Introduction to Machine Learning

Dr. Subrat K Dash

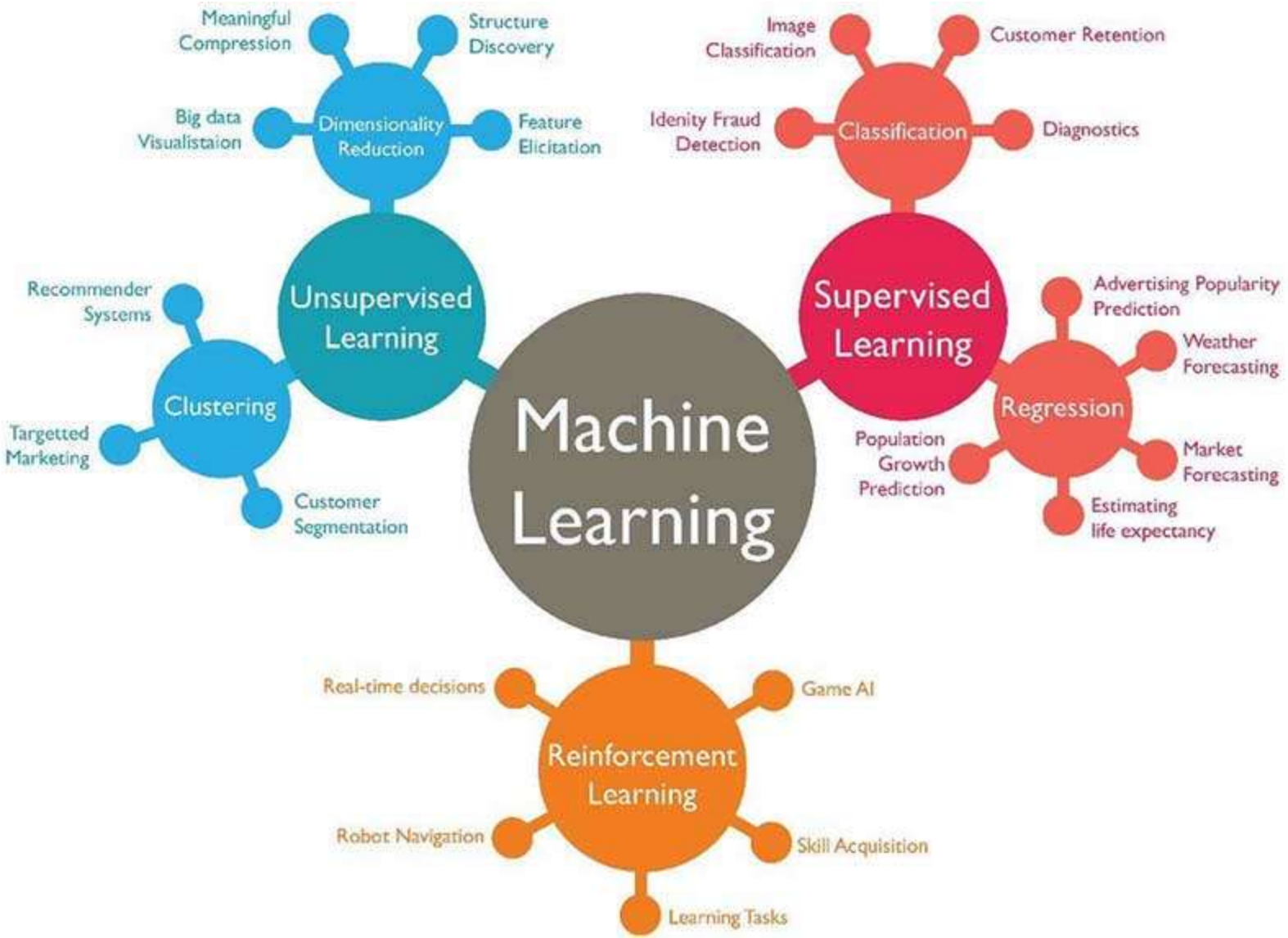
Aanlytics

- Descriptive
  - Describe or summarize raw data.
  - Provide insight into the past.
  - What has happened?
- Predictive
  - Predict what might happen.
  - Understanding the future.
  - What could happen?
- Prescriptive
  - Advise on possible outcomes.
  - What should be done?

