Introduction to Machine Learning

Sachin Tripathi

IIT(ISM), Dhanbad

Course Objectives

- This course provides a thorough understanding of machine learning techniques, algorithms, and applications.
- Students will learn the theoretical foundations of machine learning models through lectures.
- Understand the principles and methodologies of machine learning

Learning Outcomes

- Develop skills in data preprocessing, model selection, and evaluation.

- Apply machine learning techniques to solve real-world problems and analyze datasets.

- Understand the principles and methodologies of machine learning.

- Explore advanced topics in machine learning, including deep learning and reinforcement learning

Topics to be Covered (Unit-Wise)

- Introduction to Machine Learning: Definition and types of machine learning, Overview of the machine learning process, nature of machine learning tasks using motivating applications.
- Supervised Learning: Linear regression; Logistic regression, Decision trees and ensemble methods (Random Forests, Gradient Boosting); Support Vector Machines (SVM).
- Model Evaluation and Selection: Cross-validation; Evaluation metrics (accuracy, precision, recall, F1-score), Hyper parameter tuning.

- Unsupervised Learning: Clustering algorithms (K-means, Hierarchical clustering), Dimensionality reduction techniques (PCA, t-SNE), Anomaly detection.

- Introduction to Deep Learning: Neural networks architecture; Training neural networks with backpropagation; Convolutional Neural Networks (CNNs) for image classification; Recurrent Neural Networks (RNNs) for sequence modeling.

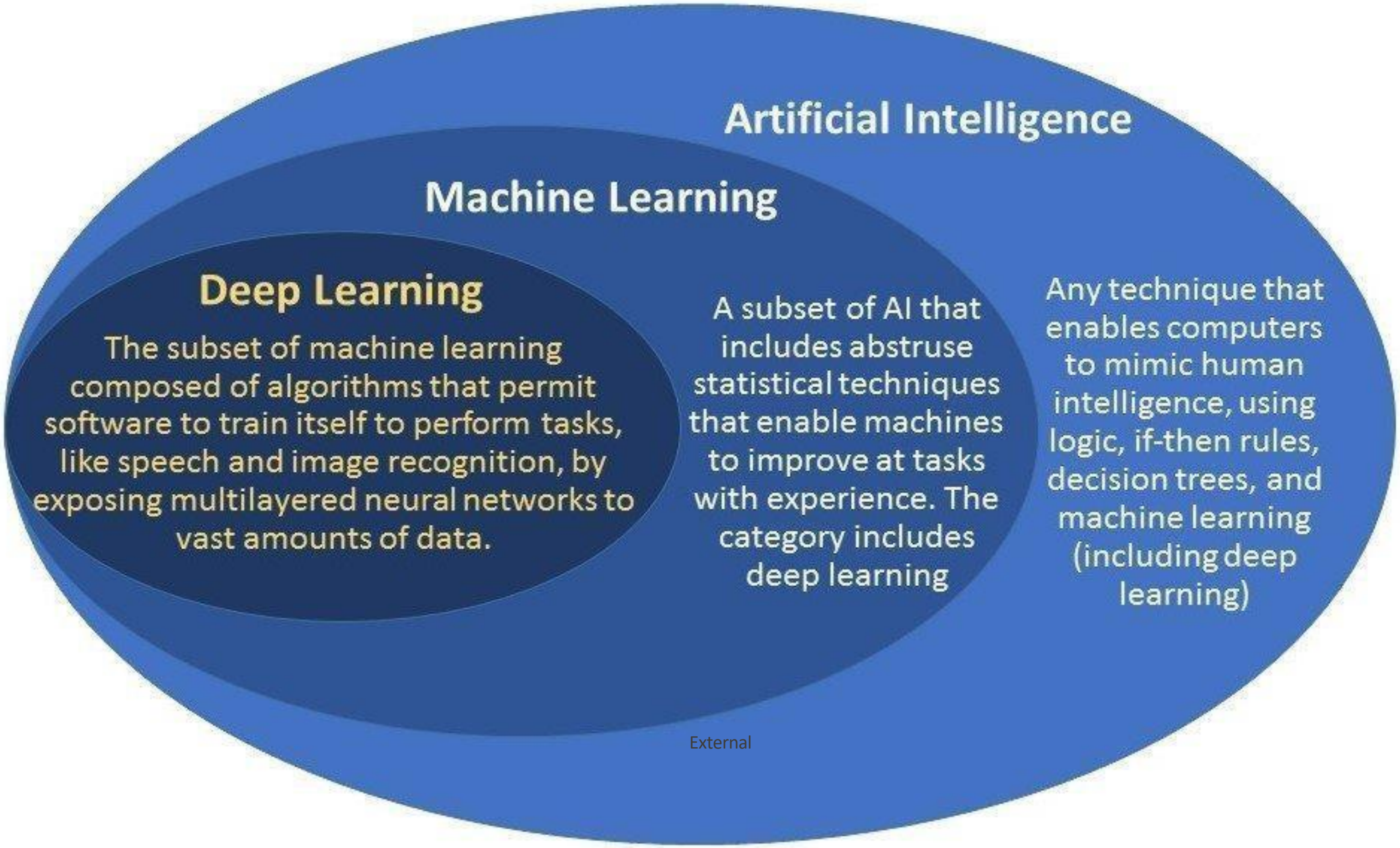
- Advanced Topics in Machine Learning; Reinforcement learning fundamentals; Deep reinforcement learning algorithms (Q-learning, Deep Q-Networks), Transfer learning and domain adaptation, Model interpretability and explainability

Text Books:

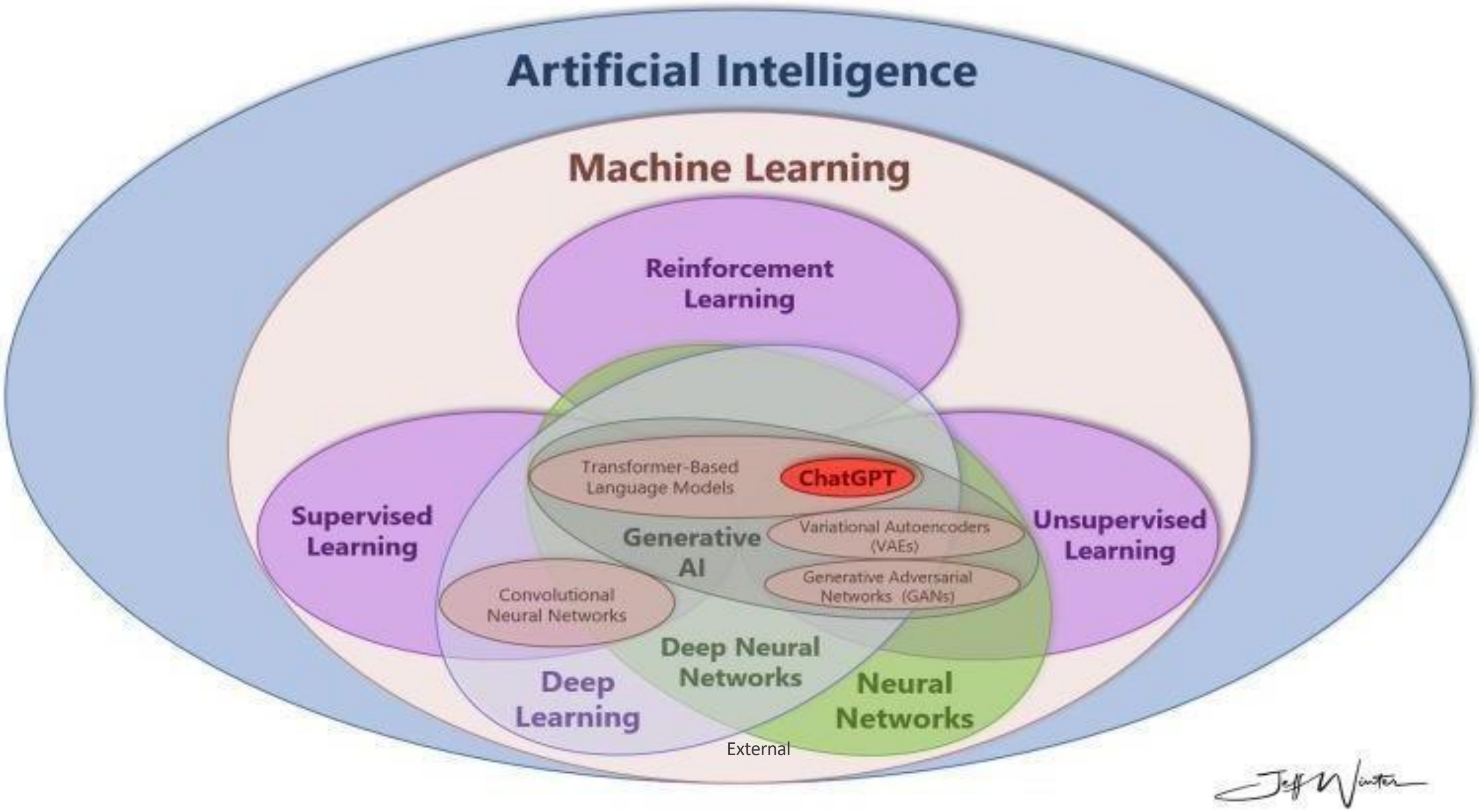
- Pattern Recognition and Machine Learning" by Christopher M. Bishop

- "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

Overview


Artificial Intelligence
Machine Learning
Deep Learning
The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data.
A subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning
Any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning)
External

Detailed



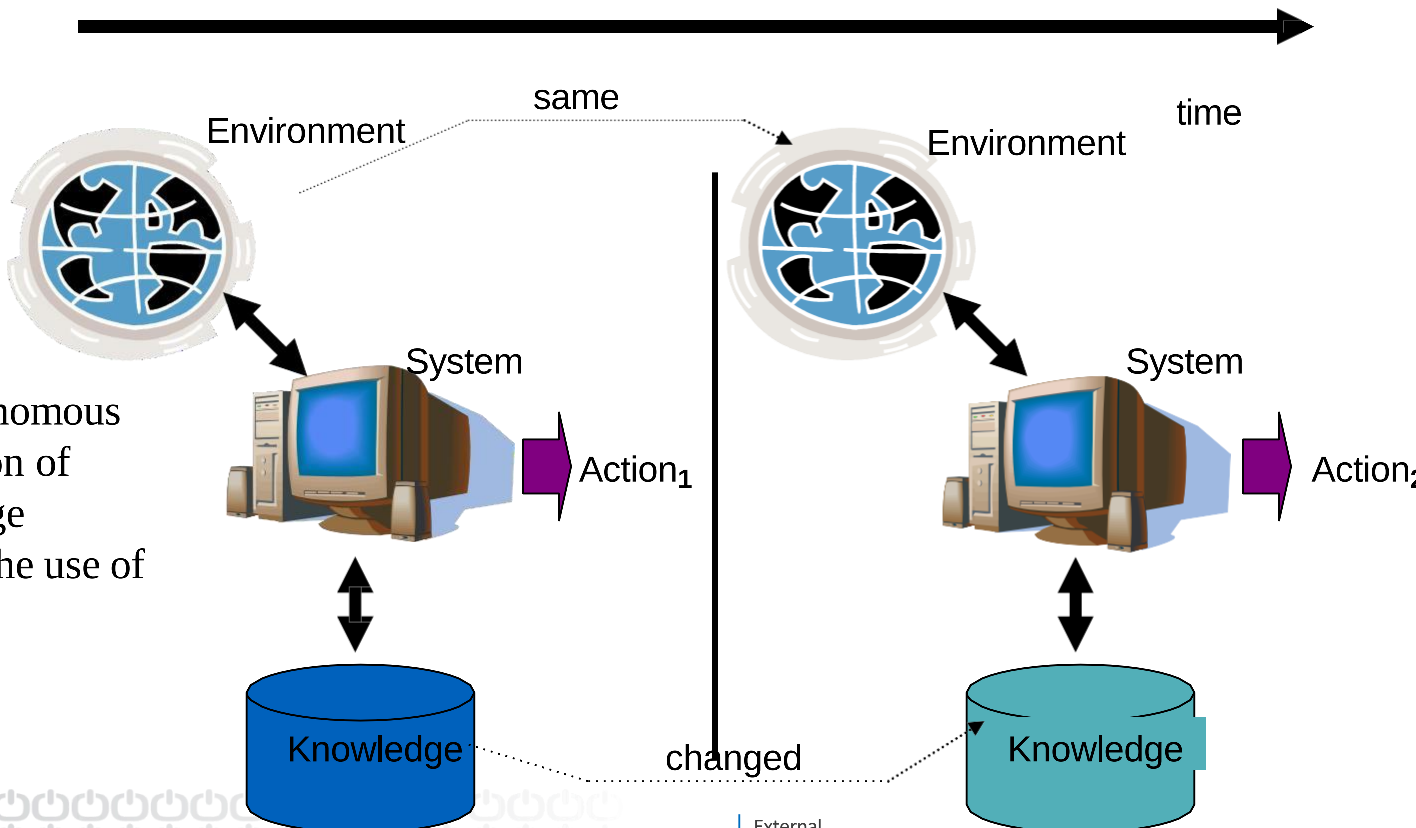
Artificial Intelligence
Machine Learning
Reinforcement Learning
Supervised Learning
Unsupervised Learning
Transformer-Based Language Models
ChatGPT
Generative AI
Variational Autoencoders (VAEs)
Generative Adversarial Networks (GANs)
Convolutional Neural Networks
Deep Neural Networks
Deep Learning
Neural Networks
External
Jeff Winter

What is Machine Learning?

- Machine Learning (ML) is a sub-field of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.
- Using algorithms that iteratively learn from data
- Allowing computers to discover patterns without being explicitly programmed where to look

What is Machine Learning?

“Logic is not the end of wisdom, it is just the beginning” --- Spock



The autonomous acquisition of knowledge through the use of computer programs

External

When do we need ML? (I)

- For tasks that are easily performed by humans but are complex for computer systems to emulate

- Vision: Identify faces in a photograph, objects in a video or still image, etc.
- Natural language: Translate a sentence from Hindi to English, question answering, identify sentiment of text, etc.

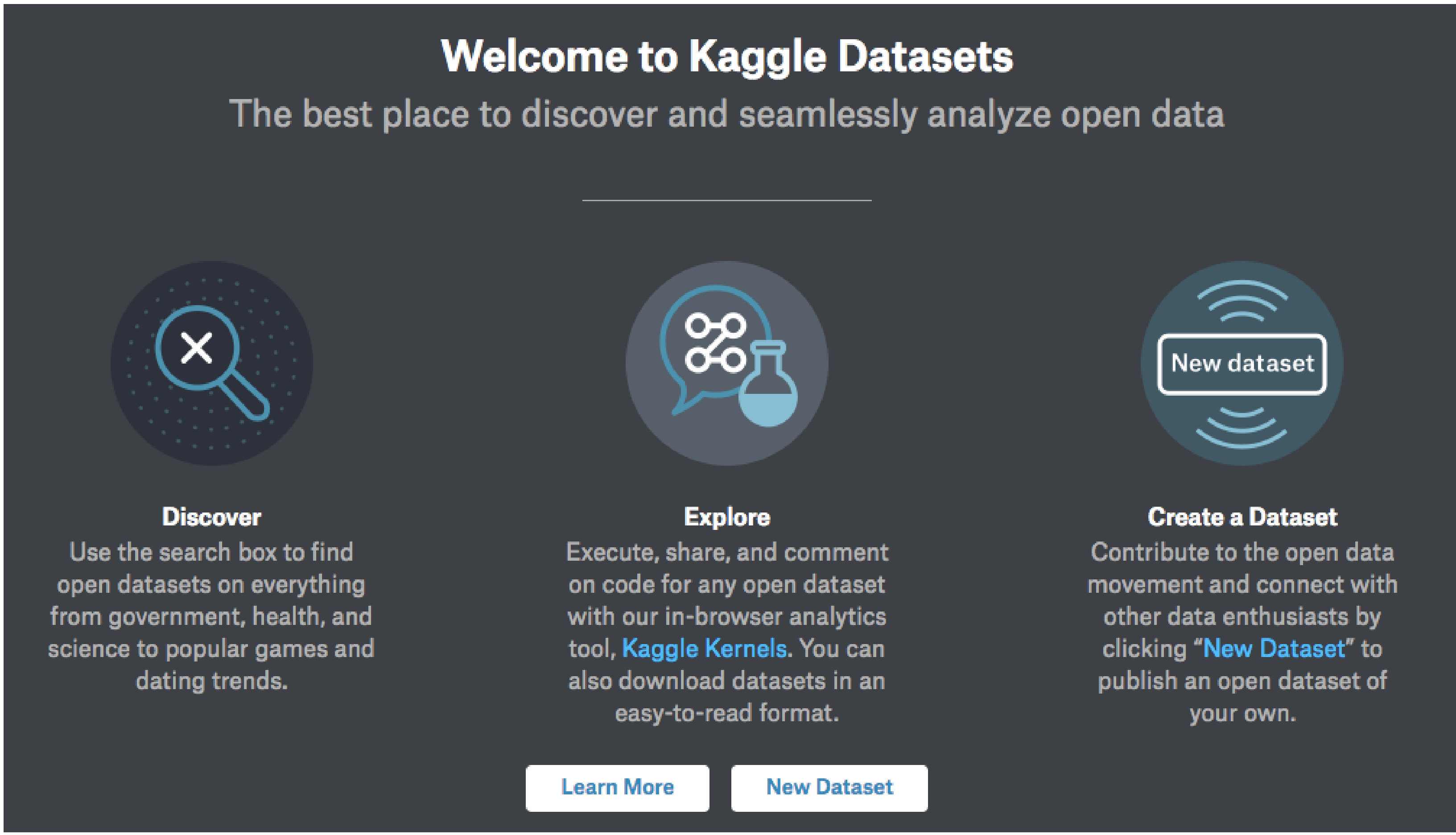
- Speech: Recognize spoken words, speaking sentences naturally
- Game playing: Play games like chess, Go, Dota.
- Robotics: Walking, jumping, displaying emotions, etc.
- Driving a car, navigating a maze, etc

Example

mite
container ship
motor scooter
leopard
mite
black widow
cockroach
tick
starfish
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
go-kart
moped
bumper car
golfcart
leopard
jaguar
cheetah
snow leopard
Egyptian cat

Datasets abound…

Kaggle: https://www.kaggle.com/datasets



- Another good resource:
  http://deeplearning.net/datasets/

- Interesting datasets for computational journalists:
  http://cjlab.stanford.edu/2015/09/30/lab-launch-and-data-sets/
- Popular resource for ML beginners:
  http://archive.ics.uci.edu/ml/index.php

- Speech and language resources:
  www.openslr.org/

ML libraries/toolkits

scikit-learn, openCV, Keras, Tensorflow, NLTK, etc.

Typical ML approach

How do we approach an ML problem?

- Modeling: Use a model to represent the task
- Decoding/Inference: Given a model, answer questions with respect to the model
- Training: The model could be parameterized and the parameters are estimated using data

How do we know if our model's any good?

Generalization: Does the trained model produce good predictions on examples beyond the training set?

- We should be careful not to over fit the training data
- Occam's Razor: All other things being equal, pick the simplest solution
- These concepts will be made more precise in later classes

Learning & Adaptation

- ”Modification of a behavioral tendency by expertise.”
(Webster 1984)
- A learning machine, broadly defined is any device whose actions are influenced by past experiences.” (Nilsson 1965)
- ”Any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population.” (Simon 1983)
- ”An improvement in information processing ability that results from information processing activity.” (Tanimoto 1990)

Ways humans learn things

- …talking, walking, running…
- Learning by mimicking, reading or being told facts
- Tutoring
- Being informed when one is correct
- Experience
- Feedback from the environment
- Analogy
- Comparing certain features of existing knowledge to new problems
- Self-reflection
- Thinking things in ones own mind, deduction, discovery

A few achievements

Programs that can:

- Recognize spoken words
- Predict recovery rates of pneumonia patients
- Detect fraudulent use of credit cards
- Drive autonomous vehicles
- Play games like backgammon – approaching the human champion!

What do we mean by learning?

- Given
  - a data set *D*,
  - a task *T*, and
  - a performance measure *M*,

A computer system/Program is said to learn from *D* to perform the task *T* if after learning the system's performance on *T* improves as measured by *M*.

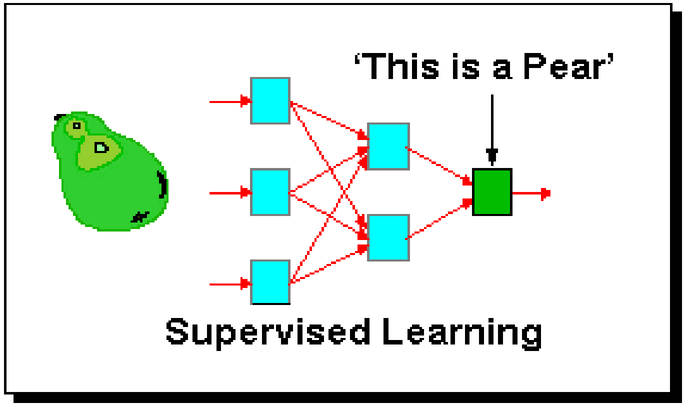
- In other words, the learned model helps the system to perform *T* better as compared to no learning.

