



CompleteNotes

Download all study materials for B.Tech, M.Tech,
Diploma and other courses for free.
Get notes & question paper PDFs completely free.



JOIN US ON TELEGRAM
t.me/completenotesofficial

www.completenotes.in

Department of Computer Science and Engineering

Subject Notes

CS 601- Machine Learning

UNIT-I

Introduction to machine learning:

Machine learning is a tool for turning information into knowledge. Machine learning techniques are used to automatically find the valuable underlying patterns within complex data that we would otherwise struggle to discover. The hidden patterns and knowledge about a problem can be used to predict future events and perform all kinds of complex decision making.

Tom Mitchell gave a “well-posed” mathematical and relational definition that “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.”

For Example:

A checkers learning problem:

Task(T): Playing checkers.

Performance measures (P): Performance of games won.

Training Experience (E): Playing practice games against itself.

Need For Machine Learning

- Ever since the technical revolution, we've been generating an immeasurable amount of data.
- With the availability of so much data, it is finally possible to build predictive models that can study and analyse complex data to find useful insights and deliver more accurate results.
- Top Tier companies such as Netflix and Amazon build such Machine Learning models by using tons of data in order to identify profitable opportunities and avoid unwanted risks.

ML Vs AI Vs DL

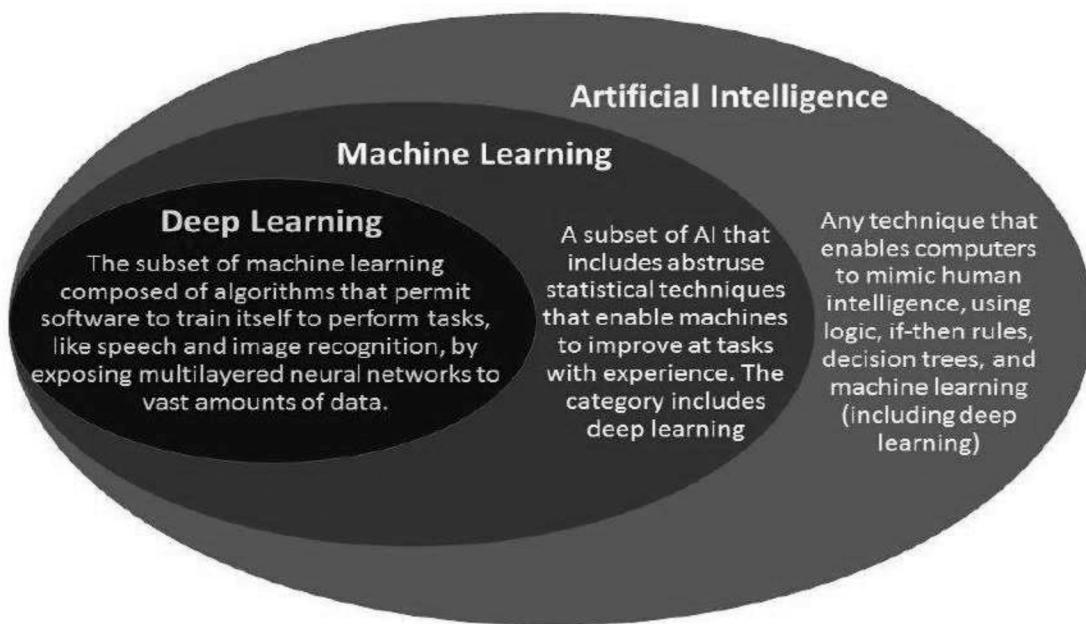


Figure: 1.1

Important Terms of Machine Learning

- **Algorithm:** A Machine Learning algorithm is a set of rules and statistical techniques used to learn patterns from data and draw significant information from it. It is the logic behind a Machine Learning model. An example of a Machine Learning algorithm is the Linear Regression algorithm.

- **Model:** A model is the main component of Machine Learning. A model is trained by using a Machine Learning Algorithm. An algorithm maps all the decisions that a model is supposed to take based on the given input, in order to get the correct output.
- **Predictor Variable:** It is a feature(s) of the data that can be used to predict the output.
- **Response Variable:** It is the feature or the output variable that needs to be predicted by using the predictor variable(s).
- **Training Data:** The Machine Learning model is built using the training data. The training data helps the model to identify key trends and patterns essential to predict the output.
- **Testing Data:** After the model is trained, it must be tested to evaluate how accurately it can predict an outcome. This is done by the testing data set.

Note: A Machine Learning process begins by feeding the machine lots of data, by using this data the machine is trained to detect hidden insights and trends. These insights are then used to build a Machine Learning Model by using an algorithm in order to solve a problem in Figure 1.2.

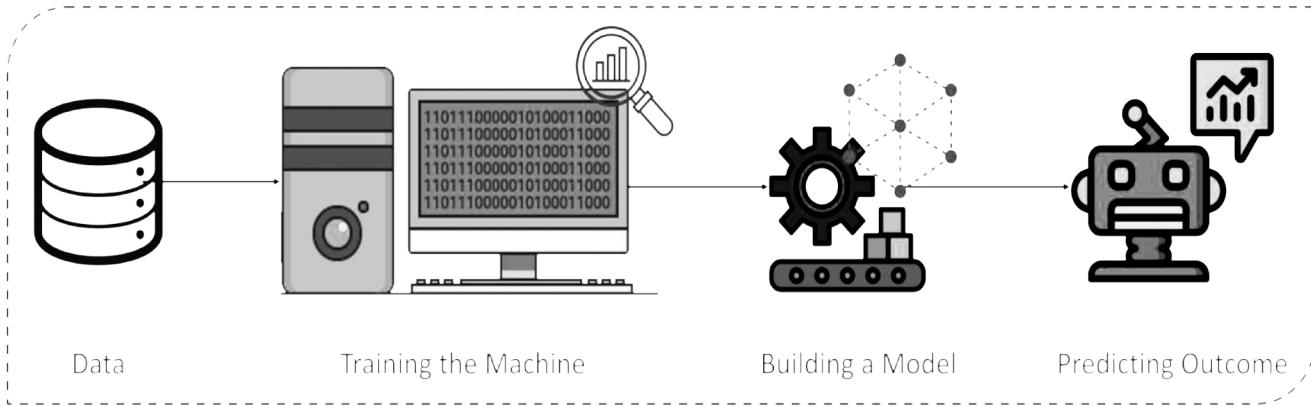


Figure: 1.2

Scope

- **Increase in Data Generation:** Due to excessive production of data, need a method that can be used to structure, analyze and draw useful insights from data. This is where Machine Learning comes in. It uses data to solve problems and find solutions to the most complex tasks faced by organizations.
- **Improve Decision Making:** By making use of various algorithms, Machine Learning can be used to make better business decisions.

For example, Machine Learning is used to forecast sales, predict downfalls in the stock market, identify risks and anomalies, etc.

- **Uncover patterns & trends in data:** Finding hidden patterns and extracting key insights from data is the most essential part of Machine Learning. By building predictive models and using statistical techniques, Machine Learning allows you to dig beneath the surface and explore the data at a minute scale. Understanding data and extracting patterns manually will take days, whereas Machine Learning algorithms can perform such computations in less than a second.
- **Solve complex problems:** Building self-driving cars, Machine Learning can be used to solve the most complex problems.

Limitations

1. What algorithms exist for learning general target function from specific training examples?
2. In what setting will particular algorithm converge to the desired function, given sufficient training data?
3. Which algorithm performs best for which types of problems and representations?

4. How much training data is sufficient?
5. When and how can prior knowledge held by the learner guide the process of generalizing from examples?
6. What is the best way to reduce the learning task to one more function approximation problem?
7. Machine Learning Algorithms Require Massive Stores of Training Data.
8. Labeling Training Data Is a Tedium Process.
9. Machines Cannot Explain Themselves.

Machine Learning Types:

A Machine can learn to solve a problem by any one of the following three approaches.

These are the ways in which a machine can learn:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Regression

Regression models are used to predict a continuous value. Predicting prices of a house given the features of house like size, price etc. is one of the common examples of Regression. It is a supervised technique.

Types of Regression

1. Simple Linear Regression
2. Polynomial Regression
3. Support Vector Regression
4. Decision Tree Regression
5. Random Forest Regression

Simple Linear Regression

This is one of the most common and interesting type of Regression technique. Here we predict a target variable Y based on the input variable X. A linear relationship should exist between target variable and predictor and so comes the name Linear Regression.

Consider predicting the salary of an employee based on his/her age. We can easily identify that there seems to be a correlation between employee's age and salary (more the age more is the salary). The hypothesis of linear regression is- $Y = a + bX$

Y represents salary, X is employee's age and a and b are the coefficients of equation. So in order to predict Y (salary) given X (age), we need to know the values of a and b (the model's coefficients).

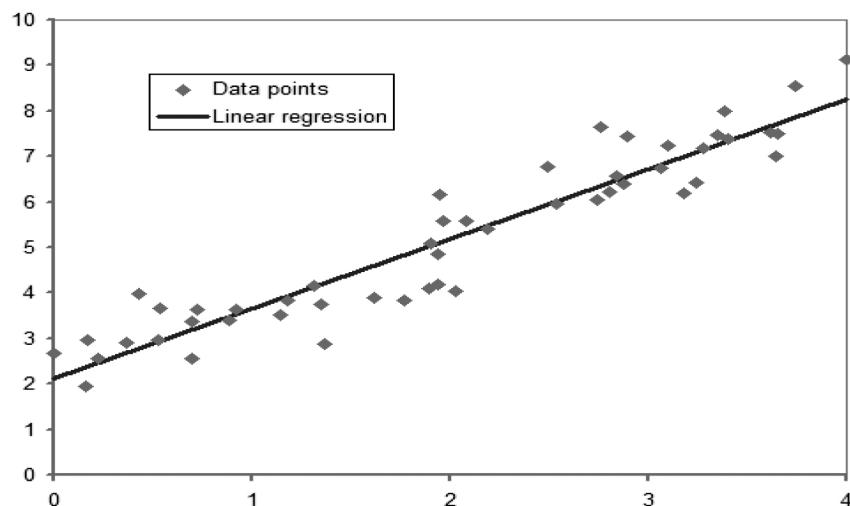


Figure: 1.3 Linear Regression

Polynomial Regression

In polynomial regression, we transform the original features into polynomial features of a given degree and then apply Linear Regression on it. Consider the above linear model $Y = a + bX$ is transformed to something like – $Y = a + bX + cX^2$

It is still a linear model but the curve is now quadratic rather than a line. Scikit-Learn provide Polynomial Features class to transform the features.

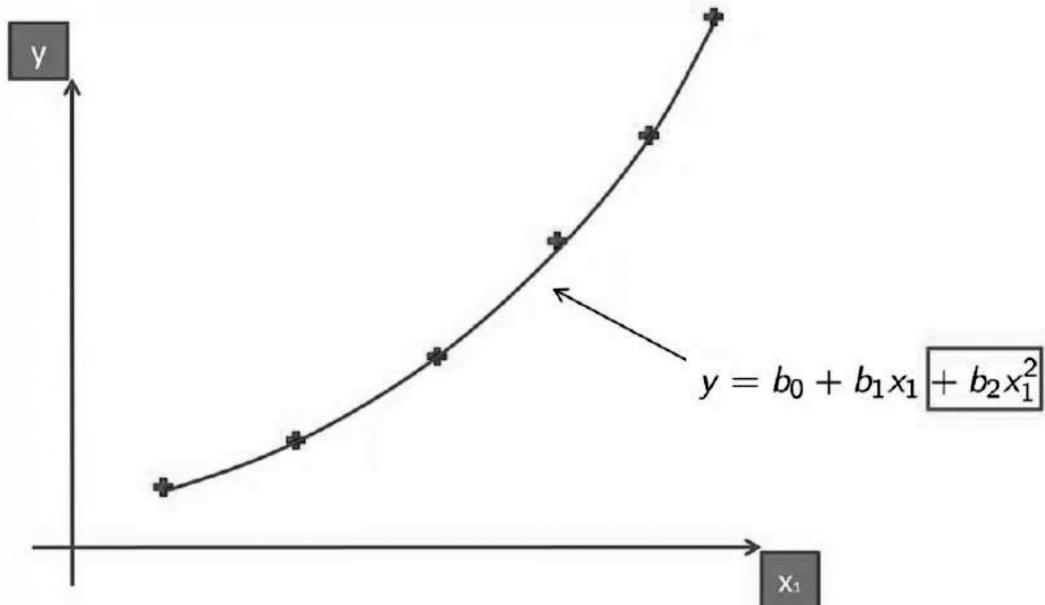
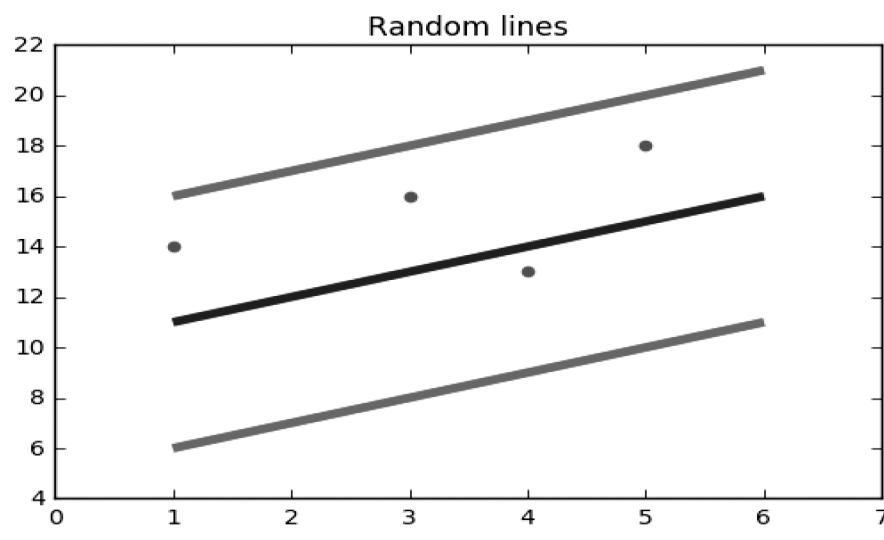


Figure: 1.4 Polynomial Regression

If we increase the degree to a very high value, the curve becomes overfitted as it learns the noise in the data as well.

Support Vector Regression

In SVR, we identify a hyper plane with maximum margin such that maximum numbers of data points are within that margin. SVRs are almost similar to SVM classification algorithm. Instead of minimizing the error rate as in simple linear regression, we try to fit the error within a certain threshold. Our objective in SVR is to basically consider the points that are within the margin. Our best fit line is the hyper plane that has maximum number of points.



Data points within the boundary line

Figure: 1.5 Support Vector Regression

Decision Tree Regression

Decision trees can be used for classification as well as regression. In decision trees, at each level we need to identify the splitting attribute. In case of regression, the ID3 algorithm can be used to identify the splitting node by reducing standard deviation.

A decision tree is built by partitioning the data into subsets containing instances with similar values (homogenous). Standard deviation is used to calculate the homogeneity of a numerical sample. If the numerical sample is completely homogeneous, its standard deviation is zero.

Random Forest Regression

Random forest is an ensemble approach where we take into account the predictions of several decision regression trees.

1. Select K random points
2. Identify n where n is the number of decision tree regressors to be created. Repeat step 1 and 2 to create several regression trees.
3. The average of each branch is assigned to leaf node in each decision tree.
4. To predict output for a variable, the average of all the predictions of all decision trees are taken into consideration.

Random Forest prevents over fitting (which is common in decision trees) by creating random subsets of the features and building smaller trees using these subsets.

Probability

probability is an intuitive concept. We use it on a daily basis without necessarily realising that we are speaking and applying probability to work.

Life is full of uncertainties. We don't know the outcomes of a particular situation until it happens. Will it rain today? Will I pass the next math test? Will my favorite team win the toss? Will I get a promotion in next 6 months? All these questions are examples of uncertain situations we live in. Let us map them to few common terminologies are-

- Experiment – are the uncertain situations, which could have multiple outcomes. Whether it rains on a daily basis is an experiment.
- Outcome is the result of a single trial. So, if it rains today, the outcome of today's trial from the experiment is "It rained"
- Event is one or more outcome from an experiment. "It rained" is one of the possible event for this experiment.
- Probability is a measure of how likely an event is. So, if it is 60% chance that it will rain tomorrow, the probability of Outcome "it rained" for tomorrow is 0.6

Random Variables

To calculate the likelihood of occurrence of an event, we need to put a framework to express the outcome in numbers. We can do this by mapping the outcome of an experiment to numbers.

Let's define X to be the outcome of a coin toss.

X = outcome of a coin toss

Possible Outcomes:

- 1 if heads
- 0 if tails

Let's take another one.

Suppose, I win the game if I get a sum of 8 while rolling two fair dice. I can define my random variable Y to be (the sum of the upward face of two fair dice)

Y can take values = (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)

A few things to note about random variables:

- Each value of the random variable may or may not be equally likely. There is only 1 combination of dice, with sum 2{(1,1)}, while a sum of 5 can be achieved by {(1,4), (2,3), (3,2), (4,1)}. So, 5 is more likely to occur as compared to 2. On the contrary, the likelihood of a head or a tail in a coin toss is equal and 50-50.
- Sometimes, the random variables can only take fixed values, or values only in a certain interval. For example in a dice, the top face will only show values between 1 and 6. It cannot take a 2.25 or a 1.5. Similarly, when a coin is flipped, it can only show heads and tails and nothing else. On the other hand, if I define my random variable to be the amount of sugar in orange. It can take any value like 1.4g, 1.45g, 1.456g, 1.4568g as so on. All these values are possible and all infinite values between them are also possible. So, in this case, the random variable is continuous with a possibility of all real numbers.
- Don't think random variable as a traditional variable (even though both are called variables) like $y=x+2$, where the value of y is dependent on x . Random variable is defined in terms of the outcome of a process. We quantify the process using the random variable.

Statistic

Machine learning and statistics are two tightly related fields of study. So much so that statisticians refer to machine learning as “applied statistics” or “statistical learning” rather than the computer-science-centric name.

Raw observations alone are data, but they are not information or knowledge.

Data raises questions, such as:

- What is the most common or expected observation?
- What are the limits on the observations?
- What does the data look like?

Although they appear simple, these questions must be answered in order to turn raw observations into information that we can use and share.

Beyond raw data, we may design experiments in order to collect observations. From these experimental results we may have more sophisticated questions, such as:

- What variables are most relevant?
- What is the difference in an outcome between two experiments?
- Are the differences real or the result of noise in the data?

Questions of this type are important. The results matter to the project, to stakeholders, and to effective decision making.

Statistical methods are required to find answers to the questions that we have about data.

We can see that in order to both understand the data used to train a machine learning model and to interpret the results of testing different machine learning models, that statistical methods are required. **Statistics is a subfield of mathematics.**

It refers to a collection of methods for working with data and using data to answer questions.

Descriptive Statistics

Descriptive statistics refer to methods for summarizing raw observations into information that we can understand and share.

Commonly, we think of descriptive statistics as the calculation of statistical values on samples of data in order to summarize properties of the sample of data, such as the common expected value (e.g. the mean or median) and the spread of the data (e.g. the variance or standard deviation).

Descriptive statistics may also cover graphical methods that can be used to visualize samples of data. Charts and graphics can provide a useful qualitative understanding of both the shape or distribution of observations as well as how variables may relate to each other.

Inferential Statistics

Inferential statistics is a fancy name for methods that aid in quantifying properties of the domain or population from a smaller set of obtained observations called a sample.

Commonly, we think of inferential statistics as the estimation of quantities from the population distribution, such as the expected value or the amount of spread.

More sophisticated statistical inference tools can be used to quantify the likelihood of observing data samples given an assumption. These are often referred to as tools for statistical hypothesis testing, where the base assumption of a test is called the null hypothesis.

Linear Algebra

Linear Algebra is a branch of mathematics that lets you concisely describe coordinates and interactions of planes in higher dimensions and perform operations on them and concerned with vectors, matrices, and linear transforms.

Although linear algebra is integral to the field of machine learning, the tight relationship is often left unexplained or explained using abstract concepts such as vector spaces or specific matrix operations.

Linear Algebra is required -

- When working with data, such as tabular datasets and images.
- When working with data preparation, such as one hot encoding and dimensionality reduction.
- The ingrained use of linear algebra notation and methods in sub-fields such as deep learning, natural language processing, and recommender systems.

Examples of linear algebra in machine learning-

1. Dataset and Data Files
2. Images and Photographs
3. Linear Regression
4. Regularization
5. Principal Component Analysis
6. Singular-Value Decomposition
7. Latent Semantic Analysis
8. Recommender Systems
9. Deep Learning

For instance-

Images and Photographs

1. Perhaps you are more used to working with images or photographs in computer vision applications.
2. Each image that you work with is itself a table structure with a width and height and one pixel value in each cell for black and white images or 3 pixel values in each cell for a color image.
3. A photo is yet another example of a matrix from linear algebra.
4. Operations on the image, such as cropping, scaling, shearing, and so on are all described using the notation and operations of linear algebra.

Linear Regression

1. Linear regression is an old method from statistics for describing the relationships between variables.
2. It is often used in machine learning for predicting numerical values in simpler regression problems.
3. There are many ways to describe and solve the linear regression problem, i.e. finding a set of coefficients that when multiplied by each of the input variables and added together results in the best prediction of the output variable.

Convex Optimization

Optimization is a big part of machine learning. It is the core of most popular methods, from least squares regression to artificial neural networks.

These methods useful in the core implementation of a machine learning algorithm. It is required to implement own algorithm tuning scheme to optimize the parameters of a model for some cost function.

A good example may be the case where we want to optimize the hyper-parameters of a blend of predictions from an ensemble of multiple child models.

Machine learning algorithms use optimization all the time. We minimize loss, or error, or maximize some kind of score functions. Gradient descent is the "hello world" optimization algorithm covered on probably any machine learning course. It is obvious in the case of regression, or classification models, but even with tasks such as clustering we are looking for a solution that *optimally* fits our data (e.g. k-means minimizes the within-cluster sum of squares). So if you want to understand how the machine learning algorithms do work, learning more about optimization helps. Moreover, if you need to do things like hyper parameter tuning, then you are also directly using optimization.

Data visualization

Data visualization is an important skill in applied statistics and machine learning.

Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding.

This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral to yourself and stakeholders than measures of association or significance.

There are five key plots that need to know well for basic data visualization. They are:

- Line Plot
- Bar Chart
- Histogram Plot
- Box and Whisker Plot
- Scatter Plot

With knowledge of these plots, you can quickly get a qualitative understanding of most data that you come across.

Line Plot

A line plot is generally used to present observations collected at regular intervals.

The x-axis represents the regular interval, such as time. The y-axis shows the observations, ordered by the x-axis and connected by a line.

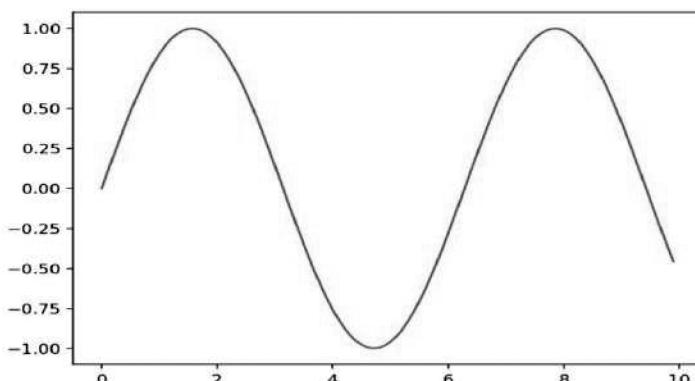


Figure: 1.6 Line Plot

Bar Chart

A bar chart is generally used to present relative quantities for multiple categories.

The x-axis represents the categories and are spaced evenly. The y-axis represents the quantity for each category and is drawn as a bar from the baseline to the appropriate level on the y-axis.

A bar chart can be created by calling the `bar()` function and passing the category names for the x-axis and the quantities for the y-axis.

Bar charts can be useful for comparing multiple point quantities or estimations.

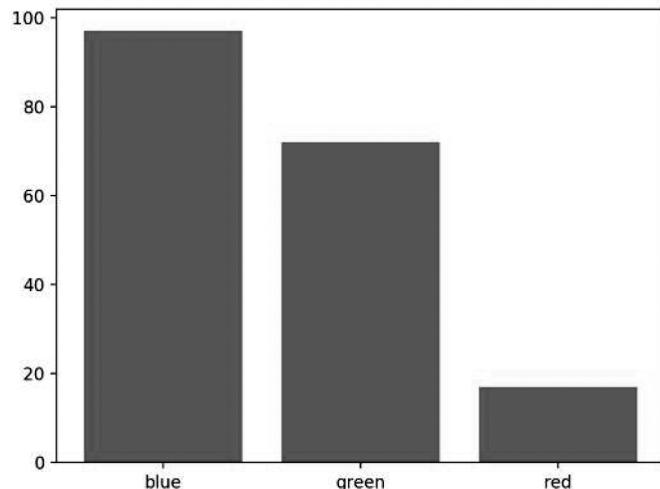


Figure: 1.7 Bar Chart

Histogram Plot

A histogram plot is generally used to summarize the distribution of a data sample.

The x-axis represents discrete bins or intervals for the observations. For example observations with values between 1 and 10 may be split into five bins, the values [1,2] would be allocated to the first bin, [3,4] would be allocated to the second bin, and so on.

The y-axis represents the frequency or count of the number of observations in the dataset that belong to each bin.

Essentially, a data sample is transformed into a bar chart where each category on the x-axis represents an interval of observation values.

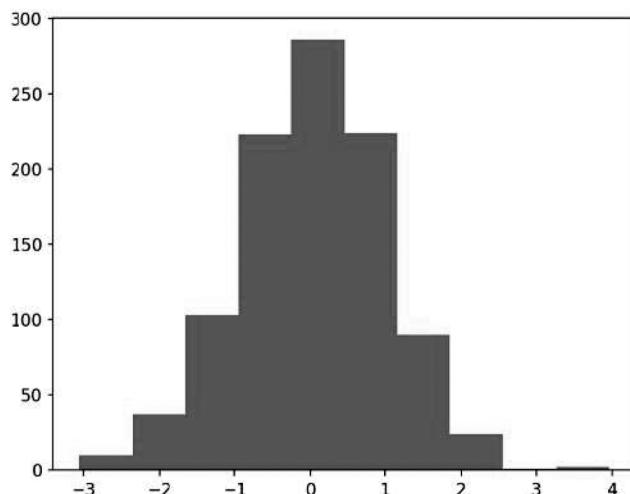


Figure: 1.7 Histogram Plot

Scatter Plot

A scatter plot (or ‘scatterplot’) is generally used to summarize the relationship between two paired data samples.

Paired data samples means that two measures were recorded for a given observation, such as the weight and height of a person.

The x-axis represents observation values for the first sample, and the y-axis represents the observation values for the second sample. Each point on the plot represents a single observation.

Scatter plots are useful for showing the association or correlation between two variables. A correlation can be quantified, such as a line of best fit, that too can be drawn as a line plot on the same chart, making the relationship clearer.

A dataset may have more than two measures (variables or columns) for a given observation. A scatter plot matrix is a cart containing scatter plots for each pair of variables in a dataset with more than two variables.

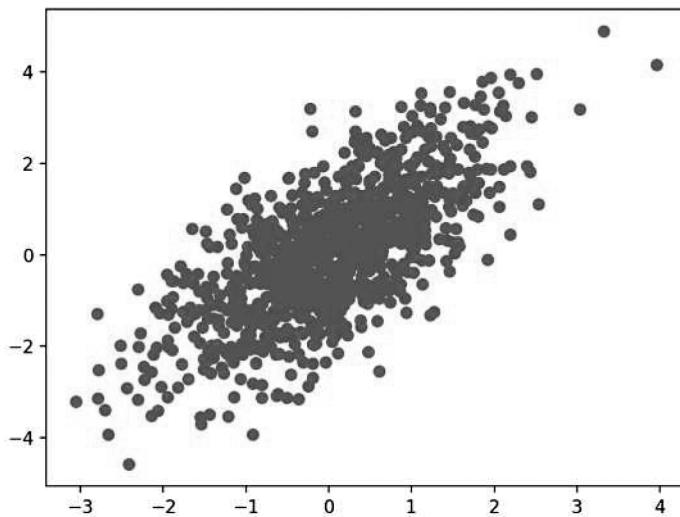


Figure: 1.7 Scatter Plot

Hypothesis function and testing

Hypothesis testing is a statistical method that is used in making statistical decisions using experimental data. Hypothesis Testing is basically an assumption that we make about the population parameter.

Ex : you say avg student in class is 40 or a boy is taller than girls.

all those example we assume need some statistic way to prove those. we need some mathematical conclusion what ever we are assuming is true.

Hypothesis testing is an essential procedure in statistics. A **hypothesis test** evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data. When **we** say that a finding is statistically significant, it's thanks to a **hypothesis test**.

The process of hypothesis testing is to draw inferences or some conclusion about the overall population or data by conducting some statistical tests on a sample.

For drawing some inferences, we have to make some assumptions that lead to two terms that are used in the hypothesis testing.

- **Null hypothesis:** It is regarding the assumption that there is no anomaly pattern or believing according to the assumption made.
- **Alternate hypothesis:** Contrary to the null hypothesis, it shows that observation is the result of real effect.

Some of widely used hypothesis testing types :-

1. T Test (Student T test)
2. Z Test
3. ANOVA Test
4. Chi-Square Test

Data Distributions

A sample of data will form a distribution, and by far the most well-known distribution is the Gaussian distribution, often called the Normal distribution.

The distribution provides a parameterized mathematical function that can be used to calculate the probability for any individual observation from the sample space. This distribution describes the grouping or the density of the observations, called the probability density function. We can also calculate the likelihood of an observation having a value equal to or lesser than a given value. A summary of these relationships between observations is called a cumulative density function.

Distributions

From a practical perspective, we can think of a distribution as a function that describes the relationship between observations in a sample space.

For example, we may be interested in the age of humans, with individual ages representing observations in the domain, and ages 0 to 125 the extent of the sample space. The distribution is a mathematical function that describes the relationship of observations of different heights.

A distribution is simply a collection of data, or scores, on a variable. Usually, these scores are arranged in order from smallest to largest and then they can be presented graphically.

Density Functions

Distributions are often described in terms of their density or density functions.

Density functions are functions that describe how the proportion of data or likelihood of the proportion of observations changes over the range of the distribution.

Two types of density functions are probability density functions and cumulative density functions.

- Probability Density function: calculates the probability of observing a given value.
- Cumulative Density function: calculates the probability of an observation equal or less than a value.

A probability density function, or PDF, can be used to calculate the likelihood of a given observation in a distribution. It can also be used to summarize the likelihood of observations across the distribution's sample space. Plots of the PDF show the familiar shape of a distribution, such as the bell-curve for the Gaussian distribution.

Data Pre-processing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm.

- Data pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

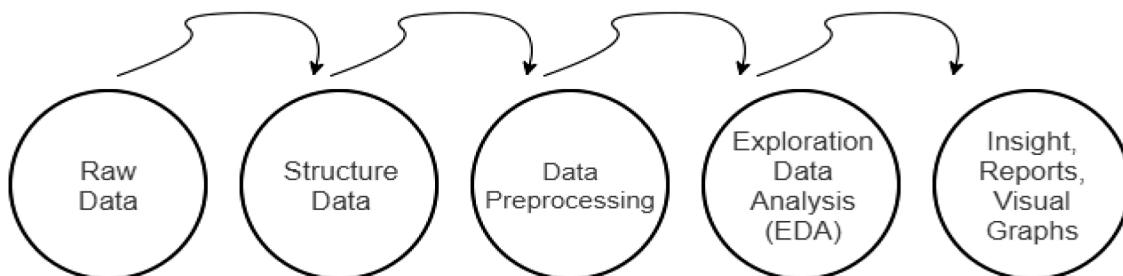


Figure: 1.8 Data Pre-Processing

Need of Data Pre-processing

- For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set.
- Another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

Data Augmentation

Data augmentation is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data. For instance we can consider image, so in image there are various ways to transform and augment the image data.

Need for data augmentation

Data augmentation is an integral process in deep learning, as in deep learning we need large amounts of data and in some cases it is not feasible to collect thousands or millions of images, so data augmentation comes to the rescue. It helps us to increase the size of the dataset and introduce variability in the dataset.

Operations in data augmentation

The most commonly used operations are-

1. Rotation
2. Shearing
3. Zooming
4. Cropping
5. Flipping
6. Changing the brightness level

Normalizing Data Sets

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values. For machine learning, every dataset does not require normalization. It is required only when features have different ranges.

The goal of normalization is to transform features to be on a similar scale. This improves the performance and training stability of the model.

Four common normalization techniques may be useful:

- scaling to a range
- clipping
- log scaling
- z-score

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. Normalization is also required for some algorithms to model the data correctly.

For example, assume your input dataset contains one column with values ranging from 0 to 1, and another column with values ranging from 10,000 to 100,000. The great difference in the scale of the numbers could cause problems when you attempt to combine the values as features during modelling. *Normalization* avoids these problems by creating new values that maintain the general distribution and ratios in the source data, while keeping values within a scale applied across all numeric columns used in the model.

Machine Learning Models

Types of classification algorithms in Machine Learning:

1. Linear Classifiers: Logistic Regression, Naive Bayes Classifier
2. Nearest Neighbour
3. Support Vector Machines
4. Decision Trees
5. Boosted Trees
6. Random Forest
7. Neural Networks

Naive Bayes Classifier (Generative Learning Model):

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Nearest Neighbour:

The k-nearest-neighbour algorithm is a classification algorithm, and it is supervised: it takes a bunch of labelled points and uses them to learn how to label other points. To label a new point, it looks at the labelled points closest to that new point (those are its nearest neighbours), and has those neighbour vote, so whichever label the most of the neighbours have is the label for the new point (the "k" is the number of neighbour it checks).

Logistic Regression (Predictive Learning Model):

It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. This is better than other binary classification like nearest neighbour since it also explains quantitatively the factors that lead to classification.

Decision Trees:

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches and a leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

Random Forest:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

Neural Network:

A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its

output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand.

Types of Machine Learning

Machine learning is sub-categorized to three types:

1. **Supervised Learning – Train Me!**
2. **Unsupervised Learning – I am self-sufficient in learning**
3. **Reinforcement Learning – My life My rules! (Hit & Trial)**

Supervised Learning

Supervised Learning is the one, where you can consider the learning is guided by a teacher. We have a dataset which acts as a teacher and its role is to train the model or the machine. Once the model gets trained it can start making a prediction or decision when new data is given to it.

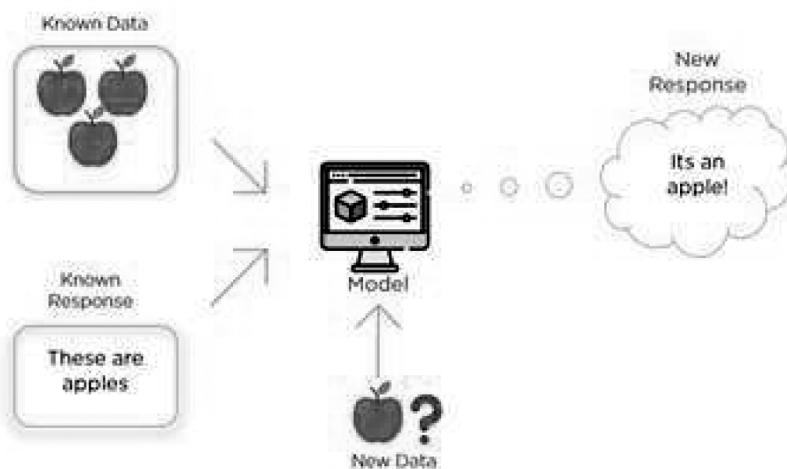


Figure 1.9 Supervised Learning

Unsupervised Learning

The model learns through observation and finds structures in the data. Once the model is given a dataset, it automatically finds patterns and relationships in the dataset by creating clusters in it. What it cannot do is add labels to the cluster; like it cannot say this a group of apples or mangoes, but it will separate all the apples from mangoes.

Suppose we presented images of apples, bananas and mangoes to the model, so what it does, based on some patterns and relationships it creates clusters and divides the dataset into those clusters. Now if a new data is fed to the model, it adds it to one of the created clusters.

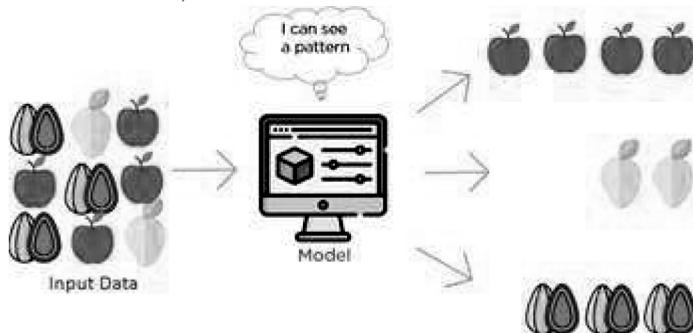


Figure 1.10 Un-Supervised Learning

Reinforcement Learning

It is the ability of an agent to interact with the environment and find out what is the best outcome. It

follows the concept of hit and trial method. The agent is rewarded or penalized with a point for a correct or a wrong answer, and on the basis of the positive reward points gained the model trains itself. And again once trained it gets ready to predict the new data presented to it.

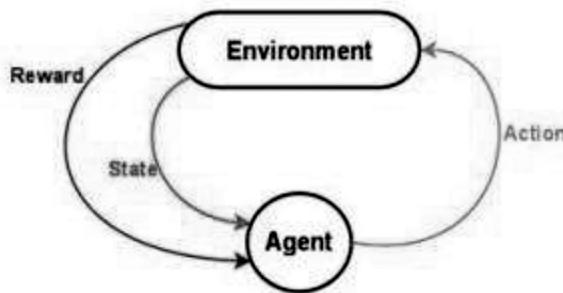


Figure 1.11 Un-Supervised Learning

Types of Machine Learning

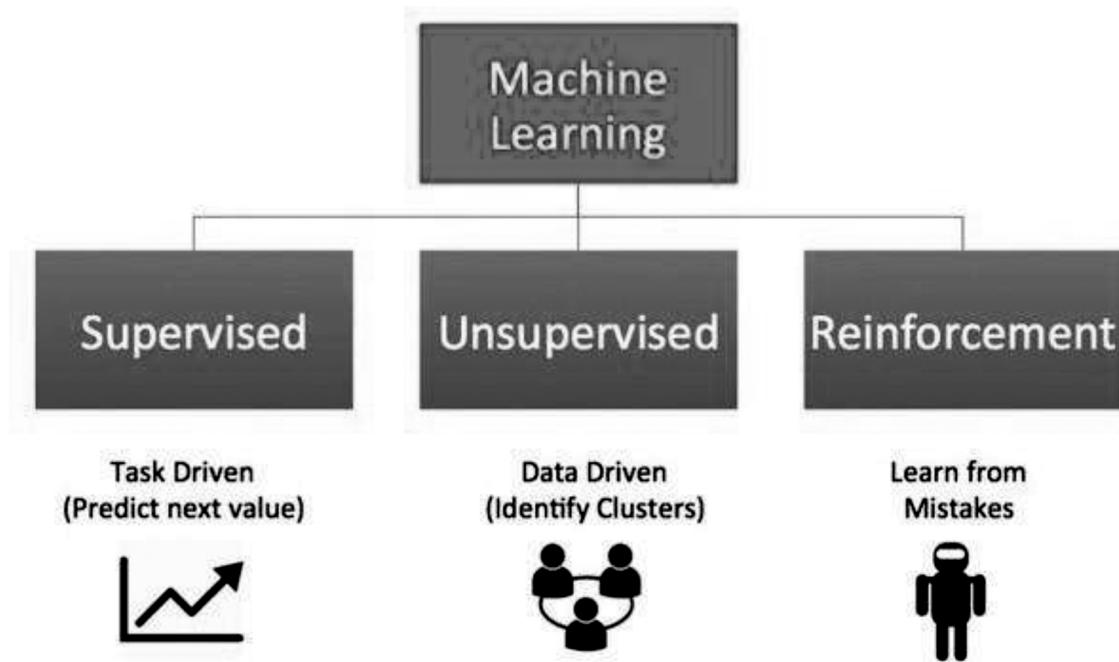


Figure: 1.12 Types of Machine Learning