UNIT IV

Machine Learning: Supervised Learning

4. What is Machine Learning?

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.
- The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.
- The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.
- Machine learning is a set of tools that, broadly speaking, allow us to “teach” computers how to perform tasks by providing examples of how they should be done.
- For example, suppose we wish to write a program to distinguish between valid email messages and unwanted spam. We could try to write a set of simple rules, for example, flagging messages that contain certain features (such as the word “viagra” or obviously-fake headers).
- However, writing rules to accurately distinguish which text is valid can actually be quite difficult to do well, resulting either in many missed spam messages, or, worse, many lost emails.
- Worse, the spammers will actively adjust the way they send spam in order to trick these strategies (e.g., writing “vi@gr@”). Writing effective rules —and keeping them up-to-date — quickly becomes an insurmountable task.
- Machine learning has provided a solution. Modern spam filters are “learned” from examples: we provide the learning algorithm with example emails which we have manually labeled as “ham” (valid email) or “spam” (unwanted email), and the algorithms learn to distinguish between them automatically.

Machine learning methods are broken into two phases:

- **1. Training:** A model is learned from a collection of training data.
- **2. Application:** The model is used to make decisions about some new test data.
- For example, in the spam filtering case, the training data constitutes email messages labeled as ham or spam, and each new email message that we receive (and which to classify) is test data.
- However, there are other ways in which machine learning is used as well.

Advantages of Machine Learning:

i. Often much more accurate than human crafted rules (since data driven)
ii. Human often incapable of expressing what they know (example: rules of English, or how to recognized letter)
iii. Don't need a human expert or programmer
iv. Cheap and flexible can apply to any learning task
v. Data input from unlimited resources
vi. Fast processing and real time predictions

Disadvantages of Machine learning:

i. Need a lot of label data
ii. Error phone, usually impossible to get perfect accuracy

4.1 Types of Machine Learning

– Some of the main types of machine learning are:

1. **Supervised Learning:**

– Supervised learning in which the training data is labeled with the correct answers, e.g.,"spam" or "ham." The two most common types of supervised learning are classification (where the outputs are discrete labels, as in spam filtering) and regression (where the outputs are real-valued).

2. **Unsupervised learning:**

– Unsupervised learning in which we are given a collection of unlabeled data, which we wish to analyze and discover patterns within.
– The two most important examples are dimension reduction and clustering.

3. **Reinforcement learning:**

– Reinforcement learning in which an agent (e.g., a robot or controller) seeks to learn the optimal actions to take based the outcomes of past actions.
– **There are many other types of machine learning as well, for example:**
  1. **Semi-supervised learning**, in which only a subset of the training data is labeled
  2. **Time-series forecasting**, such as in financial markets
  3. **Anomaly detection** such as used for fault-detection in factories and in surveillance
  4. **Active learning**, in which obtaining data is expensive, and so an algorithm must determine which training data to acquire and many others.

4.2 Application of Machine learning:

i. Virtual Personal Assistants Siri, Alexa, Google Now are some of the popular examples of virtual personal assistants.

Virtual Assistants are integrated to a variety of platforms. For example:

- Smart Speakers: Amazon Echo and Google Home
- Smartphones: Samsung Bixby on Samsung S8
- Mobile Apps: Google Allo

ii. Predictions while Commuting
    - Traffic prediction
    - Online Transportation networks

iii. Videos Surveillance

iv. Social Media Services

v. Email Spam and Malware Filtering

vi. Online Customer Support

vii. Search Engine Result Refining

viii. Product Recommendations

ix. Online Fraud Detection

4.2.1 Supervised Learning:

- Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples.
- In supervised learning, each example is a pair consisting of an input object and a desired output value (also called the supervisory signal).
- A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances.
- Supervised learning is fairly common in classification problems because the goal is often to get the computer to learn a classification system that we have created. Digit recognition, once again, is a common example of classification learning.
- More generally, classification learning is appropriate for any problem where deducing a classification is useful and the classification is easy to determine.

- In some cases, it might not even be necessary to give pre-determined classifications to every instance of a problem if the agent can work out the classifications for itself.
- This would be an example of unsupervised learning in a classification context. Supervised learning is the most common technique for training neural networks and decision trees. Both of these techniques are highly dependent on the information given by the pre-determined classifications.
- In the case of neural networks, the classification is used to determine the error of the network and then adjust the network to minimize it, and in decision trees, the classifications are used to determine what attributes provide the most information that can be used to solve the classification puzzle
    - Supervised Learning:
    - Use training data to infer model
    - apply model to test data
    - e.g. Maximum likelihood, Perceptron, SVM
    - Unsupervised Learning:
    - No training data
    - Model inference and application
    - both rely on test
    - data exclusively
    - e.g. k-means
- **For instance**, suppose you are given an basket filled with different kinds of fruits. Now the first step is to train the machine with all different fruits one by one like this:
    - If shape of object is rounded and depression at top having color Red then it will be labelled as – Apple.
    - If shape of object is long curving cylinder having color Green-Yellow then it will be labelled as – Banana.
- Now suppose after training the data, you have given a new separate fruit say Banana from basket and asked to identify it.
- Since machine has already learnt the things from previous data and this time have to use it wisely. It will first classify the fruit with its shape and color, and would confirm the fruit name as BANANA and put it in Banana category.
- Thus machine learns the things from training data (basket containing fruits) and then apply the knowledge to test data (new fruit).
-