Machine Learning (ML)

- ML is a branch of artificial intelligence:
  - Uses computing based systems to make sense out of data
    - Extracting patterns, fitting data to functions, classifying data, etc
  - ML systems can learn and improve
    - With historical data, time and experience
  - Bridges theoretical computer science and real noise data.
    - Machine learning involves automatic procedures that learn a task from a
      series of examples
    - Most convenient source of examples is data

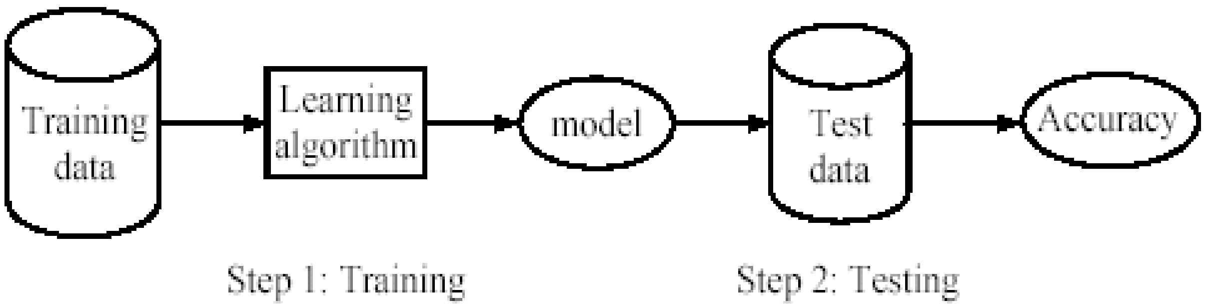
Machine Learning Paradigms

- Supervised Learning
  - For every example in the data there is always a predefined outcome
  - Models the relations between a set of descriptive features and a target (Fits data to a function)
  - 2 groups of problems:
    - Classification → Discrete
    - Regression → Continuous



Supervised learning process: two steps

- Learning (training): Learn a model using the training data
- Testing: Test the model using unseen test data to assess the model accuracy

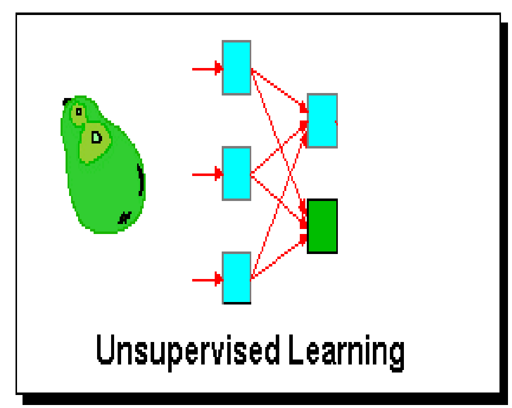


$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$

- Unsupervised Learning
  - There are not predefined and known set of outcomes
  - Look for hidden patterns and relations in the data
  - A typical example: Clustering

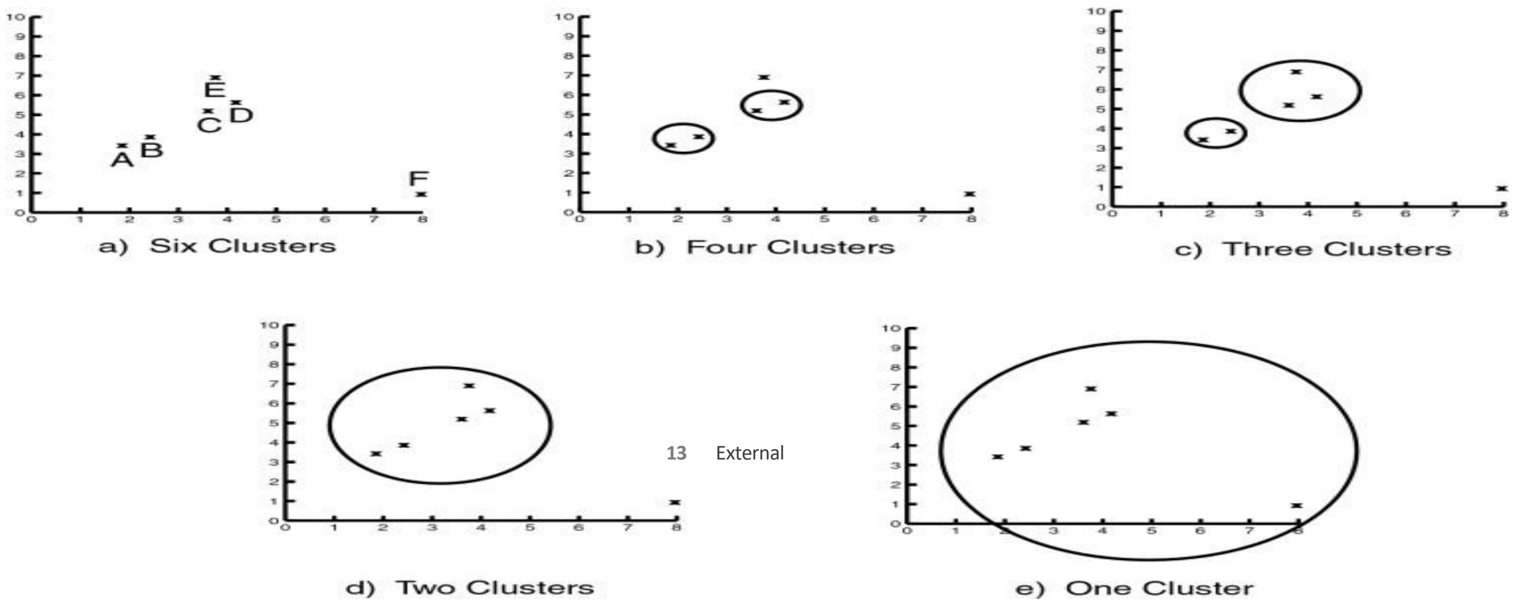
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1

External



Clustering

- Seeks to place objects into meaningful groups automatically, based on their similarity.
- Does not require the groups to be predefined. The hope in applying clustering algorithms is that they will discover useful but unknown classes of items.



❑ Semi Supervised Learning (Why ?)

- Data labeling is expensive and difficult
- Labeling is often unreliable
- Unlabeled examples
- Easy to obtain in large numbers
- e.g. webpage classification, bioinformatics, image classification

SSL Algorithms

- Self-Training
- Generative Models
- S3VMs
- Graph-Based Algorithms
- Co-training
- Multiview algorithms

How Self-Training Works:

- **Train the model on the small labeled dataset.**
  Start by training a supervised model on the labeled examples you have.
- **Use the trained model to predict labels for the unlabeled data.**
  Run the model on unlabeled data and get predicted labels with confidence scores (like probabilities).
- **Select the most confident predictions.**
  Choose unlabeled samples where the model is most confident about the predicted label (above some confidence threshold).

- **Add these confidently labeled samples to the labeled set.**
  Treat those confident predictions as if they were true labels.
- **Retrain the model on this expanded labeled set.**
  Train again with the original labeled data plus these new pseudo-labeled samples.
- **Repeat steps 2-5 iteratively.**
  Gradually the model improves as it “teaches itself” using more data.

Example Use Cases:

- Speech recognition (lots of audio, few transcriptions)
- Image classification (many images, few labeled)
- Text classification (many documents, few labeled examples)

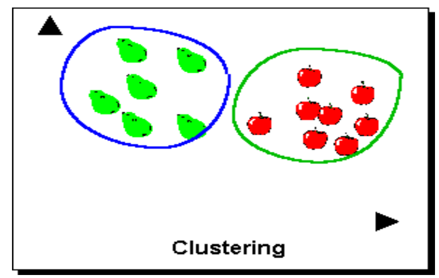
Reinforcement Learning

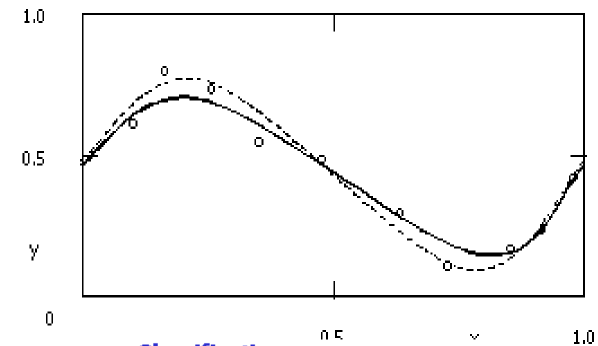
- ❑ Reinforced learning - with reward or punishment (an action is evaluated)
- ❑ Reinforcement Learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment to maximize cumulative rewards.
- ❑ Types of Reinforcement Learning Algorithms

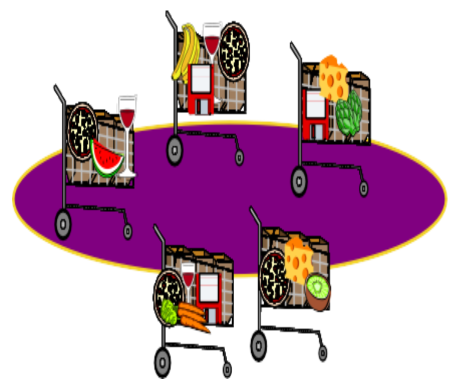
  - ➢ Model-Free RL
    Learns policy/value directly from experience
    Examples: Q-learning, SARSA, Policy Gradient
  - ➢ Model-Based RL
    Tries to model the environment and plan ahead

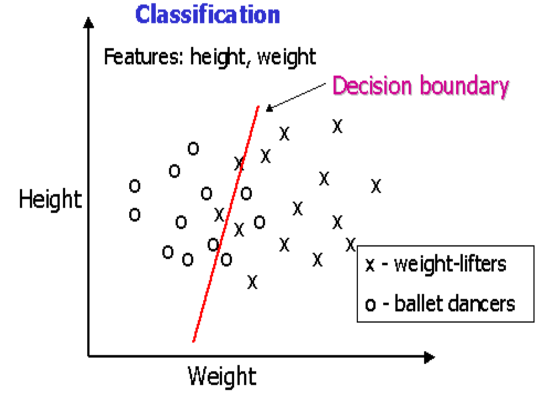
Machine Learning Tasks (Models)

- Classification
- Clustering
- Regression
- Time series analysis
- Association Analysis
- Sequence Discovery
- Anomaly Detection
- Dimensionality Reduction









Classification

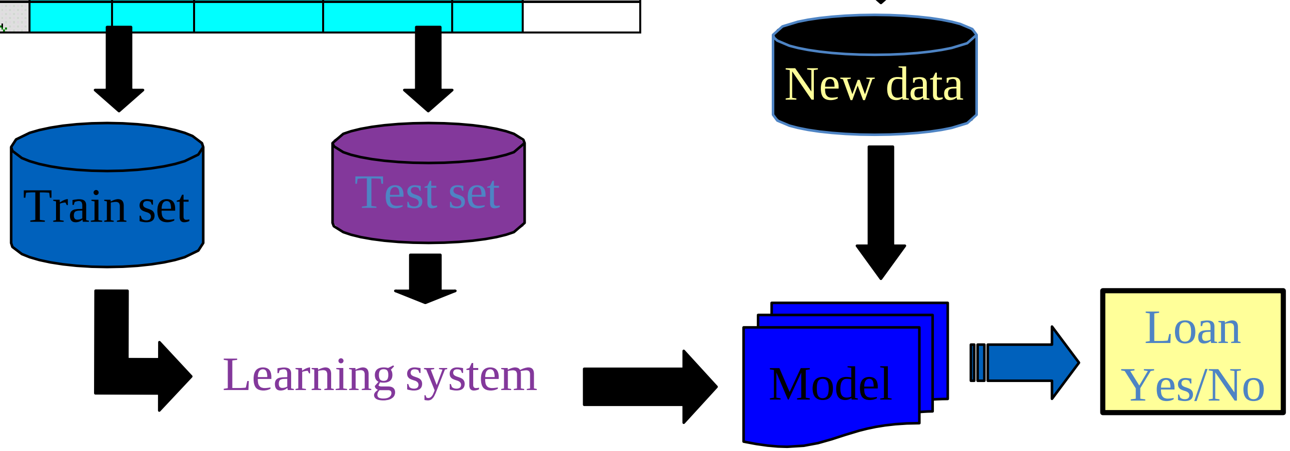
- **Goal:** Assign data points to predefined categories or classes.
- **Example:** Email spam detection (spam or not spam), disease diagnosis (sick or healthy).

- **Common Algorithms:** Logistic regression, decision trees, random forests, SVM, neural networks.

Classification example

No	Sex	Age	Marital status	Net Income	...	Loan
1	F	38	Married	45K	.	Yes
2	M	42	Married	66K	.	Yes
3	F	52	Single	43K	.	No
4	M	50	Single	70K	.	No
5	F	27	Married	40K	.	No
6	M	45	Divorced	38K	.	No
7	F	35	Widow	59K	.	Yes
8	M	32	Married	52K	.	Yes

Sex	Age	Marital status	Net Income	...	Loan
F	28	Married	44K	.	?
M	47	Divorced	95K	.	?
F	30	Single	45K	.	?
M	55	Single	69K	.	?
M	45	Married	41K	.	?



Clustering

- **Goal:** Group similar data points together without predefined labels.
- **Example:** Customer segmentation, grouping news articles by topic.
- **Common Algorithms:** K-Means, hierarchical clustering, DBSCAN.

Regression

- **Goal:** Predict a continuous value based on input features.
  - **Example:** Predict house prices, stock prices, or temperature.
- **Common Algorithms:** Linear regression, polynomial regression, support vector regression.

Time Series Analysis

- **Goal:** Analyze data points collected over time to identify trends, seasonality, and forecast future values.
- **Example:** Weather forecasting, sales prediction.
- **Common Techniques:** ARIMA, exponential smoothing, LSTM neural networks.

Association Analysis

- **Goal:** Discover rules and relationships between variables in large datasets.
  - **Example:** Market basket analysis (if someone buys bread, they also buy butter).
- **Common Algorithms:** Apriori, FP-Growth.

Sequence Discovery (Sequential Pattern Mining)

Goal: Find frequent ordered patterns or sequences in data.
Example: Customer purchase sequences, web navigation paths.
Common Algorithms: GSP, SPADE, PrefixSpan.

Others

- Anomaly Detection

  Goal: Identify unusual or rare data points that don't fit the general pattern.

- Example: Fraud detection, network intrusion detection.

- Dimensionality Reduction

  Goal: Reduce the number of input variables while preserving important information.

- Example: PCA, t-SNE.

