Chapter 4: Session 11

Estimation models -Linear regression

Regression Techniques

Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent (target) and independent variable (s) (predictor). This technique is used for forecasting, time series modelling and finding the causal effect relationship between the variables. For example, relationship between rash driving and number of road accidents by a driver is best studied through regression.

Regression analysis is an important tool for modelling and analyzing data. Here, we fit a curve / line to the data points, in such a manner that the differences between the distances of data points from the curve or line is minimized.

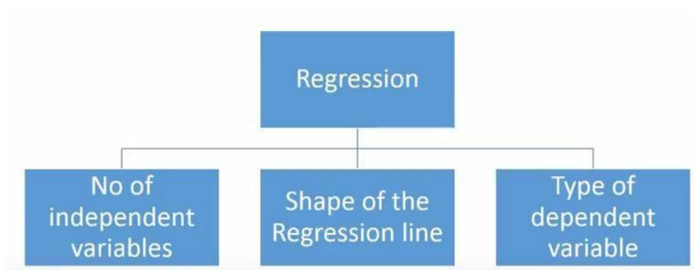
Why do we use Regression Analysis?

There are multiple benefits of using regression analysis. They are as follows:

It indicates the significant relationships between dependent variable and independent variable. It indicates the strength of impact of multiple independent variables on a dependent variable.

How many types of regression techniques do we have?

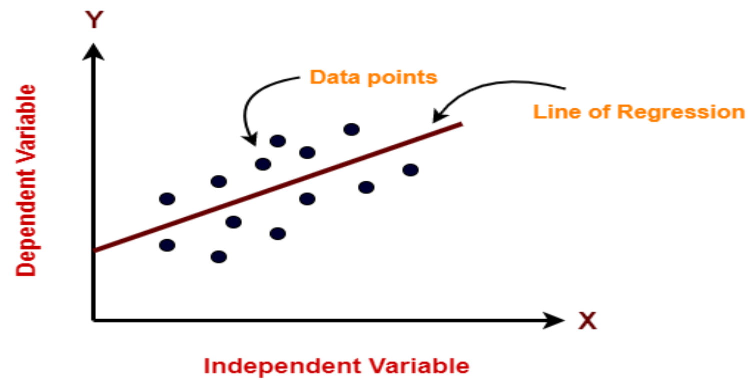
There are various kinds of regression techniques available to make predictions. These techniques are mostly driven by three metrics (number of independent variables, type of dependent variables and shape of regression line).



Linear Regression

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line).

It is represented by an equation **Y=a+b*X + e**, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).



Dependent variable: the variable we wish to explain, the main factor you are trying to understand and predict

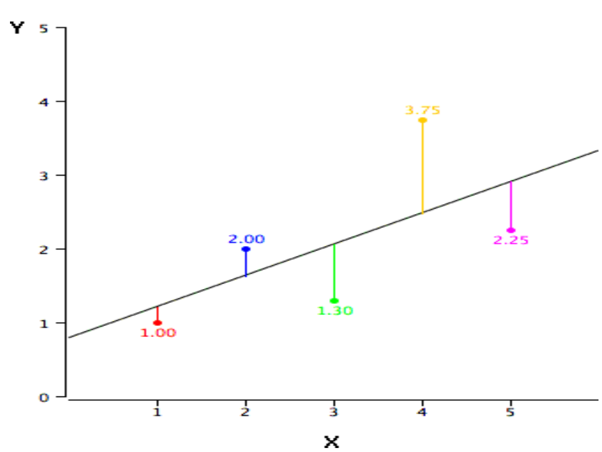
Independent variable: the variable used to explain the dependent variable, the factors that might influence the dependent variable.

How to obtain best fit line (Value of a and b)?

This task can be easily accomplished by Least Square Method. It is the most common method used for fitting a regression line. It calculates the best-fit line for the observed data by minimizing the sum of the

squares of the vertical deviations from each data point to the line. Because the deviations are first squared, when added, there is no cancelling out between positive and negative values.

$$\min_{w} ||Xw - y||_2{}^2$$



Chapter 4: Session 12

When to use Linear Regression?

Linear Regression's power lies in its simplicity, which means that it can be used to solve problems across various fields. At first, the data collected from the observations need to be collected and plotted along a line. If the difference between the predicted value and the result is almost the same, we can use linear regression for the problem.

Assumptions in linear regression

If you are planning to use linear regression for your problem then there are some assumptions you need to consider:

- The relation between the dependent and independent variables should be almost linear.
- The data is homoscedastic, meaning the variance between the results should not be too much.
- The results obtained from an observation should not be influenced by the results obtained from the previous observation.
- The residuals should be normally distributed. This assumption means that the probability density function of the residual values is normally distributed at each independent value.

You can determine whether your data meets these conditions by plotting it and then doing a bit of digging into its structure.

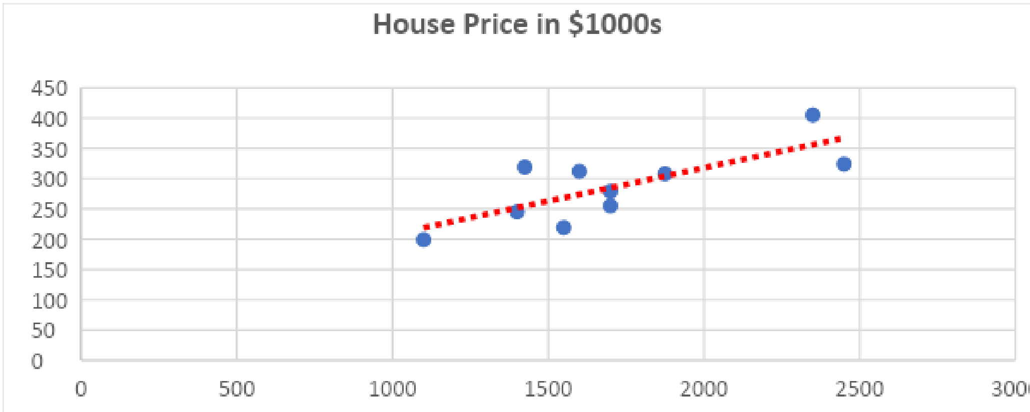
Three major uses for regression analysis are:

1. Determining the strength of predictors- To identify the strength of the effect that the independent variable(s) have on a dependent variable-What is the strength of relationship between dose and effect, sales and marketing spending, or age and income.
2. Forecasting an effect or impact of changes- How much the dependent variable changes with a change in one or more independent variables- "how much additional sales income do I get for each additional $1000 spent on marketing?"
3. Trend forecasting - "what will the price of gold be in 6 months?"

Linear Regression: Use Case

A real estate agent wishes to examine the relationship between the selling price of a home and its size (measured in square feet)

Dependent variable (Y) = House price in $1000s , Independent variable (X) = Square feet



Chapter 4: Session 13

Logistic Regression

What Is Logistic Regression?

Logistic regression is a statistical method for analysing a dataset 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). It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables.

Summary: Logistic Regression is a tool for classifying and making predictions between zero and one. Coefficients are long odds. Odds are relative so when interpreting coefficients you need to set a baseline to compare in both numeric and categorical variables.

What is the probability that your customer will return next year? What has a greater impact on conversion rates: source of acquisition or time on site? These binary / probabilistic questions can be answered by logistic regression. Logistic regression…

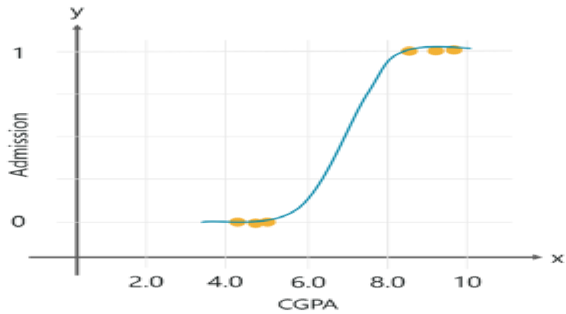
- Is used in classification problems like retention, conversion, likelihood to purchase, etc.
- Can use continuous (numeric) or categorical (buckets / categories) as independent variables (inputs to the model).
- Has a binary (0 / 1) dependent variable (variable you're trying to predict).
- Should not be trained on data with perfect separation (it breaks the mathematical underpinnings).

Why and When Do We Use Logistic Regression?

In order to understand why we use logistic regression, let's consider a small scenario.

Let's say that your little sister is trying to get into grad school, and you want to predict whether she'll get admitted in her dream establishment. So, based on her CGPA and the past data, you can use Logistic Regression to foresee the outcome.

Logistic Regression allows you to analyze a set of variables and predict a categorical outcome. Since here we need to predict whether she will get into the school or not, which is a classification problem, logistic regression would be ideal.



You might be wondering why we're not using Linear Regression in this case. The reason is that linear regression is used to predict a continuous quantity rather than a categorical one. So, when the resultant outcome can take only 2 possible values, it is only sensible to have a model that predicts the value either as 0 or 1 or in a probability form that ranges between 0 and 1.

Some Familiar Example of Logistics Regression:

Some prominent examples like:

- Email Spam Filter: Spam /No Spam
- Fraud Detection: Transaction is fraudulent, Yes/No
- Tumour: Benign/Malignant

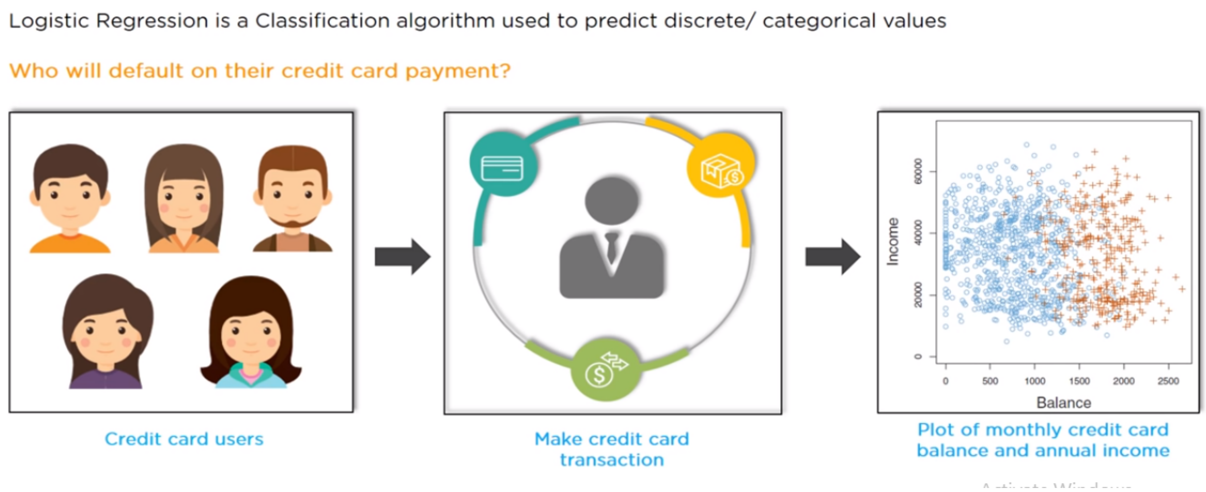
Marketing

Every day, when you browse your Facebook newsfeed, the powerful algorithms running behind the scene predict whether or not you would be interested in certain content (which could be, for instance, an advertisement). Such algorithms can be viewed as complex variations of Logistic Regression algorithms where the question to be answered is simple – will the user like this particular advertisement in his/her news feed?

Types of Logistic Regression:

- Binary logistic regression: It has only two possible outcomes. Example- yes or no
- Multinomial logistic regression: It has three or more nominal categories. Example - cat, dog, elephant.
- Ordinal logistic regression - It has three or more ordinal categories, ordinal meaning that the categories will be in a order. Example- user ratings (1–5).

Logistic Regression-Use Case :



Chapter 4: Session 14

Model Evaluation

Model Evaluation

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and overfitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation.

Hold-Out

Cross-Validation

When only a limited amount of data is available, to achieve an unbiased estimate of the model performance we use k-fold cross-validation. In k-fold cross-validation, we divide the data into k subsets of equal size. We build models k times, each time leaving out one of the subsets from training and use it as the test set. If k equals the sample size, this is called "leave-one-out".

Model evaluation can be divided to two sections:

- Classification Evaluation
- Regression Evaluation

Classification Evaluation

Confusion Matrix

A confusion matrix shows the number of correct and incorrect predictions made by the classification model compared to the actual outcomes (target value) in the data. The matrix is

NxN, where N is the number of target values (classes). Performance of such models is commonly evaluated using the data in the matrix. The following table displays a 2x2 confusion matrix for two classes (Positive and Negative).

Confusion Matrix		Target			
		Positive	Negative		
Model	Positive	a	b	*Positive Predictive Value*	a/(a+b)
	Negative	c	d	*Negative Predictive Value*	d/(c+d)
		Sensitivity	*Specificity*	**Accuracy** = (a+d)/(a+b+c+d)	
		a/(a+c)	d/(b+d)		

- **Accuracy**: the proportion of the total number of predictions that were correct.
- **Positive Predictive Value** or **Precision**: the proportion of positive cases that were correctly identified.
- **Negative Predictive Value**: the proportion of negative cases that were correctly identified.
- **Sensitivity** or **Recall**: the proportion of actual positive cases which are correctly identified.
- **Specificity**: the proportion of actual negative cases which are correctly identified.

Gain and Lift Charts

Gain or lift is a measure of the effectiveness of a classification model calculated as the ratio between the results obtained with and without the model. Gain and lift charts are visual aids for evaluating performance of classification models.

Chapter 5: Session 15

Concepts to Artificial Neural Networks

What is an Artificial neural networks? (Biologically inspired Simulations)

Artificial neural networks are the computational model that are inspired by the human brain.

The inventor of the first neuro computer, Dr. Robert Hecht-Nielsen, defines a neural network as −"a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs.”

Basic Structure of ANNs

The idea of ANNs is based on the belief that working of human brain by making the right connections, can be imitated using silicon and wires as living **neurons** and **dendrites**.

ANNs are composed of multiple **nodes**, which imitate biological **neurons** of human brain. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its **activation** or **node value.**

Each link is associated with **weight.** ANNs are capable of learning, which takes place by altering weight values

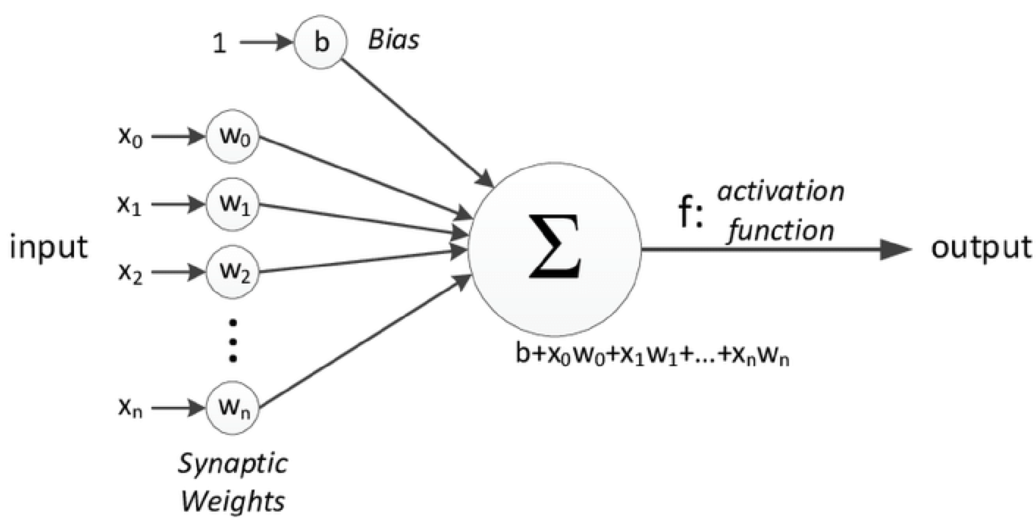
Artificial neural networks and Biological nervous system

- ANNs are programs designed to solve any problem by trying mimic the structure and the function of our nervous system.
- Neural networks are based on simulated neurons, which are joined together in a variety of ways to form networks
- Neural network resembles the human brain in the following two ways:

  i) A neural network acquires knowledge through learning

  ii) A neural network's knowledge is stored within the interconnection strengths known as synaptic weight.

The following diagram represents the general model of ANN which is inspired by a biological neuron. It is also called Perceptron. A single layer neural network is called a Perceptron. It gives a single output.



In the above figure, for one single observation, x0, x1, x2, x3...x(n) represents various inputs (independent variables) to the network. Each of these inputs is multiplied by a connection weight or synapse. The weights are represented as w0, w1, w2, w3….w(n) . Weight shows the strength of a particular node.

b is a bias value. A bias value allows you to shift the activation function up or down.

In the simplest case, these products are summed, fed to a transfer function (activation function) to generate a result, and this result is sent as output.

Mathematically, x1.w1 + x2.w2 + x3.w3 ...... xn.wn = ∑ xi.wi

Now activation function is applied $\phi(\sum xi.wi)$

Activation function

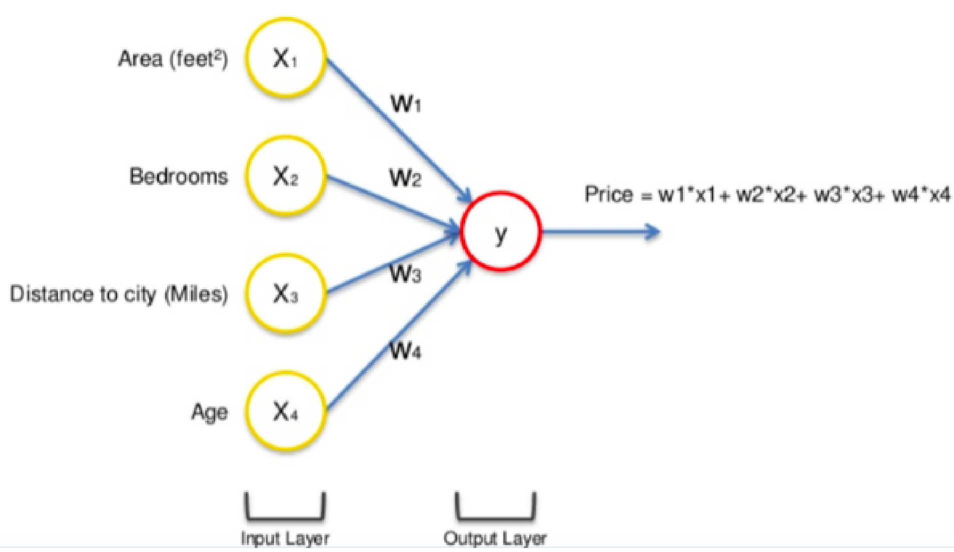
The Activation function is important for an ANN to learn and make sense of something really complicated. Their main purpose is to convert an input signal of a node in an ANN to an output signal. This output signal is used as input to the next layer in the stack.

Activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The motive is to introduce non-linearity into the output of a neuron.

If we do not apply activation function then the output signal would be simply linear function(one-degree polynomial). Now, a linear function is easy to solve but they are limited in their complexity, have less power. Without activation function, our model cannot learn and model complicated data such as images, videos, audio, speech, etc.

How does the Neural network work?

Let us take the example of the price of a property and to start with we have different factors assembled in a single row of data: Area, Bedrooms, Distance to city and Age.



The input values go through the weighted synapses straight over to the output layer. All four will be analyzed, an activation function will be applied, and the results will be produced.

Advantages:

- It involves human like thinking.
- They handle noisy or missing data.
- They can work with large number of variables or parameters.
- They provide general solutions with good predictive accuracy.

- System has got property of continuous learning.
- They deal with the non-linearity in the world in which we live.

Application of ANN

- Process modeling and control
- Machine Diagnostics
- Portfolio Management
- Target Recognition
- Medical Diagnosis
- Credit Rating
- Targeted Marketing
- Voice recognition
- Face recognition
- Financial Forecasting
- Intelligent searching
- Fraud detection

Chapter 5: Session 16

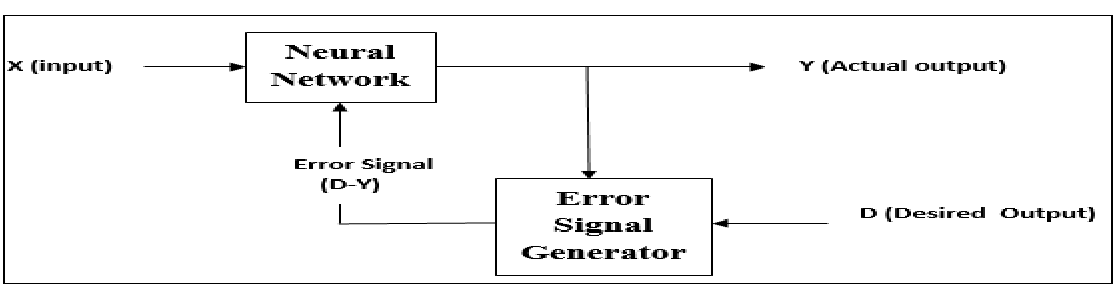
Algorithms to ANN

Learning in ANN can be classified into three categories namely supervised learning, unsupervised learning, and reinforcement learning.

Supervised Learning

As the name suggests, this type of learning is done under the supervision of a teacher. This learning process is dependent.

During the training of ANN under supervised learning, the input vector is presented to the network, which will give an output vector. This output vector is compared with the desired output vector. An error signal is generated, if there is a difference between the actual output and the desired output vector. On the basis of this error signal, the weights are adjusted until the actual output is matched with the desired output.

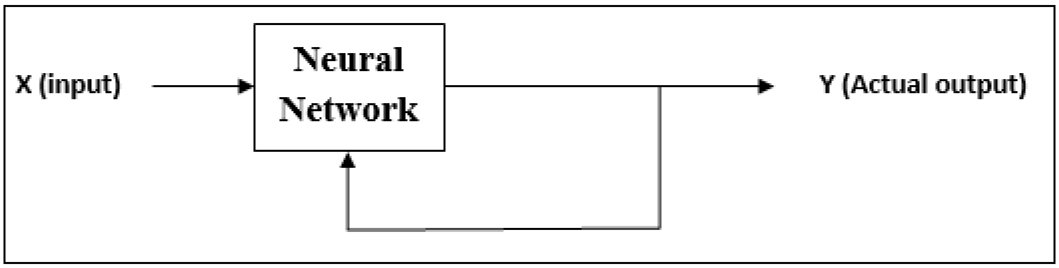


Unsupervised Learning

As the name suggests, this type of learning is done without the supervision of a teacher. This learning process is independent.

During the training of ANN under unsupervised learning, the input vectors of similar type are combined to form clusters. When a new input pattern is applied, then the neural network gives an output response indicating the class to which the input pattern belongs.

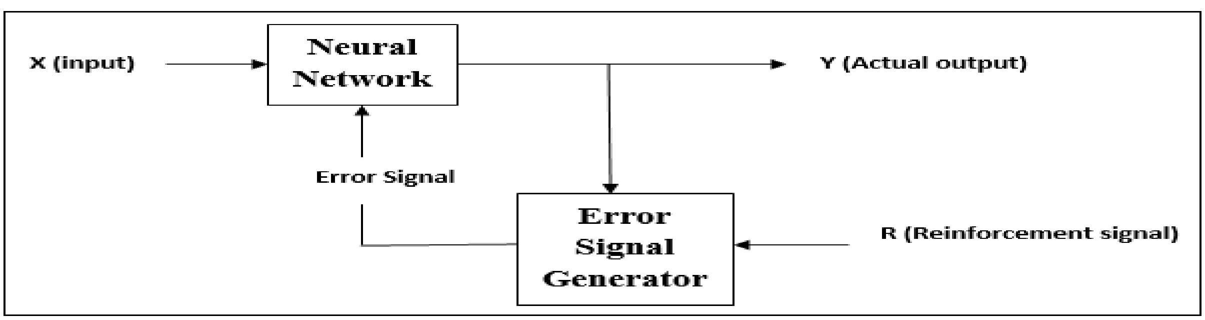
There is no feedback from the environment as to what should be the desired output and if it is correct or incorrect. Hence, in this type of learning, the network itself must discover the patterns and features from the input data, and the relation for the input data over the output.



Reinforcement Learning

As the name suggests, this type of learning is used to reinforce or strengthen the network over some critic information. This learning process is similar to supervised learning, however we might have very less information.

During the training of network under reinforcement learning, the network receives some feedback from the environment. This makes it somewhat similar to supervised learning. However, the feedback obtained here is evaluative not instructive, which means there is no teacher as in supervised learning. After receiving the feedback, the network performs adjustments of the weights to get better critic information in future.



Chapter 5: Session 17

PREPROCESSING DATA FOR NEURAL NETWORKS

Ways to handle input data effectively and efficiently in developing neural networks.

INPUT DATA SELECTION

Data selection can be a demanding and intricate task. After all, a neural network is only as good as the input data used to train it. If important data inputs are missing, then the effect on the neural network's performance can be significant. Developing a workable neural network application can be considerably more difficult without a solid understanding of the problem domain. When selecting input data, the implications of following a market theory should be kept in mind. Existing market inefficiencies can be noted quantitatively by making use of artificial intelligence tools.

Individual perspective on the markets also influences the choice of input data. Technical analysis suggests the use of only single-market price data as inputs, while conversely, fundamental analysis concentrates solely on data inputs that reflect supply/ demand and economic factors. In today's global environment, neither approach alone is sufficient for financial forecasting. Instead, synergistic market analysis combines both approaches with intermarket analysis within a quantitative framework using neural networks. This overcomes the limitations of interpreting intermarket relationships through simple visual analysis of price charts and carries conceptualization of intermarket analysis to its logical conclusion.

PREPROCESSING INPUT DATA

Once the most appropriate raw input data has been selected, it must be pre-processed; otherwise, the neural network will not produce accurate forecasts. The decisions made in this phase of development are critical to the performance of a network.

Transformation and normalization are two widely used pre-processing methods. Transformation involves manipulating raw data inputs to create a single input to a net, while normalization is a transformation performed on a single data input to distribute the data evenly and scale it into an acceptable range for the network. Knowledge of the domain is important in choosing pre-processing methods to highlight underlying features in the data, which can increase the network's ability to learn the association between inputs and outputs.

Some simple pre-processing methods include computing differences between or taking ratios of inputs. This reduces the number of inputs to the network and helps it learn more easily. In financial forecasting, transformations that involve the use of standard technical indicators should also be considered. Moving averages, for example, which are utilized to help smooth price data, can be useful as a transform.

Data normalization is the final pre-processing step. In normalizing data, the goal is to ensure that the statistical distribution of values for each net input and output is roughly uniform. In addition, the values should be scaled to match the range of the input neurons. This means that along with any other transformations performed on network inputs, each input should be normalized as well.

Chapter 5: Session 18

Backpropagation

How do Neural networks learn?

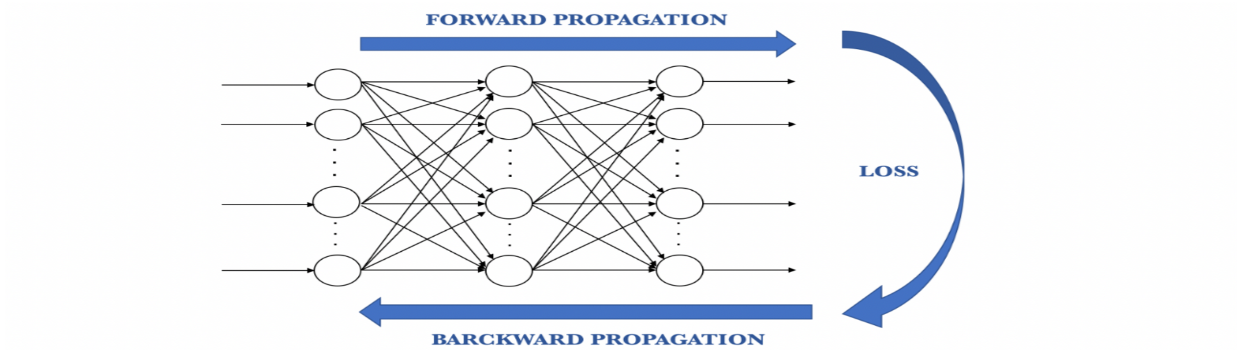
Cost Function: One half of the squared difference between actual and output value.

For each layer of the network, the cost function is analyzed and used to adjust the threshold and weights for the next input. Our aim is to minimize the cost function. The lower the cost function, the closer the actual value to the predicted value. In this way, the error keeps becoming marginally lesser in each run as the network learns how to analyze values.

We feed the resulting data back through the entire neural network. The weighted synapses connecting input variables to the neuron are the only thing we have control over.

As long as there exists a disparity between the actual value and the predicted value, we need to adjust those wights. Once we tweak them a little and run the neural network again, A new Cost function will be produced, hopefully, smaller than the last.

We need to repeat this process until we scrub the cost function down to as small as possible.



The procedure described above is known as Back-propagation and is applied continuously through a network until the error value is kept at a minimum.

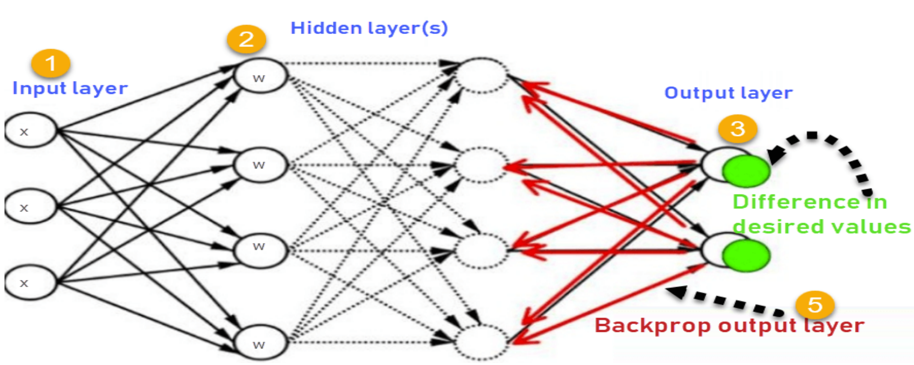
What is Backpropagation?

Back-propagation is the essence of neural net training. It is the method of fine-tuning the weights of a neural net based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and to make the model reliable by increasing its generalization.

Backpropagation is a short form for "backward propagation of errors." It is a standard method of training artificial neural networks. This method helps to calculate the gradient of a loss function with respects to all the weights in the network.

How Backpropagation Works: Simple Algorithm

Consider the following diagram



1. Inputs X, arrive through the preconnected path
2. Input is modeled using real weights W. The weights are usually randomly selected.
3. Calculate the output for every neuron from the input layer, to the hidden layers, to the output layer.
4. Calculate the error in the outputs

$Error_B$= Actual Output – Desired Output

5. Travel back from the output layer to the hidden layer to adjust the weights such that the error is decreased.

Keep repeating the process until the desired output is achieved

Why We Need Backpropagation?
Most prominent advantages of Backpropagation are:

- Backpropagation is fast, simple and easy to program
- It has no parameters to tune apart from the numbers of input
- It is a flexible method as it does not require prior knowledge about the network
- It is a standard method that generally works well
- It does not need any special mention of the features of the function to be learned.

Chapter 6: Session 19

Time Series Modeling

We all must have heard that people are saying that the price of different objects has decreased or increased with time, these different objects could be anything like petrol, diesel, gold, silver, eatable things, etc.

Also, the rate of interest fluctuates in banks and different for different kinds of loans. What are all this data, how it is useful? These types of data are time-series data that go through analysis for forecasts.

Because of the tremendous variety of conditions, time-series used by both nature and human beings for communication, description, and data visualizations. Also, time is the physical quantity, and elements, coefficients, parameters, and characteristics of time-series data are mathematical quantities, so time-series can have real-time or real-world interpretations.

Introduction

We are going to examine what is time series analysis, its scope in the future, how this can be used in several repetitions of financial data and services, and time series analysis using machine learning.

In the broad form, it is analyzed to obtain inference what has occurred in the past with the data point series and endeavor to predict what is going to appear in the coming time.

Now the questions arise how do people get to know that the price of an object as increased or decreased over time, they do so by comparing the price of an object over a set of the time period.

An ordered set of observations with respect to time periods is a time series. In simple words, a sequential organization of data accordingly to their time of occurrence is termed as time series.

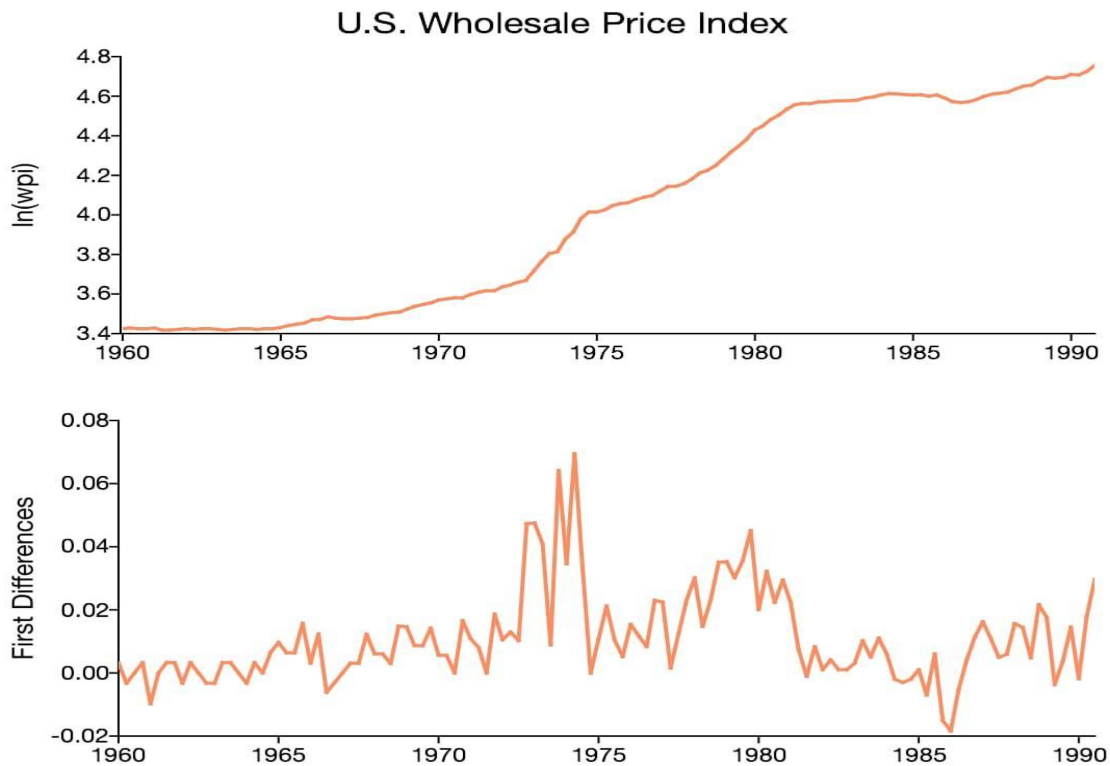
Time acts as a reference point in relation to the entire procedure. It can be noticed that time-series always depicts a relationship between two variables in which one is time and the other one is any quantitative variable, not necessarily there is an increment in the change of variable with respect to time in the observations, it also exhibited decrement in variable-time observational data.

For example, the temperature of a particular area at a particular time increases or decreases accordingly.

Time series data can be found in economics, social sciences, finance, epidemiology, and the physical sciences.

Field	**Example topics**
Economics	Gross Domestic Product (GDP), Consumer Price Index (CPI), S&P 500 Index, and unemployment rates
Social sciences	Birth rates, population, migration data, political indicators
Epidemiology	Disease rates, mortality rates, mosquito populations
Medicine	Blood pressure tracking, weight tracking, cholesterol measurements, heart rate monitoring
Physical sciences	Global temperatures, monthly sunspot observations, pollution levels.

Time Series Visualization



An overview of Statistical Time Series Analysis

A time-series contains sequential data points mapped at a certain successive time duration, it incorporates the methods that attempt to surmise a time series in terms of understanding either the underlying concept of the data points in the time series or suggesting or making predictions.

Forecasting data using time-series analysis comprises the use of some significant model to forecast future conclusions on the basis of known past outcomes. An example of a restaurant in which prediction is made on the number of customers as when will more customers appear in the restaurant at a specified time duration based on the previous appearance of customers with time.

Broadly specified time-series models are **Autoregressive (AR) Models, Integrated (I) models, Moving Average(MA) models**, and some other models are the combination of these models such as **Autoregressive Moving Average (ARMA) models, and Autoregressive Integrated Moving Average (ARIMA) models.**

These models reflect measurements near concurrently in time will be more closely relevant as compared to measurements distant apart.

Implementing Time Series Analysis in Machine Learning

It is a well-known fact that Machine Learning is a powerful technique in imagining, speech and natural processing for a huge explicated dataset available, on the other hand, problems based on time series do not have usually interpreted datasets, even as data is collected from various sources so exhibit substantial variations in terms of features, properties, attributes, temporal scales, and dimensionality.

Time series analysis requires such sorting algorithms that can allow it to learn time-dependent patterns across multiples models different from images and speech. Various machine learning tools such as classification, clustering, forecasting, and anomaly detection depend upon real-world business applications.

Among various defined applications, discussing here Time series forecasting, it is an important area of machine learning because there are multiple problems involving time components for making predictions.

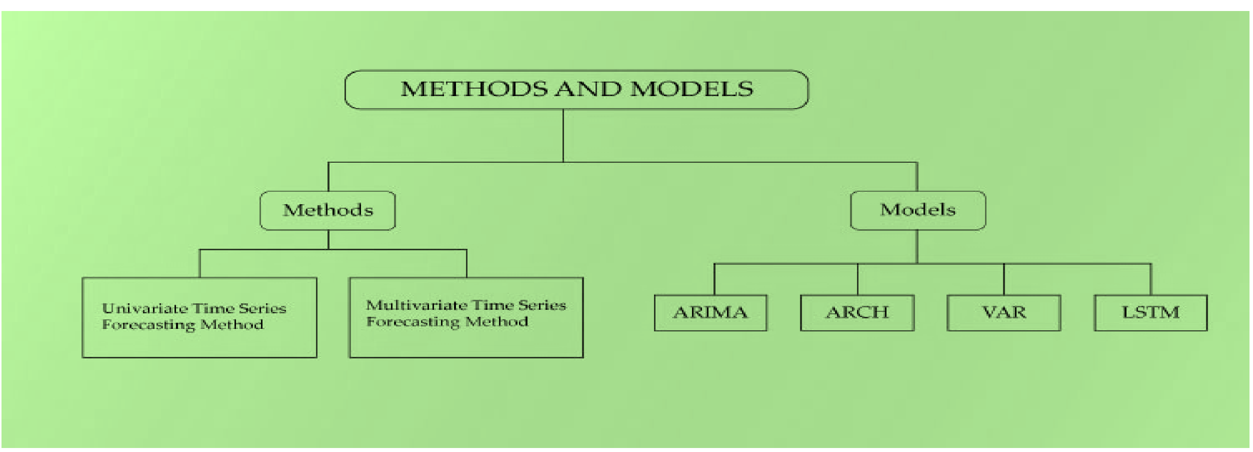
There are multiple models and methods used as approaches for time series forecasting, let's understand them more clearly;

Methods

In the **Univariate Time-series Forecasting method**, forecasting problems contain only two variables in which one is time and the other is the field we are looking to forecast. For example, if you want to predict the mean temperature of a city for the coming week, now one parameter is time(week) and the other is a city.

On the other hand, in the **Multivariate Time-series Forecasting method**, forecasting problems contain multiple variables keeping one variable as time fixed and others will be multiple in parameters.

Consider the same example, predicting the temperature of a city for the coming week, the only difference would come here now temperature will consider impacting factors such as rainfall and time duration of raining, humidity, wind speed, precipitation, atmospheric pressure, etc, and then the temperature of the city will be predicted accordingly. All these factors are related to temperature and impact it vigorously.



Models

ARIMA Model: As mentioned in the above section, it is a combination of three different models itself, AR, MA and I, where "AR" reflects the evolving variable of interest is regressed on its own prior values, "MA" infers that the regression error is the linear combination of error terms values happened at various stages of time priorly, and "I" shows the data values are replaced by the difference between their values and the previous values. Combinedly "ARIMA" tries to fit the data into the model, and also ARIMA depends on the accuracy over a broad width of time series.

ARCH/GARCH Model: Being the extended model of its common version GARCH, Autoregressive Conditional Heteroscedasticity (ARCH) is the most volatile model for time series forecasting, and are well trained for catching dynamic variations of volatility from time series.

Vector Autoregressive Model or VAR model: It gives the independencies between various time-series data which as a generalization of the Univariate Autoregression Model.

LSTM: Long-Short Term Memory (LSTM) is a deep learning model, it is a kind of Recurrent Neural Network(RNN) to read the sequence dependencies. It enables us to handle long structures during training the dataset and creates predictions according to previous data.

Conclusion

We can use Time Series for multiple investigations to predict future as circadian rhythms, seasonal behaviors, trends, changes, etc. to interrogate the questions like predicted values, what is leading and lagging, connections and association, control, repetitions, and hidden pattern, etc. Time series analysis is basically the recording of data at a regular interval of time, which could lead to taking a versed decision, crucial for trade and so have multiple applications such as Stock Market and Trends analysis, Financial forecasting, Inventory analysis, Census Analysis, Yield prediction, Sales forecasting, etc.

Multiple applications of the Time-Series Analysis

Chapter 6: Session 20

Time series approach and the steps

Time series analysis is the use of statistical methods to analyze time series data and extract meaningful statistics and characteristics of the data.

Time series analysis is the collection of data at specific intervals over a period of time, with the purpose of identifying trends, cycles, and seasonal variances to aid in the forecasting of a future event. Data is any observed outcome that's measurable. Unlike in statistical sampling, in time series analysis, data must be measured over time at consistent intervals to identify patterns that form trends, cycles, and seasonal variances. Measurements at random intervals lose the ability to predict future events.

There are two main goals of time series analysis:
(a) identifying the nature of the phenomenon represented by the sequence of observations,
(b) forecasting (predicting future values of the time series variable).
Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once the pattern is established, we can interpret and integrate it with other data (i.e., use it in our theory of the investigated phenomenon, e.g., seasonal commodity prices). Regardless of the depth of our understanding and the validity of our interpretation (theory) of the phenomenon, we can extrapolate the identified pattern to predict future events.

Problem Statement
There is a company X which has been keeping a record of monthly sales of shampoo for the past 3 years. Company X wants to forecast the sale of the shampoo for the next 4 months so that the demand and supply gap can be managed by the organisation. Our main job here is to simply predict the sales of the shampoo for the next 4 months.

Dataset comprises of only two columns. One is the Date of the month and other is the sale of the shampoo in that month.

Stages in Time Series Forecasting
Solving a time series problem is a little different as compared to a regular modelling task. A simple/basic journey of solving a time series problem can be demonstrated through the following processes. We will understand about tasks which one needs to perform in every stage.

Steps are –

Visualising time series
In this step, we try to visualise the series. We try to identify all the underlying patterns related to the series like trend and seasonality. You can say that this is more a type of an exploratory analysis of time series data.

Stationarising time series

A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary (i.e., "stationarised") through the use of mathematical transformations. A stationarised series is relatively easy to predict: you simply predict that its statistical properties will be the same in the future as they have been in the past! Another reason for trying to stationarise a time series is to be able to obtain meaningful sample statistics such as means, variances, and correlations with other variables. Such statistics are useful as descriptors of future behaviour only if the series is stationary.
For example, if the series is consistently increasing over time, the sample mean and variance will grow with the size of the sample, and they will always underestimate the mean and variance in future periods. And if the mean and variance of a series are not well-defined, then neither are its correlations with other variables

Finding the best parameters for our model
We need to find optimal parameters for forecasting models one's we have a stationary series. Hence, this stage is more about plotting above two graphs and extracting optimal model parameters based on them.

Fitting model
Once we have our optimal model parameters, we can fit an ARIMA model to learn the pattern of the series. Always remember that time series algorithms work on stationary data only hence making a series stationary is an important aspect

Predictions
After fitting our model, we will be predicting the future in this stage. We will find out the sales of the shampoo for the next 4 months.