Machine Learning

- Coined in 1959 by Arthur Samuel.
- “A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*” – Tom M. Mitchell.
- ML: build mathematical model(s) based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.
- Example domains: Email filtering, Computer vision, Social media, Recommender system etc.

Machine Learning



Supervised Learning

- Given a collection of records (**training set**)
  - Each record is characterized by a tuple (***x***,*y*), where ***x*** is the attribute set and *y* is the class label
    - ***x***: attribute, predictor, independent variable, input
    - *y*: class, response, dependent variable, output
- Task:
  - **Learn a model** (build a classifier) that maps each attribute set ***x*** into one of the predefined class labels *y*

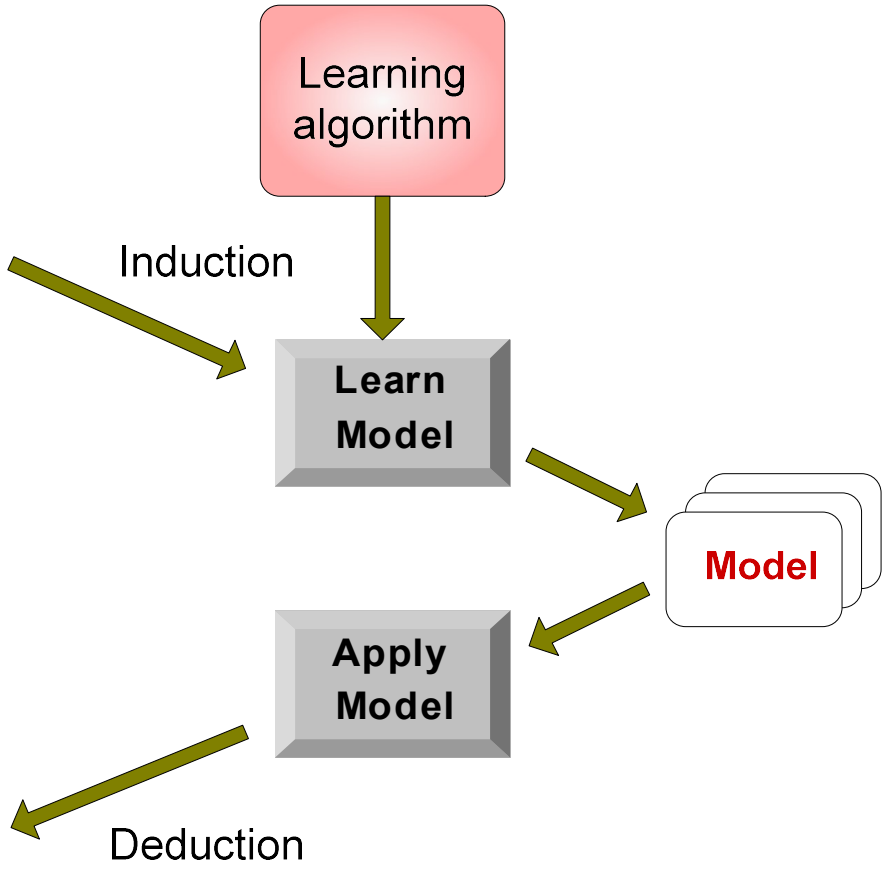
General Approach for Building Classification Model

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Unsupervised Learning

- Given a collection of records, which are unlabelled, learn pattern out of them.
  - Group them according to their similarities/ patterns/ differences.

- Example: Clustering

Reinforcement Learning

- Learning the optimal behaviour in an environment to obtain maximum reward.
  - It is concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward.