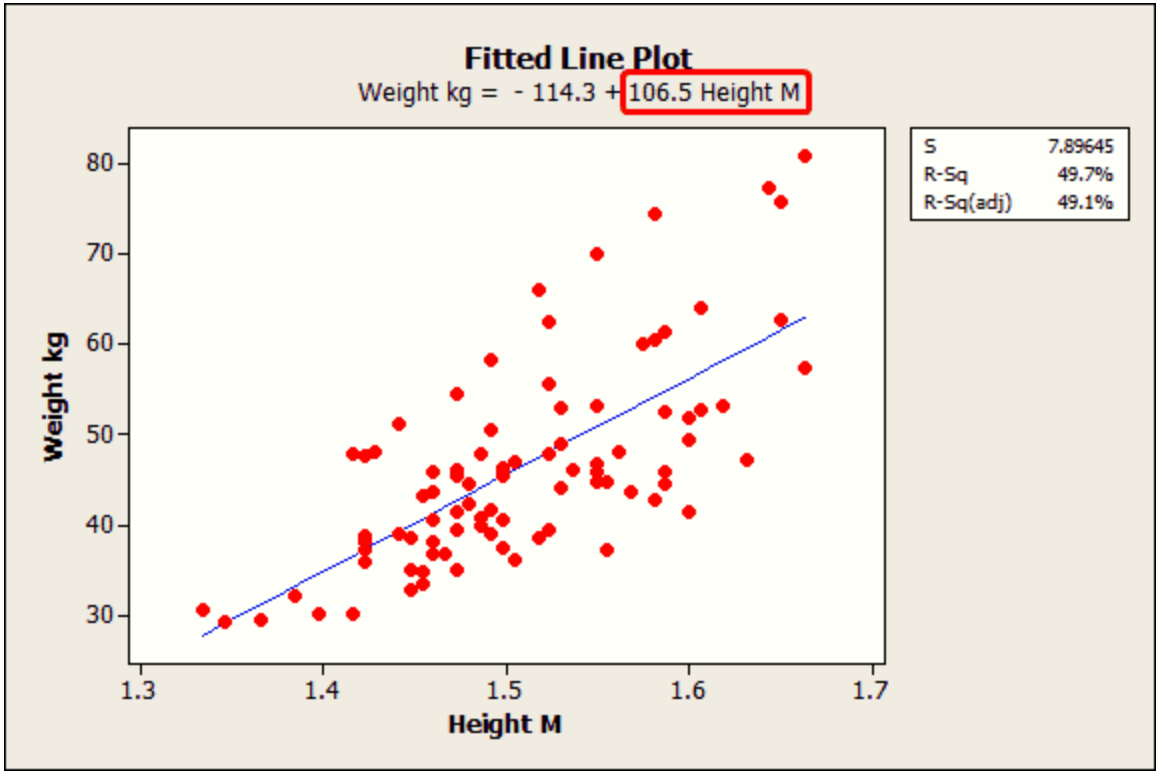
❖ A low p-value (< 0.05) indicates that you can reject the null hypothesis.

❖ In other words, a predictor that has a low p-value is likely to be a meaningful addition to your model because changes in the predictor's value are related to changes in the response variable.

- **Regression Coefficients**

Regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant

Regression Coefficients



Linear Regression Calculation

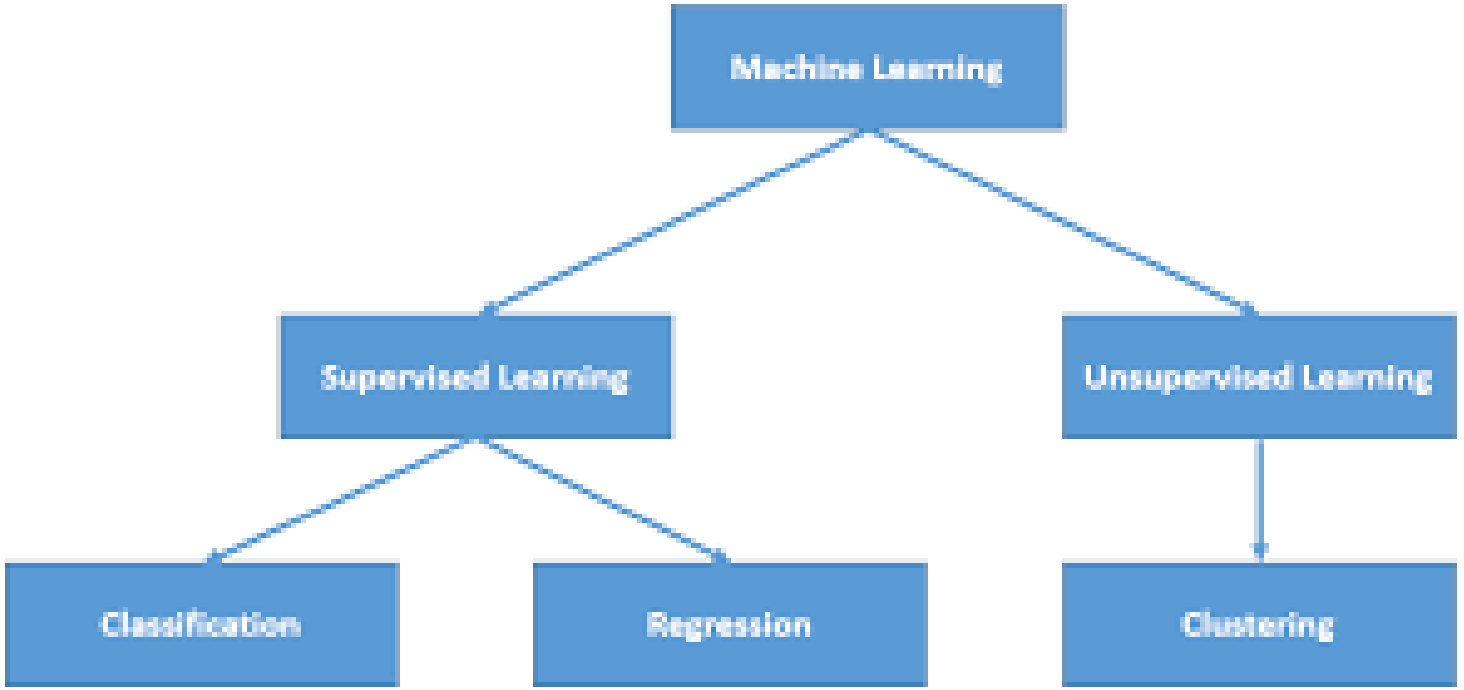
- http://www.statisticshowto.com/how-to-find-a-linear-regression-equation/

Error Metric for Linear Regression

1.Root Mean Squared Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

Mind MAP


Machine Learning
Supervised Learning
Unsupervised Learning
Classification
Regression
Clustering

Formative Assessment question

- You go bag-packing to a new country, you did not know much about it - their food, culture, language etc. However from day 1, you start making sense there, learning to eat new cuisines including what not to eat, find a way to that beach etc.
- Scenario 2 is an example for
- a.Semi Supervised
- **b.Supervised**
- c.Unsupervised
- d. Neural Network

Stimulating question

- 1.From ---------- technique, data is labeled and the algorithms learn to predict the output from the input data.
- a.Semi Supervised
- **b.Supervised**
- c.Unsupervised
- d. Neural Network