Inductive learning

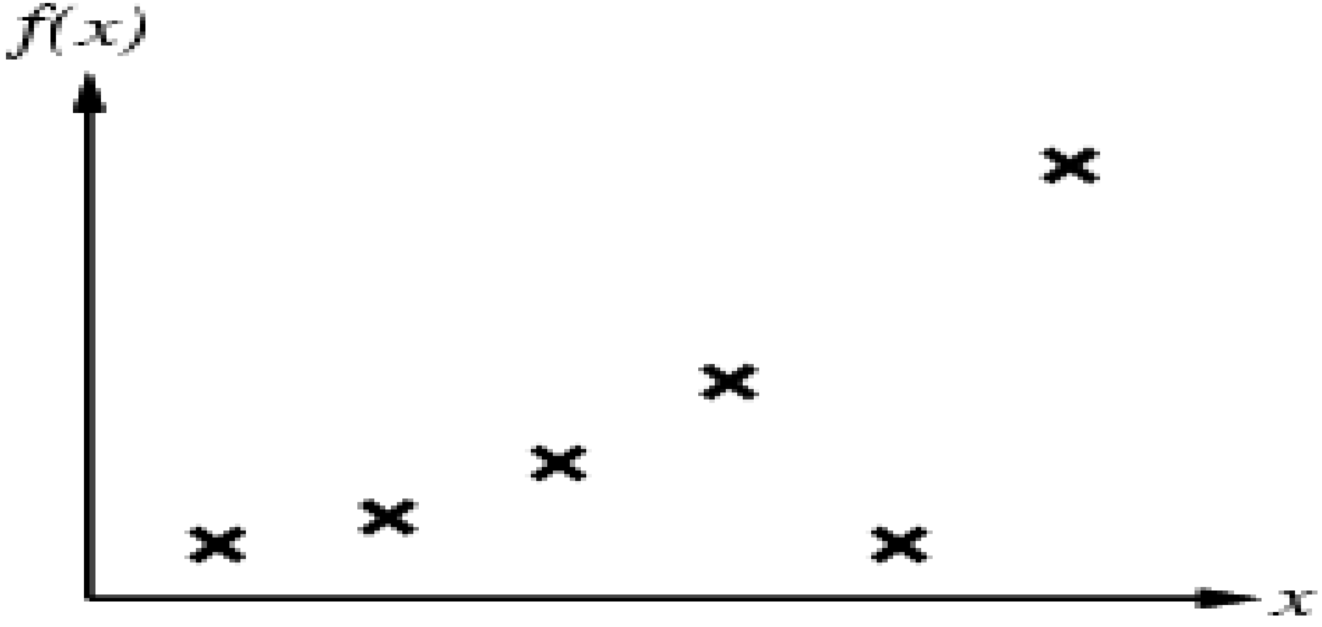
- Simplest form: learn a function from examples

  f is the target function An example is a pair (x, $f(x)$)

Problem: find a hypothesis h
such that $h \approx f$
given a training set of examples

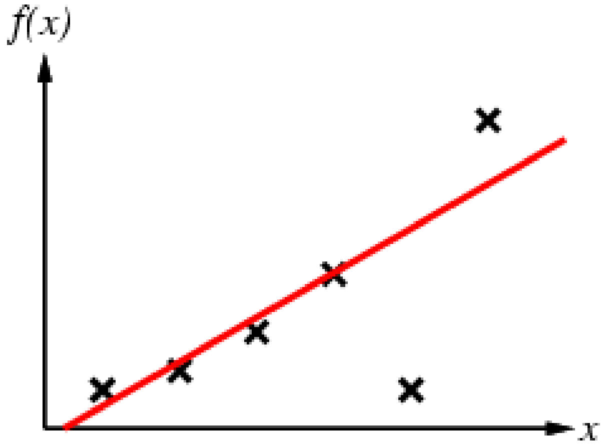
Inductive learning method

- Construct/adjust *h* to agree with *f* on training set (*h* is consistent if it agrees with *f* on all examples)
- E.g., curve fitting:



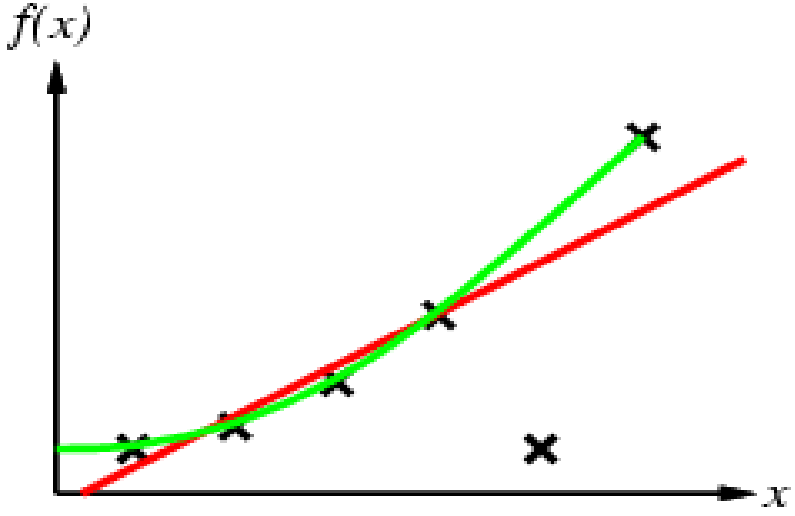
Inductive learning method

- Construct/adjust *h* to agree with *f* on training set (*h* is consistent if it agrees with *f* on all examples)
- E.g., curve fitting:



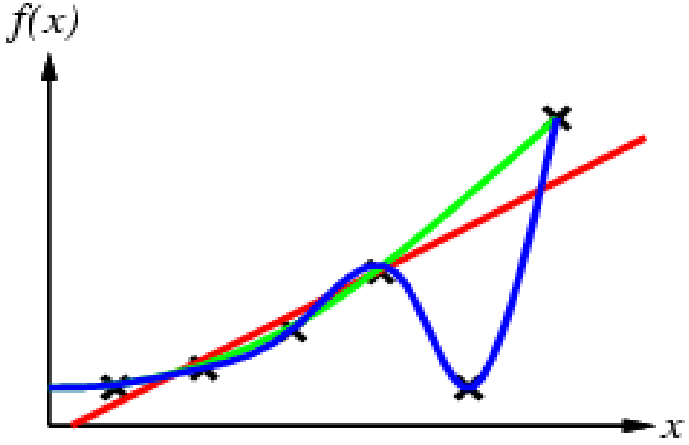
Inductive learning method

- Construct/adjust *h* to agree with *f* on training set (*h* is consistent if it agrees with *f* on all examples)
- E.g., curve fitting:



Inductive learning method

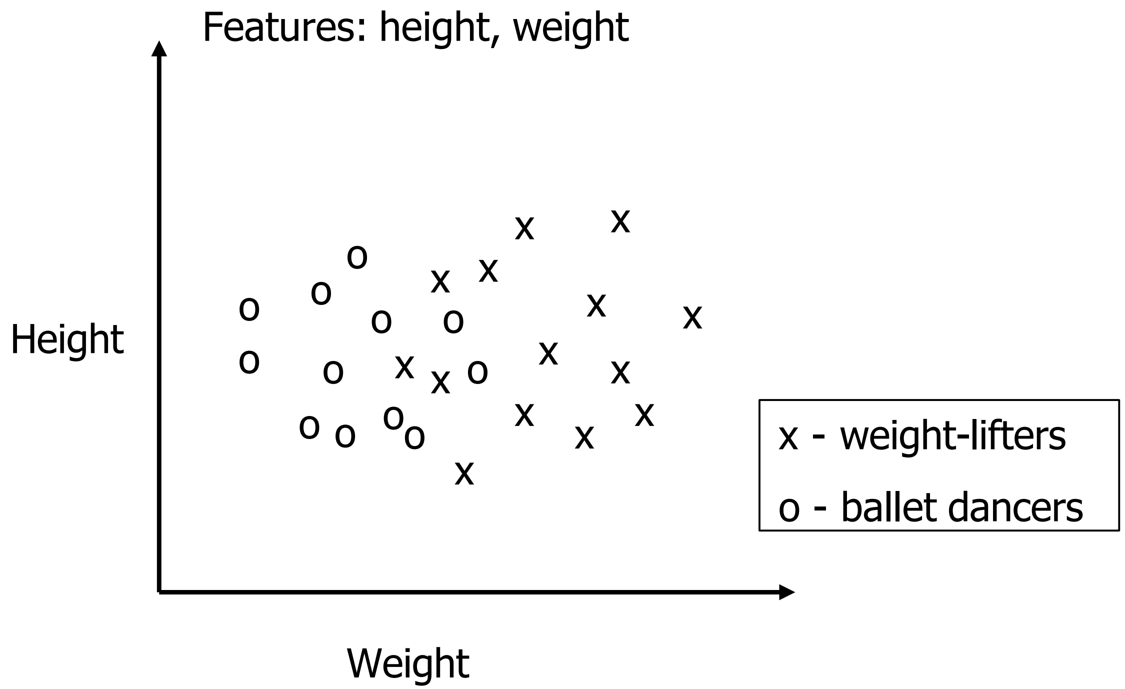
- Construct/adjust *h* to agree with *f* on training set (*h* is consistent if it agrees with *f* on all examples)
- E.g., curve fitting:



Machine Learning Methods

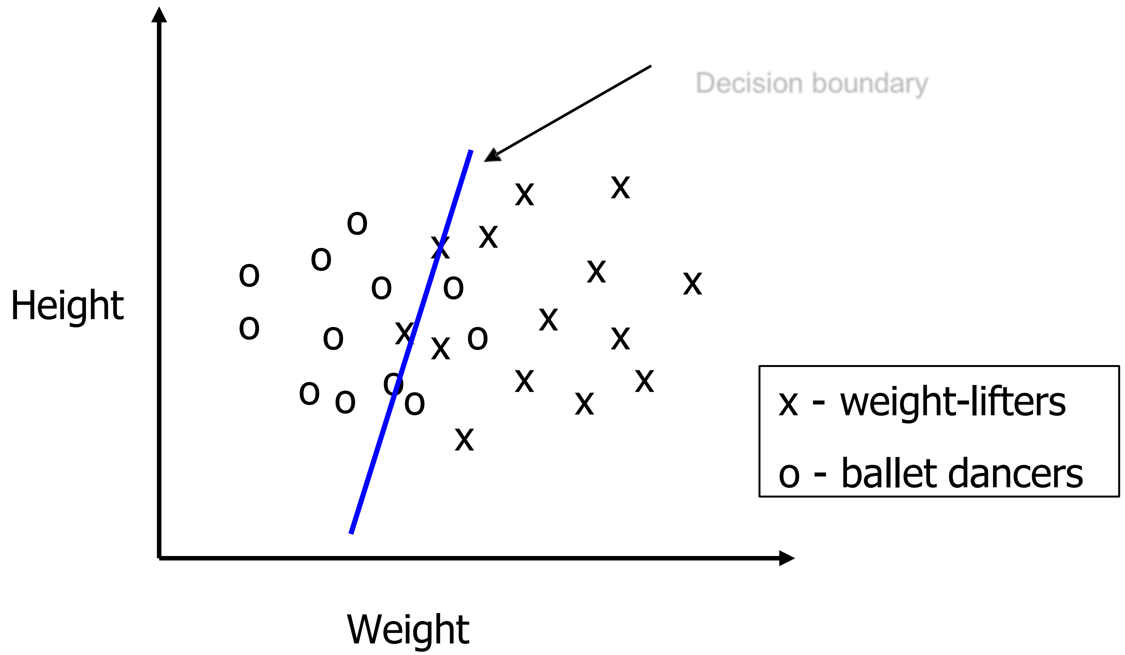
- Instance Based Methods (CBR, k-NN)
- Decision Trees
- Artificial Neural Networks
- Bayesian Networks
- Naïve Base
- **Evolutionary Strategies**
- Support Vector Machines
- Ensemble

Classification example

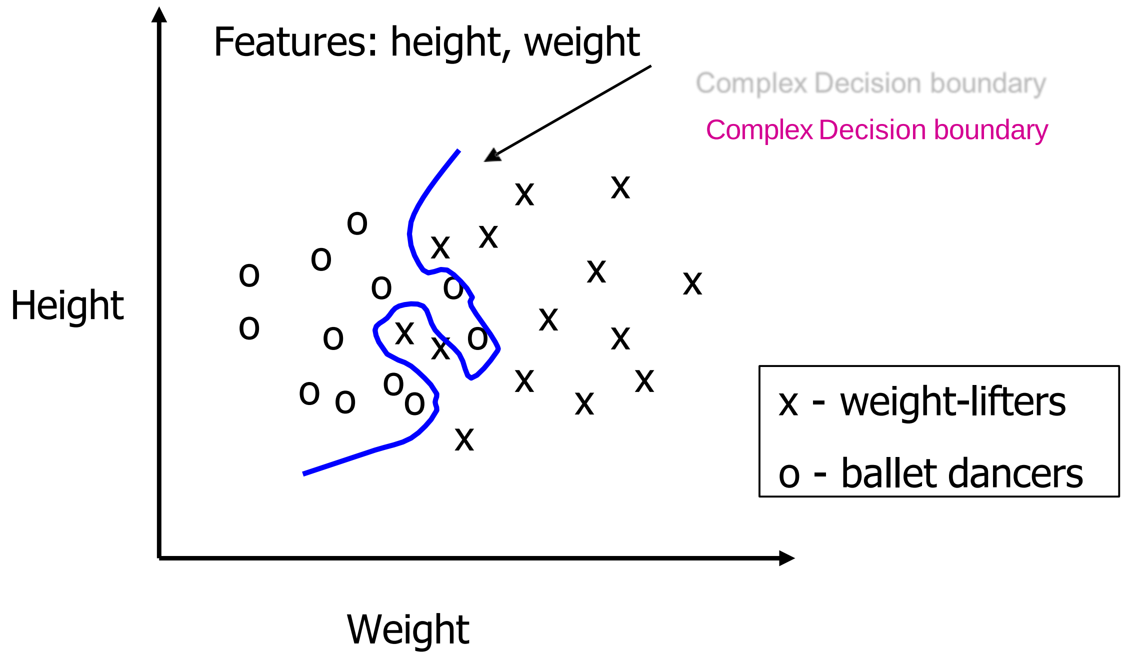


Classification example - Simple Model

Features: height, weight



Classification example - Complex model



Note: A simple decision boundary is better than a complex one - It GENERALIZES better.

Machine Learning Methodologies

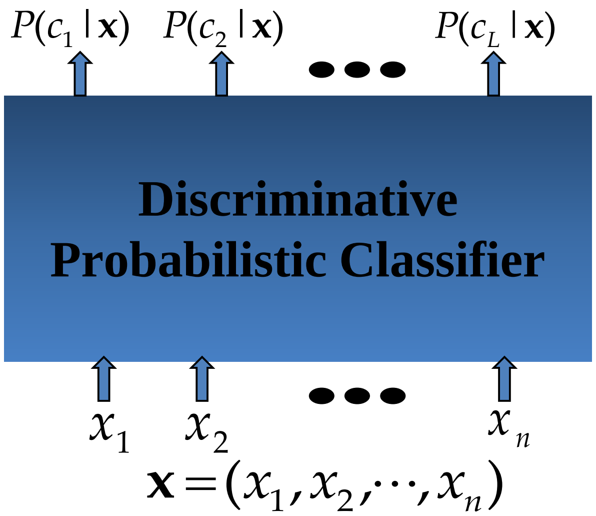
- There are three methodologies:
  - Model a classification rule directly
    Examples: k-NN, linear classifier, SVM, neural nets, …
  - Model the probability of class memberships given input data
    Examples: logistic regression, probabilistic neural nets (softmax),…
  - Make a probabilistic model of data within each class
    Examples: naive Bayes

- Important ML taxonomy for learning models
  probabilistic models vs non-probabilistic models
  discriminative models vs generative models

Probabilistic Classification Principle

- Establishing a probabilistic model for classification
  - **Discriminative model**

$$\boxed{P(c\,|\,\mathbf{x}) \quad c = c_1, \cdots, c_L, \mathbf{x} = (x_1, \cdots, x_n)}$$



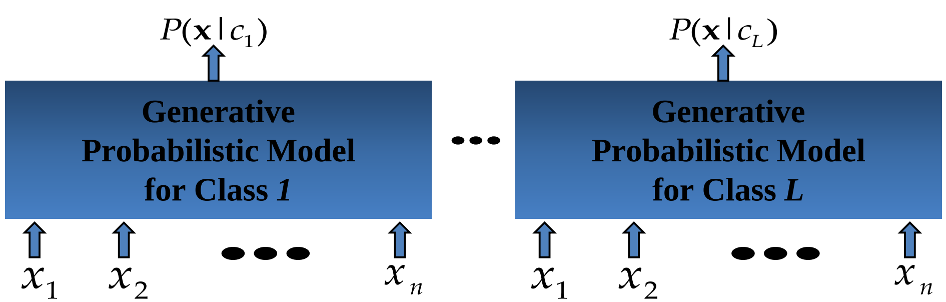
- To train a discriminative classifier (regardless its probabilistic or non-probabilistic nature), all training examples of different classes must be jointly used to build up a single discriminative classifier.

- Output L probabilities for L class labels in a probabilistic classifier while a single label is achieved by a non-probabilistic discriminative classifier .

Probabilistic Classification Principle

- Establishing a probabilistic model for classification (cont.)
  - **Generative model (must be probabilistic)**

$$\boxed{P(\mathbf{x}\,|\,c) \quad c = c_1, \cdots, c_L,\ \mathbf{x} = (x_1, \cdots, x_n)}$$



$$\mathbf{x} = (x_1, x_2, \cdots, x_n)$$

- L probabilistic models have to be trained independently
- Each is trained on only the examples of the same label
- Output L probabilities for a given input with L models
- "Generative" means that such a model can produce data subject to the distribution via sampling.

. Generative Classification

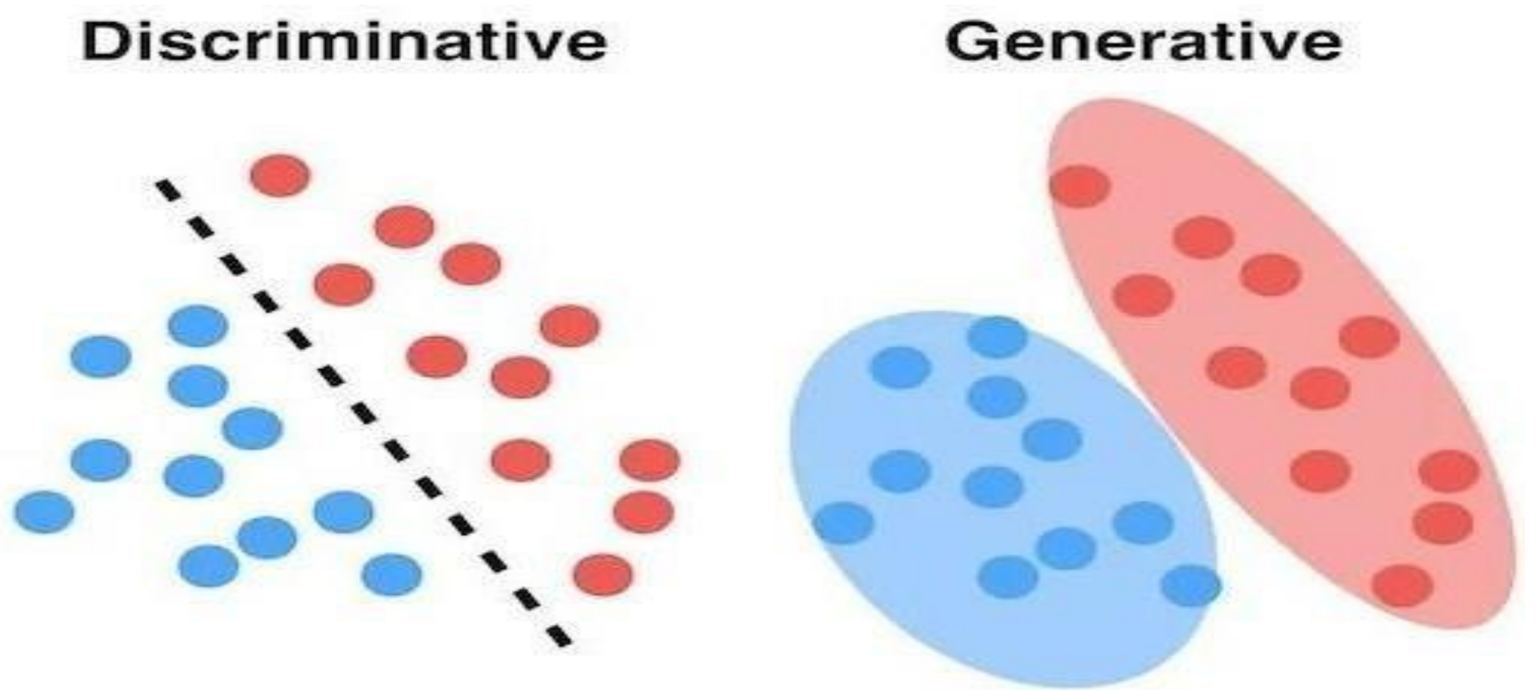
- Generative Classifiers tries to model class, i.e., what are the features of the class.
  An example of such classifiers is Naive Bayes.

- Mathematically, generative models try to learn the joint probability distribution, p(x,y), of the inputs x and label y, and make their prediction using Bayes rule to calculate the conditional probability, p(y|x), and then picking a most likely label.

Discriminative

- Discriminative Classifiers learn what the features in the input are most useful to distinguish between the various possible classes. So, if given images of dogs and cats, and all the dog images have a collar, the discriminative models will learn that having a collar means the image is of dog.

Discriminative Vs. Generative Classification



Continued..

- Both methods use conditional probability to classify but learn different types of probabilities to generate conditional probability.
- It is seen, in most of the classification tasks, discriminative classifiers are often more accurate, so they are most commonly used. One of the reasons for the discriminative classifier is more accurate as it tries to directly solve the classification task.