Supervised learning classified into two categories of algorithms:

1. **Classification:** A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.
2. **Regression:** A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

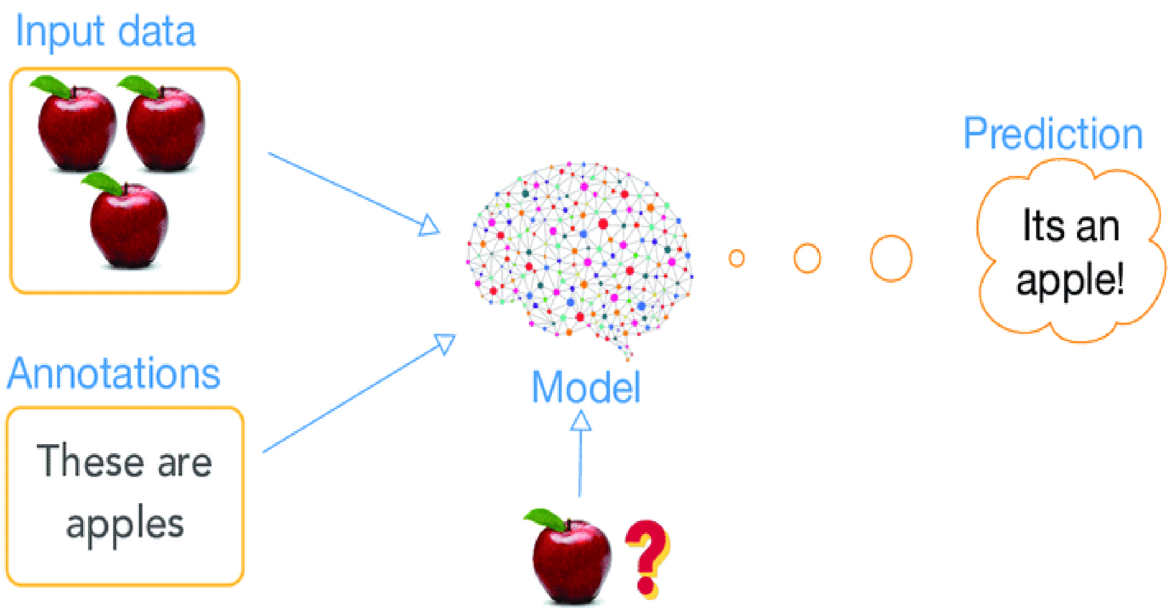

Figure: Supervised Learning

4.2.2 Unsupervised Learning:

- Unsupervised learning seems much harder: the goal is to have the computer learn how to do something that we don't tell it how to do! There are actually two approaches to unsupervised learning.
- The first approach is to teach the agent not by giving explicit categorizations, but by using some sort of reward system to indicate success.
- Note that this type of training will generally fit into the decision problem framework because the goal is not to produce a classification but to make decisions that maximize rewards.
- This approach nicely generalizes to the real world, where agents might be rewarded for doing certain actions and punished for doing others. Often, a form of reinforcement learning can be used for unsupervised learning, where the agent bases its actions on the previous rewards and punishments without necessarily even learning any information about the exact ways that its actions affect the world.

- In a way, all of this information is unnecessary because by learning a reward function, the agent simply knows what to do without any processing because it knows the exact reward it expects to achieve for each action it could take.
- This can be extremely beneficial in cases where calculating every possibility is very time consuming (even if all of the transition probabilities between world states were known).
- On the other hand, it can be very time consuming to learn by, essentially, trial and error.
- **For instance**, suppose it is given an image having both dogs and cats which have not seen ever.
- Thus machine has no any idea about the features of dogs and cat so we can't categorize it in dogs and cats. But it can categorize them according to their similarities, patterns and differences i.e., we can easily categorize the above picture into two parts.
- First first may contain all pics having **dogs** in it and second part may contain all pics having **cats** in it. Here you didn't learn anything before, means no training data or examples.

Unsupervised learning classified into two categories of algorithms:

1. **Clustering:** A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
2. **Association:** An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

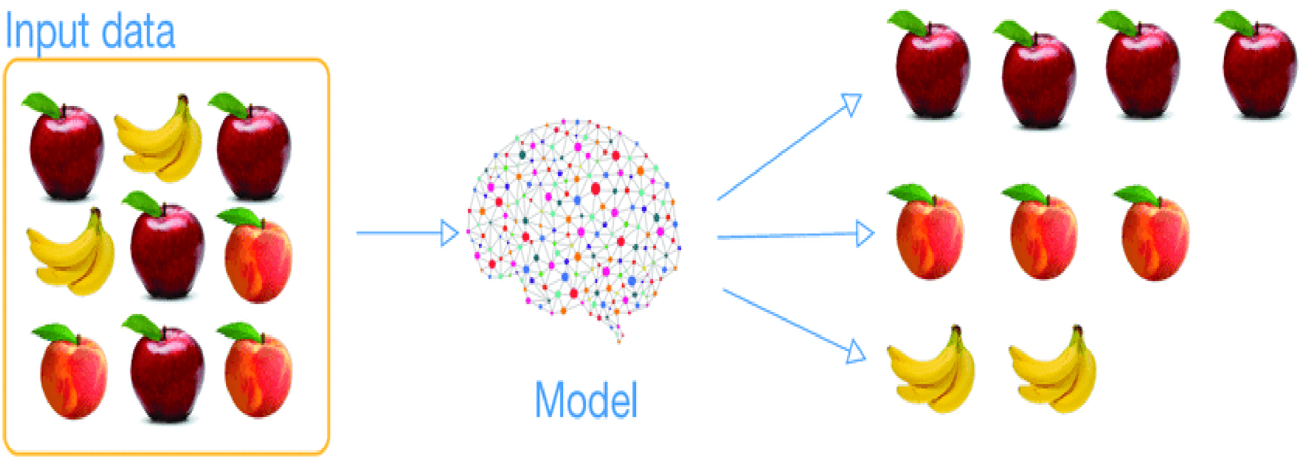

Figure: Unsupervised Learning

Difference between Supervised Learning and Unsupervised Learning:

PARAMETER	SUPERVISED LEARNING	UNSUPERVISED LEARNING
Input Data	Uses Known and Labeled Data as input	Uses Unknown Data as input
Computational Complexity	Very Complex	Less Computational Complexity
Real Time	Uses off-line analysis	Uses Real Time Analysis of Data
Number of Classes	Number of Classes are known	Number of Classes are not known
Accuracy of Results	Accurate and Reliable Results	Moderate Accurate and Reliable Results

In order to solve a given problem of supervised learning, one has to perform the following steps

1. Determine the type of training examples. Before doing anything else, the user should decide what kind of data is to be used as a training set. In the case of handwriting analysis, for example, this might be a single handwritten character, an entire handwritten word, or an entire line of handwriting.
2. Gather a training set. The training set needs to be representative of the real-world use of the function. Thus, a set of input objects is gathered and corresponding outputs are also gathered, either from human experts or from measurements.
3. Determine the input feature representation of the learned function. The accuracy of the learned function depends strongly on how the input object is represented. Typically, the input object is transformed into a feature vector, which contains a number of features that are descriptive of the object. The number of features should not be too large, because of the curse of dimensionality; but should contain enough information to accurately predict the output.
4. Determine the structure of the learned function and corresponding learning algorithm. For example, the engineer may choose to use support vector machines or decision trees.
5. Complete the design. Run the learning algorithm on the gathered training set. Some supervised learning algorithms require the user to determine certain control parameters. These

parameters may be adjusted by optimizing performance on a subset (called a *validation* set) of the training set, or via cross-validation

6. Evaluate the accuracy of the learned function. After parameter adjustment and learning, the performance of the resulting function should be measured on a test set that is separate from the training set.

– A wide range of supervised learning algorithms is available, each with its strengths and weaknesses. There is no single learning algorithm that works best on all supervised learning problems.

4.3 Structure of Regression Model:

– Regression is a data mining (machine learning) technique used to fit an equation to a dataset. The simplest form of regression, linear regression, uses the formula of a straight line (y = mx + b) and determines the appropriate values for m and b to predict the value of y based upon a given value of x.
– Advanced techniques, such as multiple regressions, allow the use of more than one input variable and allow for the fitting of more complex models, such as a quadratic equation.
– The two basic types of regression are linear regression and multiple regressions. Linear regression uses one independent variable to explain and/or predict the outcome of Y, while multiple regressions use two or more independent variables to predict the outcome.
– The general form of each type of regression is:

Linear Regression: $Y = a + bX + u$

Multiple Regression: $Y = a + b_1X_1 + b_2X_2 + B_3X_3 + ... + B_tX_t + u$

Where:

Y= the variable that we are trying to predict

X= the variable that we are using to predict Y

a= the intercept

b= the slope

u= the regression residual

In multiple regression the separate variables are differentiated by using subscripted numbers.

– Regression takes a group of random variables, thought to be predicting Y, and tries to find a mathematical relationship between them.

- This relationship is typically in the form of a straight line (linear regression) that best approximates all the individual data points. Regression is often used to determine how much specific factors such as the price of a commodity, interest rates, particular industries or sectors influence the price movement of an asset.

4.8 support Vector Machine:

- In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.
- Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting).
- An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall
- In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.
- When data is unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups.
- Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in.
- There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes.
- So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized.
- If such a hyperplane exists, it is known as the maximum-margin hyperplane and the linear classifier it defines is known as a maximum-margin classifier; or equivalently, the perceptron of optimal stability.

- support-vector machine constructs a hyperplane or set of hyperplane in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection.
- Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier.