Introduction To Machine Learning

Prerna Preeti
Software Developer, IBM

About Me

I am a Master of Science (CS) graduate from Indiana University ,Bloomington. Coming from the capital city of Patna, Bihar I work as a Software Development Engineer in IBM Data & AI organization in the San Francisco Bay Area, California. I am currently working on AI powered projects and I am eager to help others working on projects related to Machine Learning and AI. Before moving to the United States, I also worked on data warehousing and ETL technologies.

Introduction to Machine Learning

- What is Machine Learning?
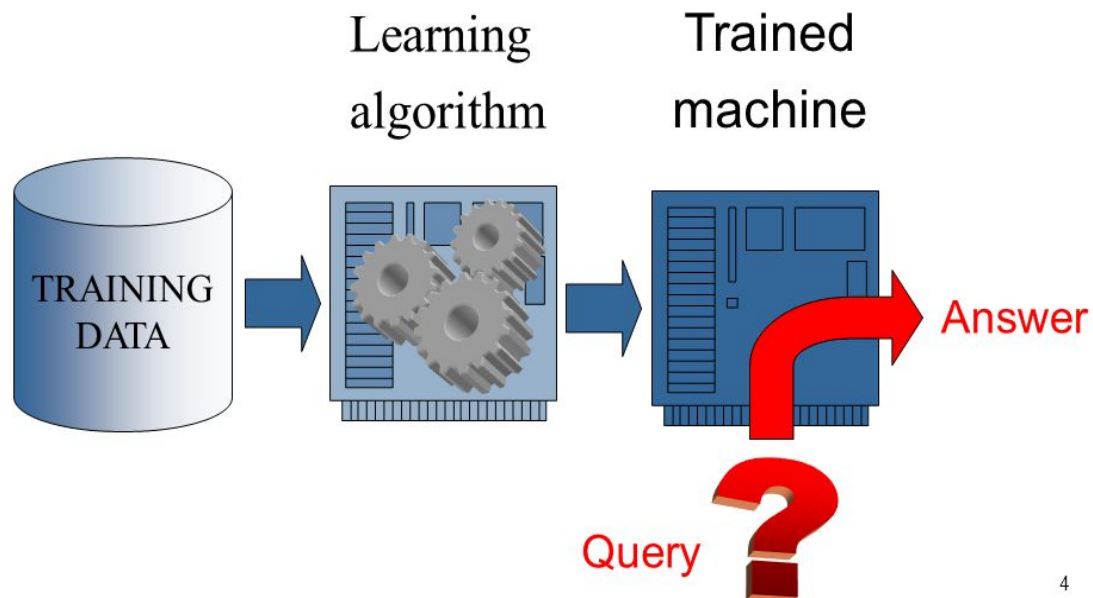
Introduction to Machine Learning

- We store and process data

- Example Domain: Educational Institutions
  - Hundreds of institutes in the country
  - Tailor their educational content for competitive entrance exams
  - Record the details : date, student details, who cracked the exams and their scores, where did they go to school , what was the training method , how often they tested for skills, etc.
  - Gigabytes of data is generated everyday
  - Turn this data into information for prediction

Introduction to Machine Learning

- Do we know who is more likely to crack IIT?
- Which training method works best for most students?

Introduction to Machine Learning

- Answer is : NO

- Can learning from the data help us to predict the answer to our questions? : YES

Introduction to Machine Learning

- We can collect data

- Try to extract answers to these similar questions

- There is a process to explain the data we observe and it is not completely random

- Random students do not crack IIT

- There are patterns in the data we need to identify

Introduction to Machine Learning

- Niche of Machine Learning

    – We cannot identify the process completely

    – We can construct good approximation

    – Detect certain pattern and irregularities

    – Use these patterns for predictions assuming the near future will not be much different than the past

- "*Machine Learning is programming computers to optimize a performance criterion using example data or past experience*"

  - Uses theory of statistics to build mathematical models
  - Core task is to make inferences from a sample

