Machine Learning

Dr.G.RAVI KUMAR
Assistant Professor
Dept.of Computer Science
Rayalaseema University
Kurnool

- Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data.

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.

- Machine learning is the about the construction and study of systems that can learn from data. This is very different than traditional computer programming.

- **According to Tom Mitchell** "Machine learning is a program can be 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**."

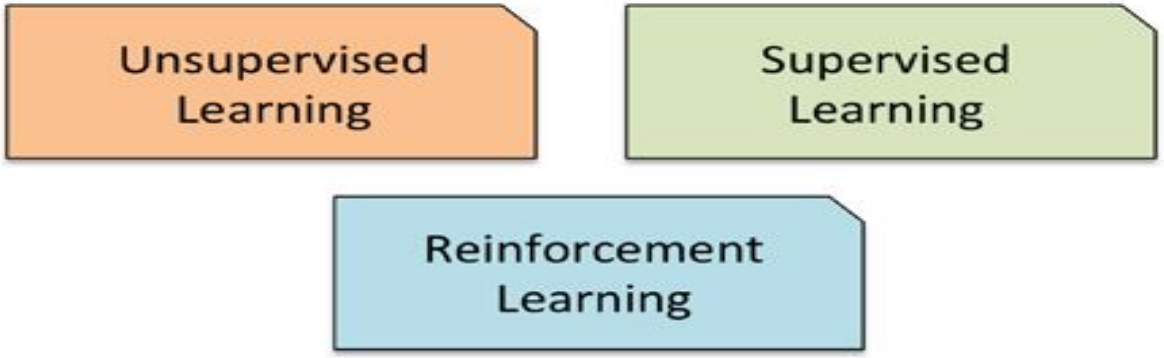
- Machine learning offers a more efficient alternative for capturing the knowledge in data to gradually improve the performance of predictive models, and make data-driven decisions.

- Not only is machine learning becoming increasingly important in computer science research but it also plays an ever greater role in our everyday life.

- In machine learning, data plays an indispensable role, and the learning algorithm is used to discover and learn knowledge or properties from the data. The quality or quantity of the dataset will affect the learning and prediction performance.
- Designing a machine learning approach involves a number of design choices, including choosing the type of training experience, the target function to be learned, a representation for this target function, and an algorithm for learning the target function from training examples.
- Machine learning is the design and study of software artifacts that use past experience to make future decisions; it is the study of programs that learn from data.
- The fundamental goal of machine learning is to generalize, or to induce an unknown rule from examples of the rule's application.

The different types of machine learning

The three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.



In **supervised learning** problems, a program predicts an output for an input by learning from pairs of labeled inputs and outputs; that is, the program learns from examples of the right answers.

- In **unsupervised learning**, a program does not learn from labeled data. Instead, it attempts to discover patterns in the data.

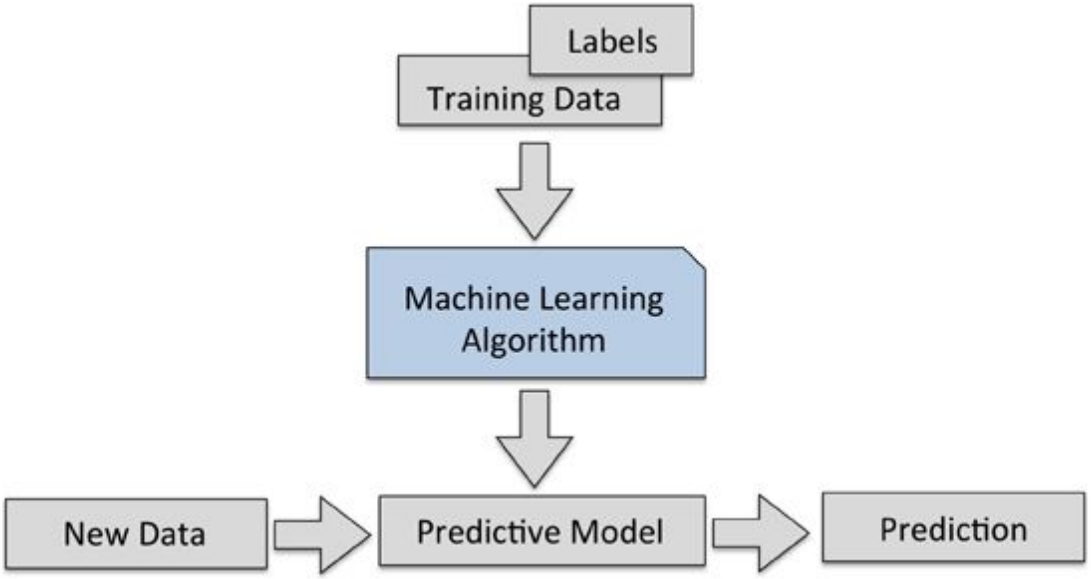
- An example of an unsupervised learning problem is dividing the data points into groups. A program might produce groups that correspond to men and women, or children and adults.

- **Reinforcement learning:** A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle), without a teacher explicitly telling it whether it has come close to its goal or not.

- Another example is learning to play a game by playing against an opponent.

Supervised learning

- The main goal in supervised learning is to learn a model from labeled training data that allows us to make predictions about unseen or future data. Here, the term supervised refers to a set of samples where the desired output signals (labels) are already known.

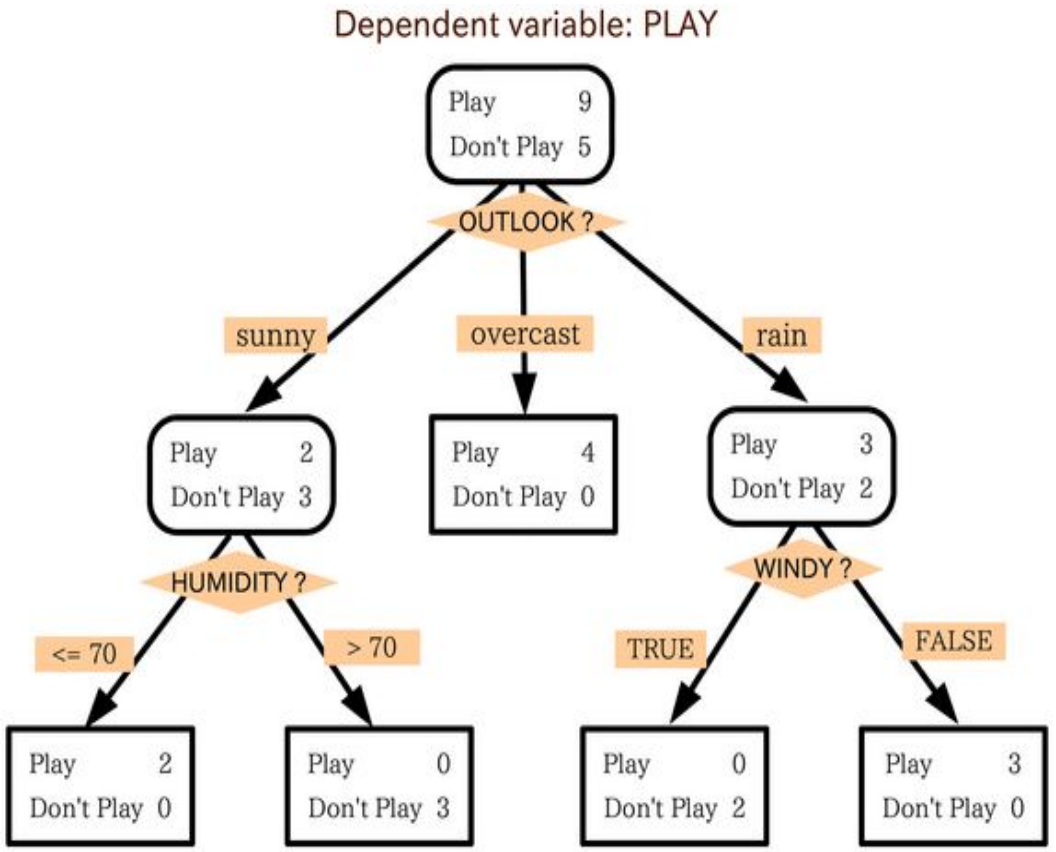


Decision trees

- Example: decision trees tools that create rules
- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Play golf dataset

Independent variables				Dep. var
OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
sunny	85	85	FALSE	Don't Play
sunny	80	90	TRUE	Don't Play
overcast	83	78	FALSE	Play
rain	70	96	FALSE	Play
rain	68	80	FALSE	Play
rain	65	70	TRUE	Don't Play
overcast	64	65	TRUE	Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
rain	75	80	FALSE	Play
sunny	75	70	TRUE	Play
overcast	72	90	TRUE	Play
overcast	81	75	FALSE	Play
rain	71	80	TRUE	Don't Play

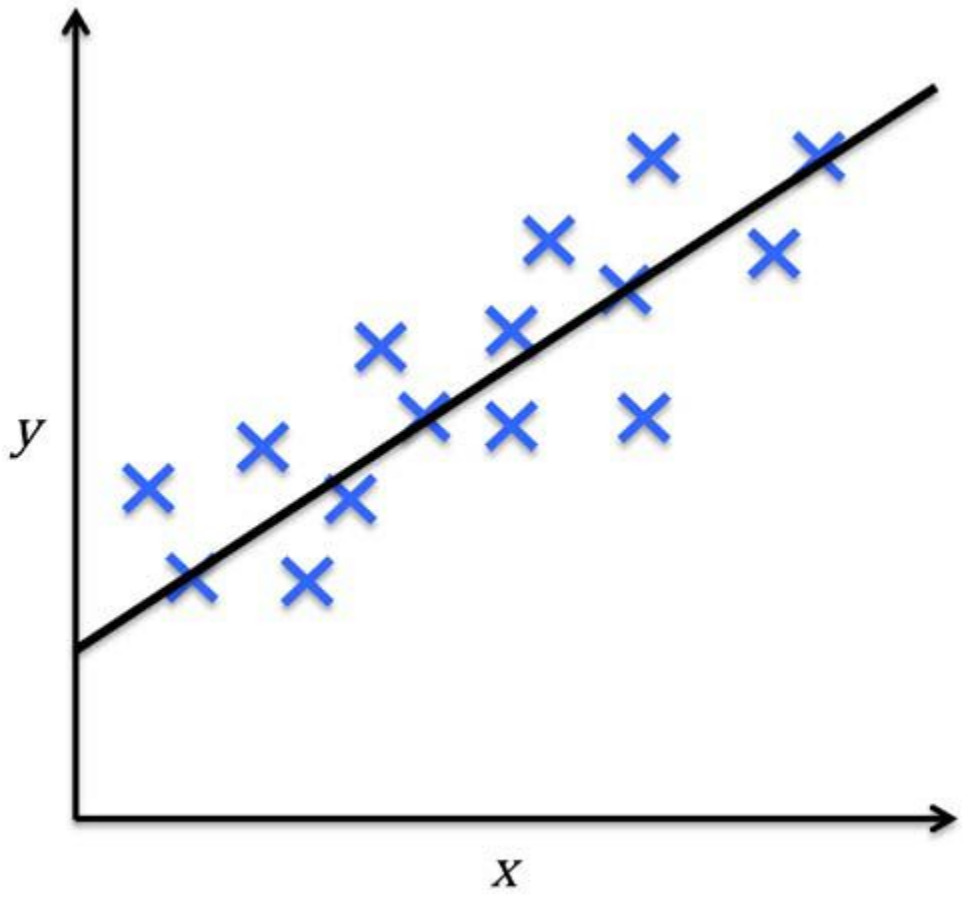


Classification Examples

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Medical diagnosis: From symptoms to illnesses
- Web Advertizing: Predict if a user clicks on an ad on the Internet.

Regression for predicting continuous outcomes

- A second type of supervised learning is the prediction of continuous outcomes, which is also called regression analysis. In regression analysis, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome), and we try to find a relationship between those variables that allows us to predict an outcome.
- Example
- Given a predictor variable x and a response variable y, we fit a straight line to this data that minimizes the distance—most commonly the average squared distance—between the sample points and the fitted line.
- We can now use the intercept and slope learned from this data to predict the outcome variable of new data:


y
x

Unsupervised Learning

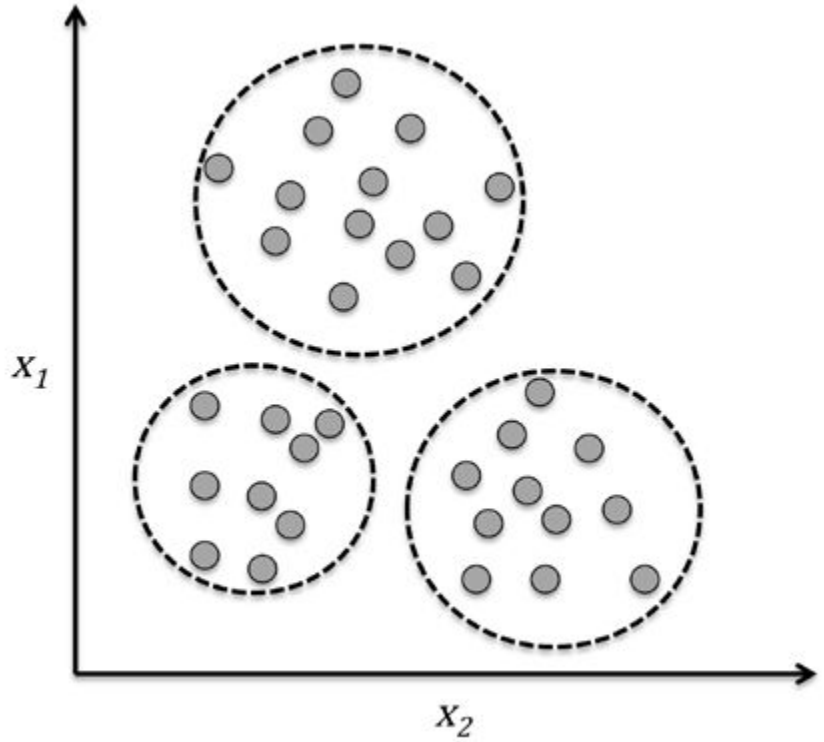
- In unsupervised learning, however, we are dealing with unlabeled data or data of unknown structure.
- Using unsupervised learning techniques, we are able to explore the structure of our data to extract meaningful information without the guidance of a known outcome variable or reward function.
- Clustering is considered to be the most important unsupervised learning problem. Deals with finding structure in unlabeled data
- Clustering is “the process of organizing objects into groups whose members are similar in some way”. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

- The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data.
- Clustering is a great technique for structuring information and deriving meaningful relationships among data,
- For example, it allows marketers to discover customer groups based on their interests in order to develop distinct marketing programs.
- **K-Means Cluster Algorithm**

The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid
3. When all objects have been assigned, recalculate the positions of the K centroids
4. Repeat steps 2 and 3 until the centroids no longer move.

The figure below illustrates how clustering can be applied to organizing unlabeled data into three distinct groups based on the similarity of their features x1 and x2:



Associations

- In the case of retail—for example, a supermarket chain—one application of machine learning is basket analysis, which is finding associations between products bought by customers:
- If people who buy X typically also buy Y, and if there is a customer who buys X and does not buy Y, he or she is a potential Y customer. Once we find such customers, we can target them for cross-selling.

- Basket analysis:

P (Y | X) probability that somebody who buys X also buys Y where X and Y are products/services.

- Example: P (chips | beer) = 0.7

- Learning from labelled data (supervised learning)

Eg. Classification, regression, prediction, function approx.

- Learning from unlabelled data (unsupervised learning)

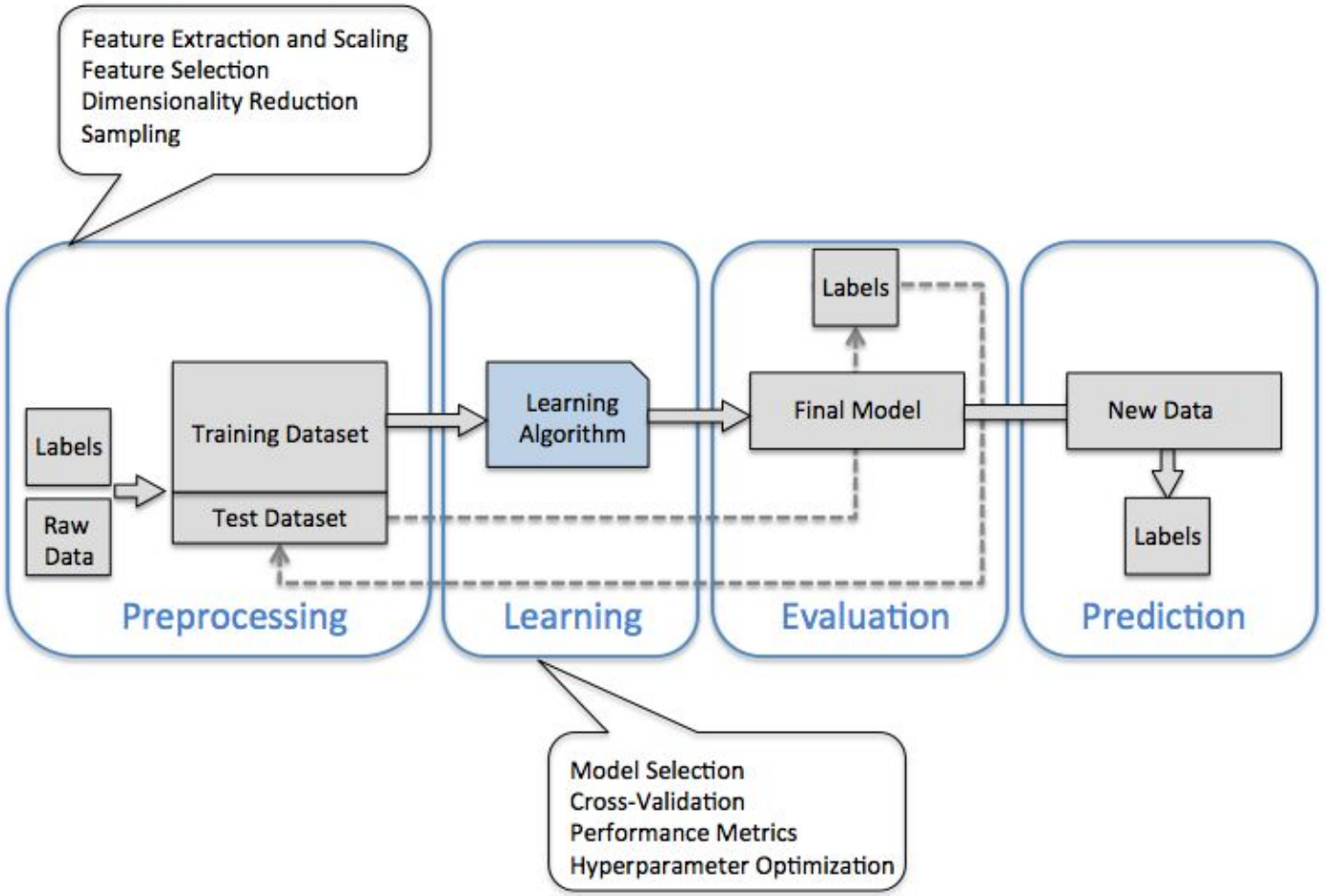
Eg. Clustering, visualisation, dimensionality reduction

- Learning from sequential data

Eg. Speech recognition, DNA data analysis

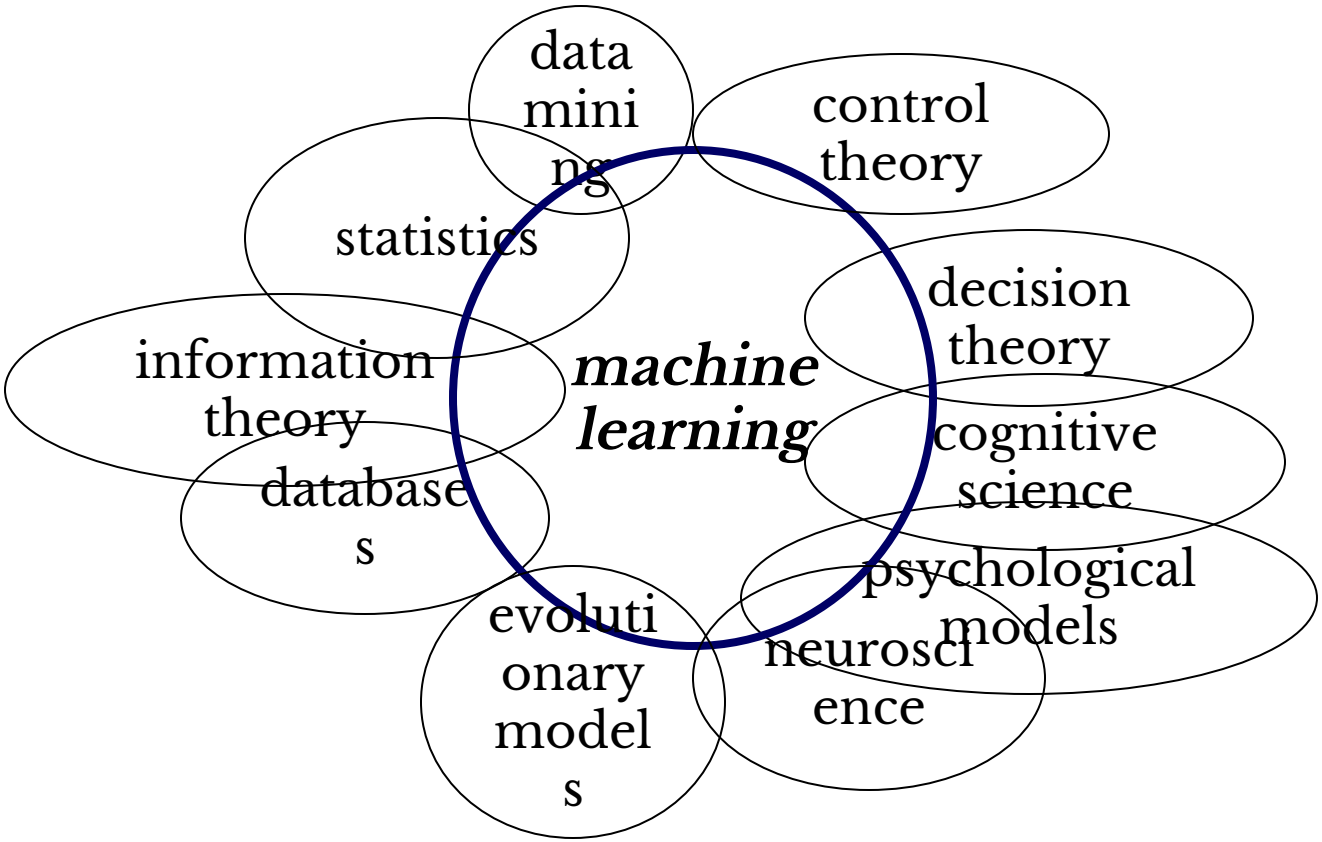
- Associations
- Reinforcement Learning

workflow diagram for using machine learning in predictive modeling



- We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience.
- The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both.

Related Fields



Machine learning is primarily concerned with the accuracy and effectiveness of the *computer system*.

Common Applications:

- **Marketing**: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records
- **Biology:** classification of plants and animals given their features
- **Insurance:** Fraud Detection
- **City-planning:** identifying groups of houses according to their house type, value and geographical location;
- **Earthquake studies:** clustering observed earthquake epicenters to identify dangerous zones;
- **WWW:** document classification; clustering clickstream data to discover groups of similar access patterns. Creating recommender systems

Conclusion

- In this presentation, a broad overview of machine learning is presented. Machine learning is generally composed of modeling and optimization, and the necessary part to perform machine learning is a suitable dataset for knowledge learning.

- Machine learning explores the construction and study of algorithms that can learn from and make predictions on data.

References

1. Python Machine Learning, Sebastian Raschka, Packt Publishing
2. Introduction to Machine Learning, Thomas Dietterich, Editor Christopher Bishop, David Heckerman, Michael Jordan, and Michael Kearns, Associate Editors, Second Edition

Thank You