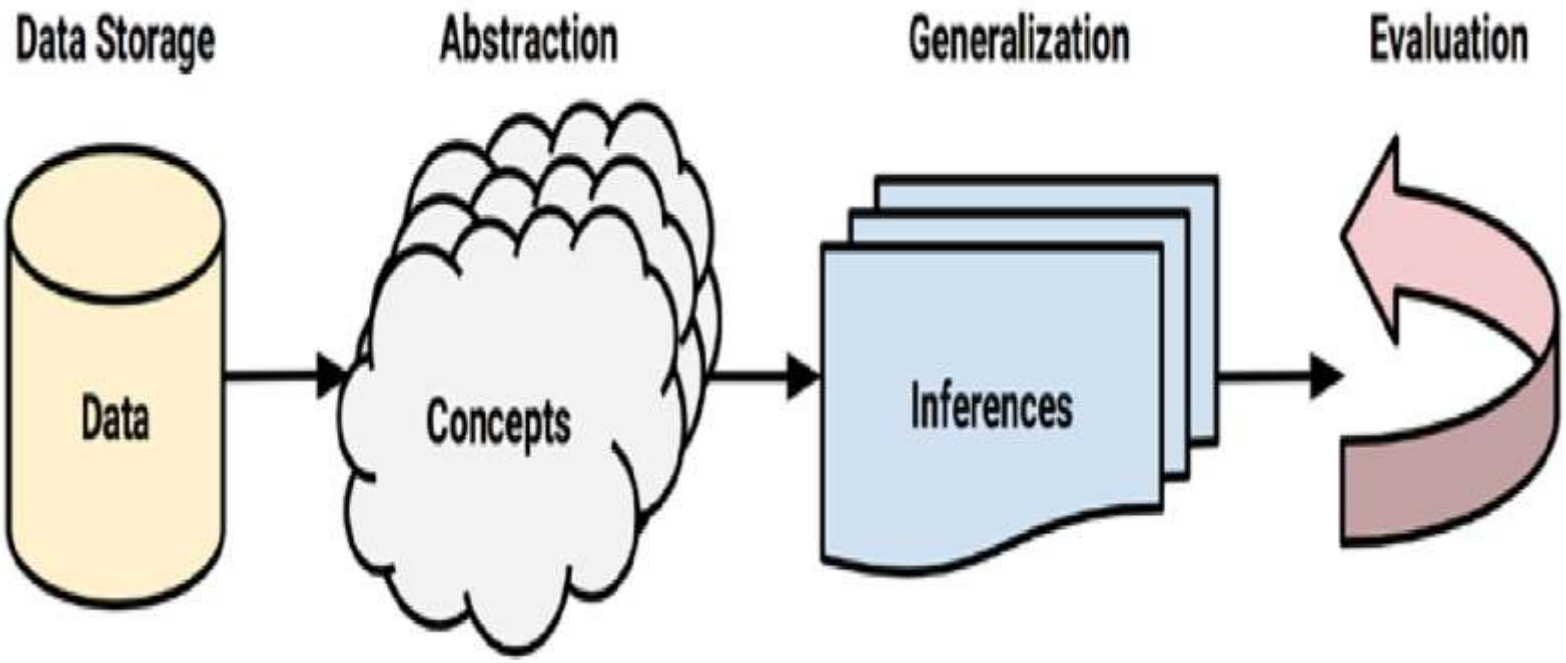
MACHINE LEARNING

How machines learn

- According to Tom M Mitchel

  A machine learns whenever it is able to utilize an experience such that its performance improves on similar experience in future
- Data storage
- Abstraction
- Generalization
- Evaluation

The following figure illustrates the steps in the learning process:



Data storage

- Utilizes observation, memory and develop strategies to relate the stored information
- Data stored on different types of memory, RAM with CPU

Abstraction

- Translation of stored data into broader representation and concepts
- Assigning meaning to stored raw data
- Identifying unseen relationship in the data
- Basis of Knowledge representation results in the formation of logical structures that turn raw sensory information into a meaningful form

- During knowledge representation raw data is summarized using a model which is an explicit description of patterns within the data
- Models include
  - Mathematical equations
  - Relational diagrams such as trees and graphs
  - Logical if/else rules
  - Grouping of data called clusters
- Process of fitting the model to a dataset is called training
- Model summarizes the original information

Observations → **Data** → **Model**

Distance	Time
4.9m	1s
19.6m	2s
44.1m	3s
78.5m	4s

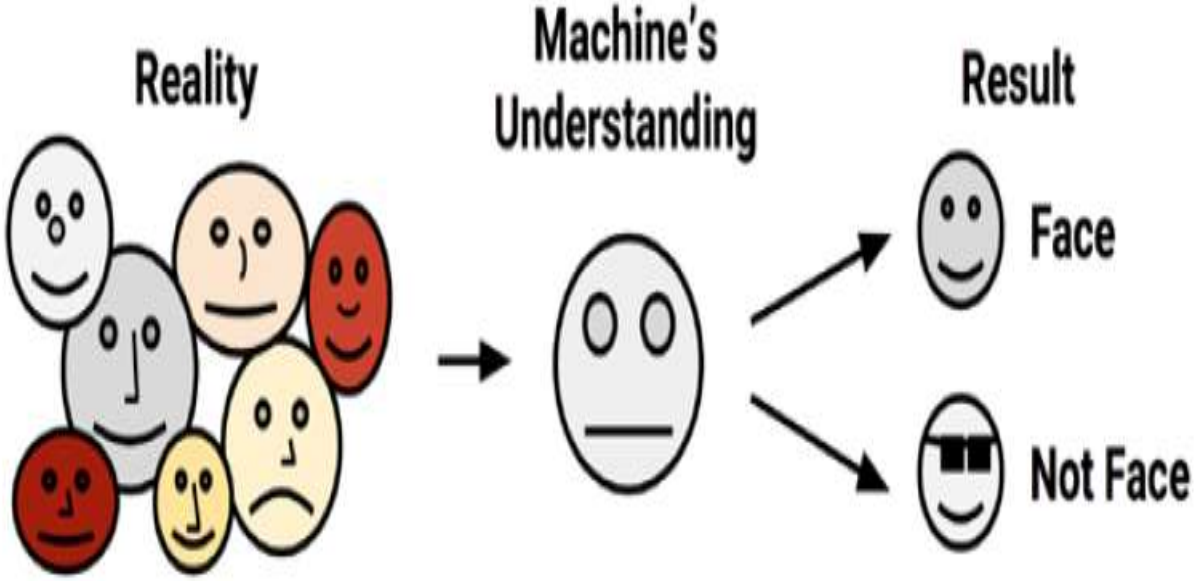
$$g = 9.8m/s^2$$

Generalization

- Using the abstracted knowledge for future action on similar tasks as seen during training
- Limiting the patterns to those that will be most relevant to its future tasks
- May be biased resulting in wrong conclusions
- Machine learning algorithms choose among the countless ways to understand a set of data
- Algorithms employ heuristics to reduce the search space

Evaluation

- Measure the learner's success in spite of its biases and use this information for additional training if required
- Can be done after the model is trained on initial training data set.
- Evaluation is done on new unseen data
- Fail to perfectly generalize due to noise
  - Unexplainable variation in data


Reality
Machine's Understanding
Result
Face
Not Face

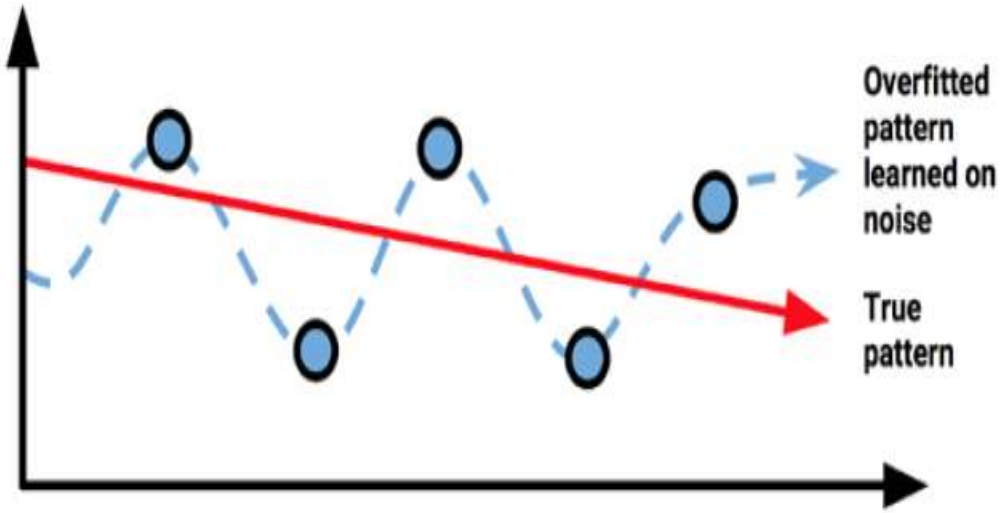
Overfitting in Machine Learning

- refers to a model that models the training data too well.
- Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data. This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. The problem is that these concepts do not apply to new data and negatively impact the models ability to generalize.
-

Underfitting in Machine Learning

- Underfitting refers to a model that can neither model the training data nor generalize to new data.
- An underfit machine learning model is not a suitable model and will be obvious as it will have poor performance on the training data.

overfitting



Machine learning in practice

- Data collection-collecting learning material for the algorithm to generate actionable knowledge
- Data exploration and preparation-prepare the data for learning process-cleaning the data-eliminating unnecessary data
- Model training-pattern identified and model selected

- Model evaluation
  - testing the accuracy of model using test data
  - Identifying how well the algorithm learns from experience
  - Develop measures of performance
- Model improvement
  - More data
  - A different model

Supervised Learning

- **Classification**:- Problem of identifying by which set of categories of new observation belong on the basis of training set of data containing observations.
- Eg:-Consider the following data:-

Score1:	29	22	10	31	17
Score2:	43	29	47	50	18
Result:	pass	fail	fail	pass	fail

- Given a training set of data
- Two attributes Score1 and Score2
- Class label: Result
- Class label has 2 possible value “pass” and “fail”
- The data can be divided into two categories

(1) Set of data for which the class label is pass

(2) Set of data for which the class label is fail.

- If we have some new data ,for eg:score1:25 and score2:35.which value should be assigned to Result?

- To find out the above question using the given data alone, we need to find the rule, formula or method.
- The problem of finding this rule or formula or method is called classification problem.

- Regression:-
- ✓ Problem of predicting the value of a numeric variable based on observed value of variable.
- ✓ Consider the example below:-

Price	Age	Distance	Weight
13500	23	46936	1165
13750	23	72937	1165
13950	24	48711	1165

- Here, we are required to estimate the price of a car aged 25 years with distance 53240km and weight 1200 pounds
- This is an example of regression problem because we have to predict the value of the numeric variable “price”.
- Let ‘x’ denote the set of input variable, y denotes the output variable.

- General approach to regression is to assume a model ie a mathematical relation between x, y and Θ.

$$y=f(x, \Theta)$$

- Aim:-Estimate the value of the dependent variable ‘y’ are as close as possible to correct values given in the training set.

  y=f(x, Θ)

  price=$a_0 + a_1$(age)+a_2(distance)+a_3(weight)

- Regression model can be classified as below:
- ✓ Simple linear regression (y=a+bx)
- ✓ Multivariate linear regression
- ✓ Polynomial regression
- ✓ Logistic regression

- Based on the three different aspects
- ✓ The number and the type of independent variables.
- ✓ Type of dependent variable.
- ✓ Shape of regression line

- Association rule learning:
- ✓ Example of unsupervised Learning
- ✓ Discover interesting relation called association rule between variable in large dataset
- ✓ Example: Consider two sets,

{ Onion, Potato}

{burger}

- We can define an association between the set of products {onion, potato} and the set{burger}. This association is represented in the form of a rule as follows
- { onion,potato}=>{burger}
- Conditional probability =Measure of how likely a customer who has bought onion and potato to buy burger

- In finding an association rule x=>y, we areusing a conditional probability of form p(y|x)

where y is the product the customer may buy and x is the product or set of product the customer has already bought.

- P(y|x,D):To make distinction among customer, we may estimate p(y|x,D), where D is the set of customer attribute like gender,age etc.

- Cluster: The descriptive modeling task of dividing a dataset into homogeneous group is called clustering.

- Apply KNN algorithm to the given data and predict the category of tomato with sweetness is 3,crunchiness is 7(consider k=1).Use manhattan distance formula for calculating the distance.

Ingredient	Sweetness	Crunchiness	Food Type
Grape	8	5	Fruit
Green bean	3	7	Vegetable
Nuts	3	6	Protein
Orange	7	3	Fruit

- Different distance metric used in KNN algorithm, inaddition to Euclidean Distance:

✓ Manhattan distance

✓ Hamming distance

✓ Minkowski distance

✓ Kullback-Leiber (KL) divergence

- The Manhattan Distance between two points **(P1,P2)** and (Q1,Q2) is given by |P1-Q1| + |P2 – Q2|.

- Based on the survey conducted in an institution, students are classified based on the two attributes academic excellence and other activities.Given the following data,identify the classification of a student with X=5,Y=7 using knn algorithm(consider k=3)