Introduction to Machine Learning

- Role of Computer Science

  - To store and process massive amounts of data

  - Training to solve optimization problems

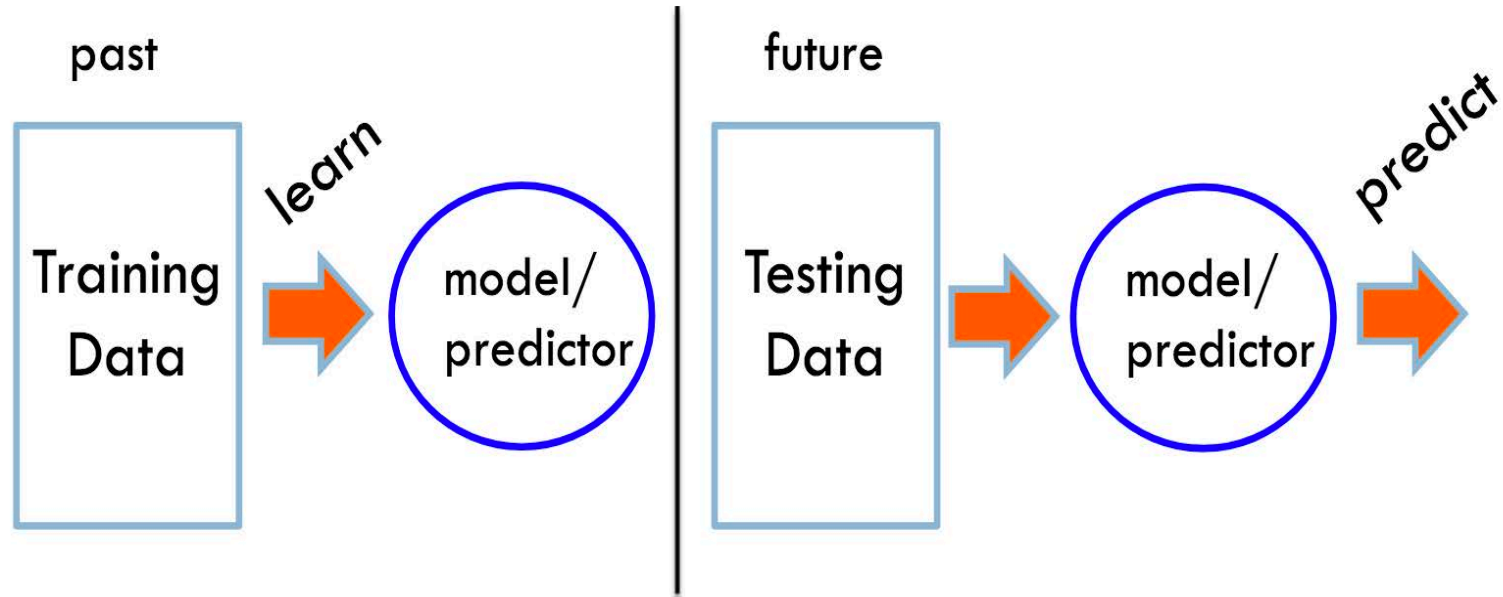 - Representation of data and algorithmic solutions for inferences

Training, Testing and Performance

- Training is the process of making the system able to learn

- Several factors affect the performance of the model
  - Algorithm used for training since they extract useful information from training examples
  - Feedback provided

# Training, Testing and Performance

Machine Learning Algorithms

- Supervised Learning
  - Classification
  - Regression

- Unsupervised Learning
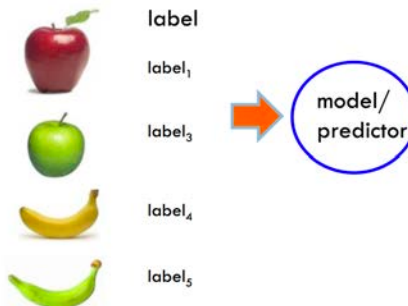 - Clustering
  - Learning Association

Supervised Learning

- There is an input and an output label
- Learn mapping from input to output with data samples
- Model defined up to a set of parameters:
    - $y = g(x \mid P)$
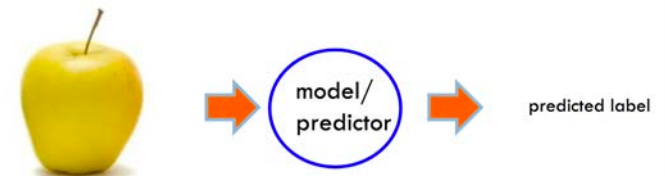   - $g(.)$ is the model and P are its parameters
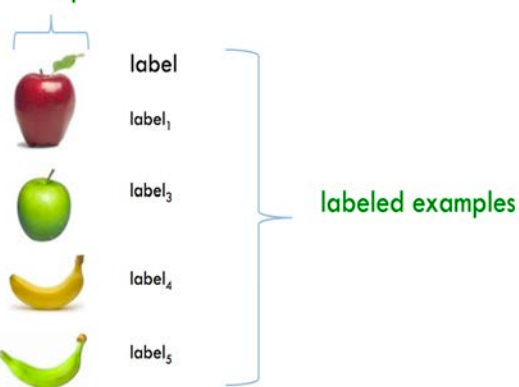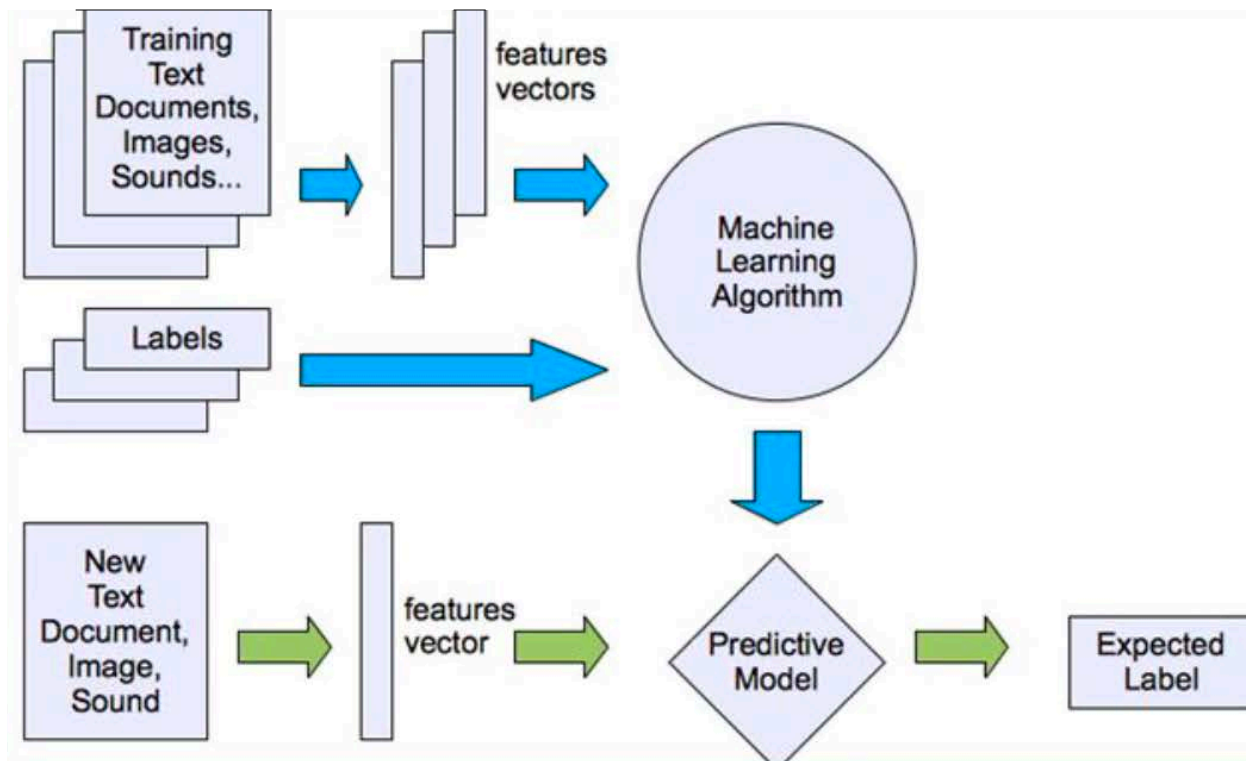

Learn from labeled data   Define the model   Model predicts new samples

Supervised Learning

Classification

- Educational Institute Example – IIT Analysis
  - Predict who will potentially crack IIT
  - Targeting the right students can help make profit and reputation
  - Finding the right training method can help more students

- Calculating Probability to crack competitive exams
  - Calculate the probability based on recorded student data and their past scores
  - Form a rule
  - Fit a model to the past data and calculate the probability for a new student

Classification

- Classes
  - High Probability (H)
  - Low Probability (L)
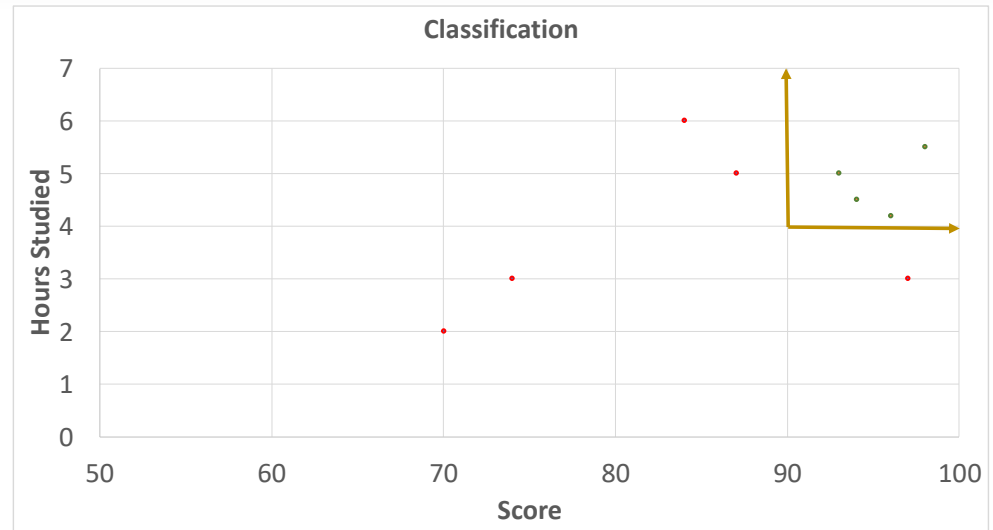
- Rule(Prediction)
  - IF CBSE/ICSE/BB score > 90% AND average daily hours spent on IIT exam practice > 4 THEN "H" ELSE "L"
 - Example of discriminant

Regression

- Trying to map input to Real valued output labels
- Lets say we want to predict car prices

| ##		Price	Age	KM
##	1	13500	23	46986
##	2	13750	23	72937
##	3	13950	24	41711
##	4	14950	26	48000
##	5	13750	30	38500

Regression

- Data transformation into numeric values
- Find a linear pattern and predict
- As you can see from graph the price(P) decreases with a so lesser the age of car more expensive it is
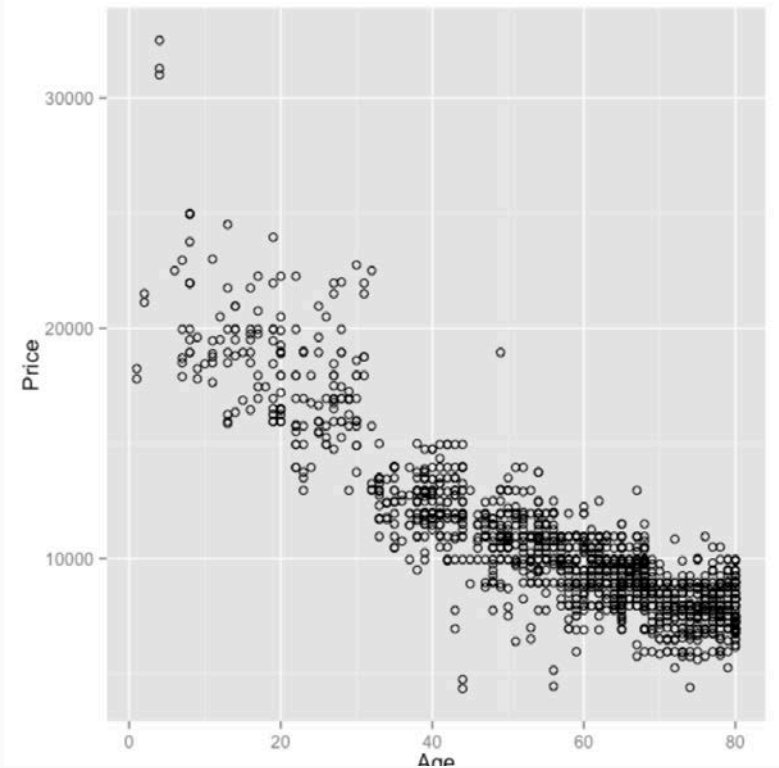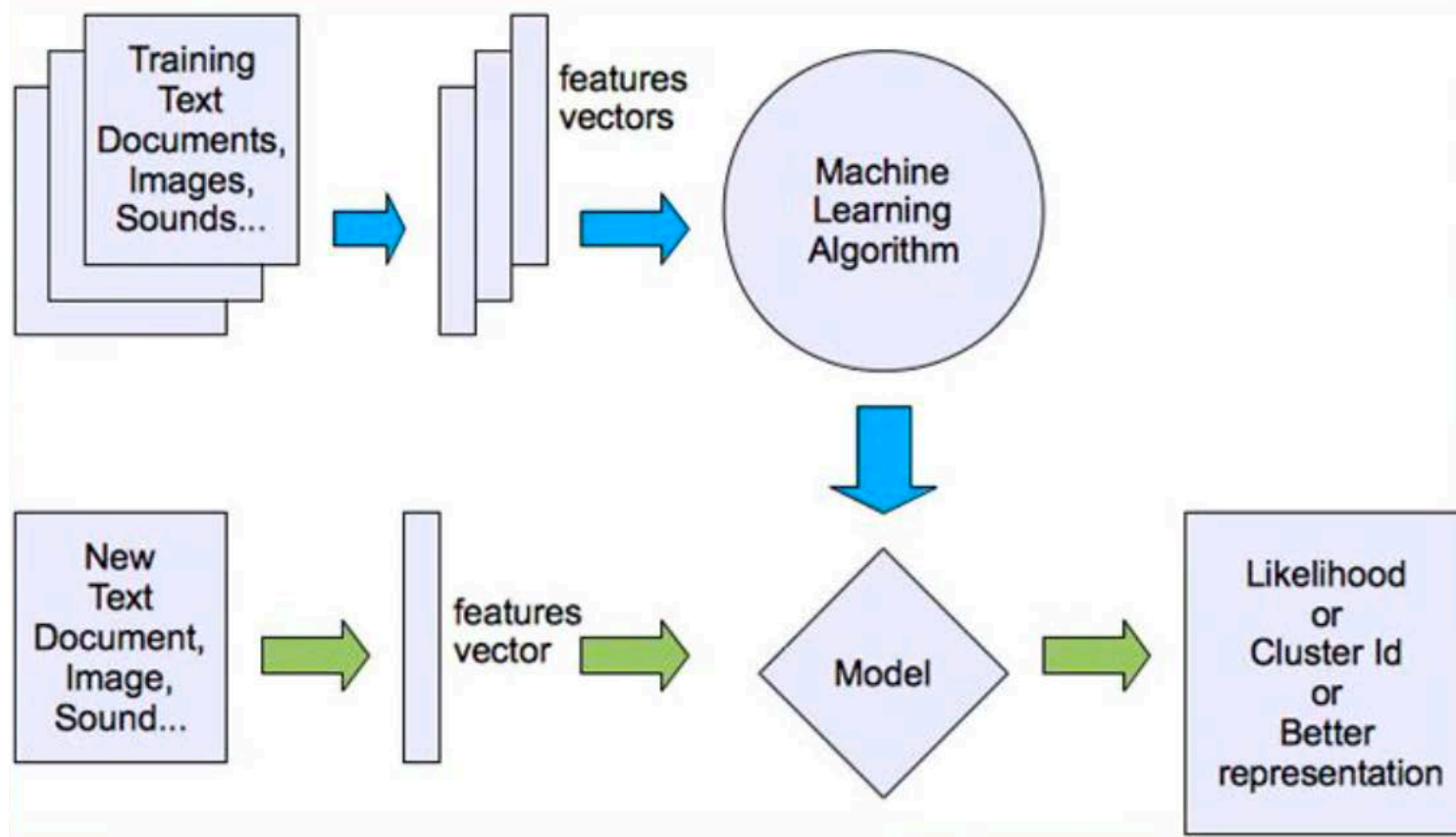
Regression

- And you can see from this graph that the price(P) decreases the more the car is driven i.e. more KMs it has less expensive it is!

- So we can come up with a linear regression model which can tell us the price of the car based on its age and KMs driven
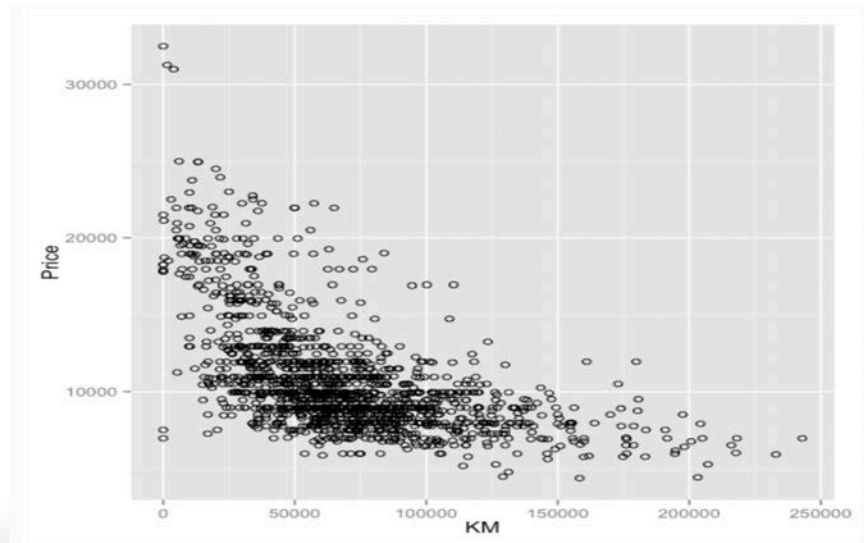
Unsupervised Learning

Learning Associations

- Retail – Cart Analysis

    – Finding associations between products bought by customers
    – If a customer buys X, they typically buy Y
    – To find potential customers for product Y, target them from cross selling

Learning Associations

- **Retail – Cart Analysis**
  - Finding associations between products bought by customers
  - If a customer buys X, they typically buy Y
  - To find potential customers for product Y, target them from cross selling

- **Association Rule**
  - $P(Y|X)$ where Y is the product we condition on X and X is the product the customer has already purchased
  - $P(Coke|Pizza) = 0.8$ then 80% who buy Pizza also buy Coke

- **Distinction Attribute**
  - $P(Y|X,D)$ D is set of customer attributes like gender, age etc.

Thank you!