Online Shopper's Intention : Implementation

Design and Implementation of Neural network models (Perceptron , SVM , LVQ and SOM) on Online Shoppers Purchasing Intention Dataset.

Anshal Singh

Lovely Professional University , Phagwara , Punjab

Abstract

In this project we have to use the dataset and try to implement certain models on the Dataset namely Perceptron , SVM (Support Vector Machines) , LVQ (Learning Vector Quantization) and SOM (Self Organizing Maps) .

Introduction

The data set was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

"Administrative", "Administrative Duration", "Informational", "Informational Duration" and "Product Related" represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories.

"Page Value" feature represent the metrics measured by "Google Analytics" for each page in the e-commerce site.

"Special Day" represents the section where we take into consideration the no of people visiting on different days throughout the year which also include all the special days like "Valentine's Day , "Mother's Day" , "Diwali" and "Christmas" etc.

So we are going to try to implement all the given models to the univariate dataset that we have extracted .

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.

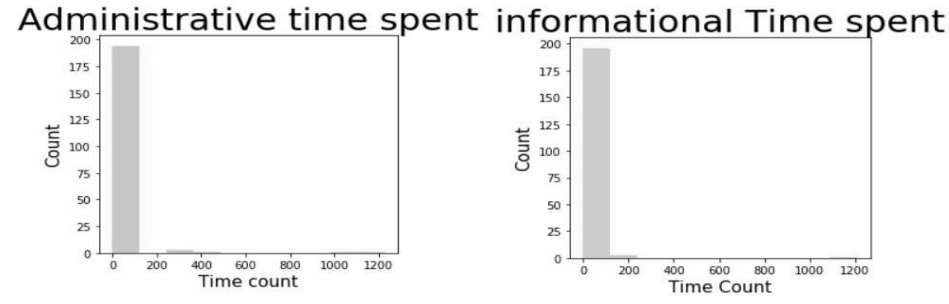
We need Data Pre-processing for achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set.

To start off everything we first read the data using the pandas library of Python .

After that we will import some useful libraries that we require for the implementation and the rest we will import as we go along the way.

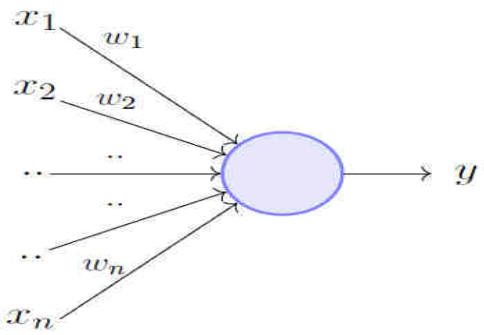
First we get some basic idea about the dataset we are dealing with such as the no of rows and columns in the dataset , count of all the values in the dataset , mean of all the values , standard deviation on the dataset and at last minimum and the maximum values of the dataset respectively.

We can also visualize some of the data.



Secondly , we extract the data from the csv file and sort it into the arrays we need for the models and so we proceed for the model implementation.

Perceptron was conceptualized by Frank Rosenblatt in the year 1957 and is the most primitive form of artificial neural networks.



$$y = 1 \quad if \sum_{i=1}^{n} w_i * x_i \geq \theta$$

$$= 0 \quad if \sum_{i=1}^{n} w_i * x_i < \theta$$

Rewriting the above,

$$y = 1 \quad if \sum_{i=1}^{n} w_i * x_i - \theta \geq 0$$

$$= 0 \quad if \sum_{i=1}^{n} w_i * x_i - \theta < 0$$

The perceptron model is a more general computational model than McCulloch-Pitts neuron. It takes an input, aggregates it (weighted sum) and returns 1 only if the aggregated sum is more than some threshold else returns 0.

In a perceptron model the learning parameters are as follows:

- Weights of input parameters
- Bias
- Threshold value

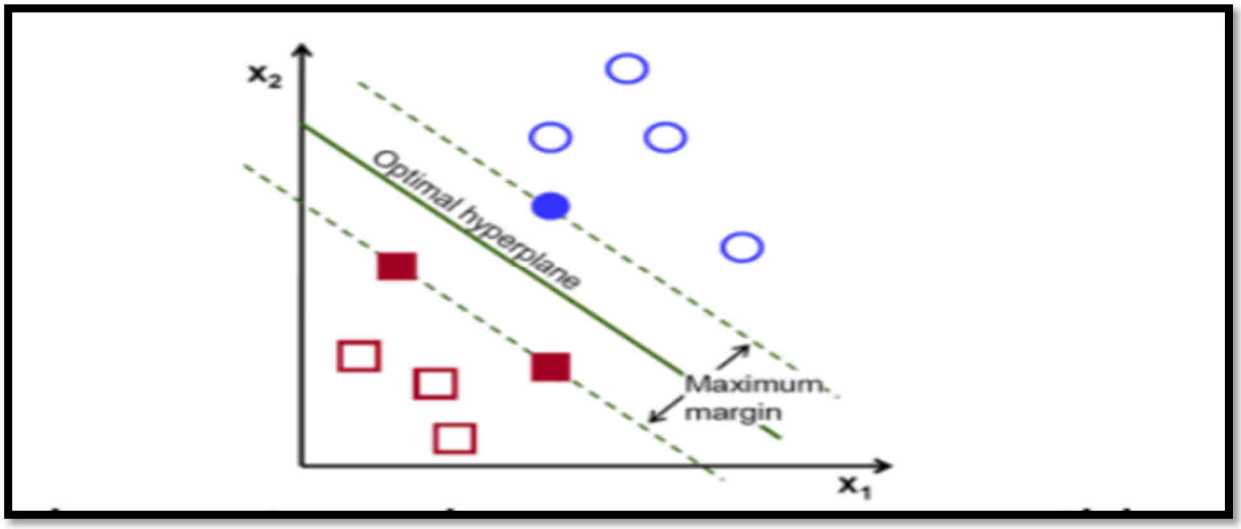
During the training phase , the weights are initialized randomly and with each data , the error is calculated and the weights are adjusted accordingly to reduce the error.

Actual	Output	Error = Actual - Output	Weight Adjustment
0	0	0	No Adjustment
1	0	1	Output is less than actual so increase the weight
0	1	-1	Output is more than actual so decrease the weight
1	1	0	No Adjustment

In the perceptron model we take the learning rate as 0.01 and number of iterations as 50 or 30 and after it is done implementing we get an accuracy score of 0.72 which also could vary if we decided to use a different dataset and a different part of the data.

Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms which are used both for classification and regression. But generally, they are used in classification problems. In 1960s, SVMs were first introduced but later they got refined in 1990. SVMs have their unique way of implementation as compared to other machine learning algorithms. Lately, they are extremely popular because of their ability to handle multiple continuous and categorical variables.

An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).

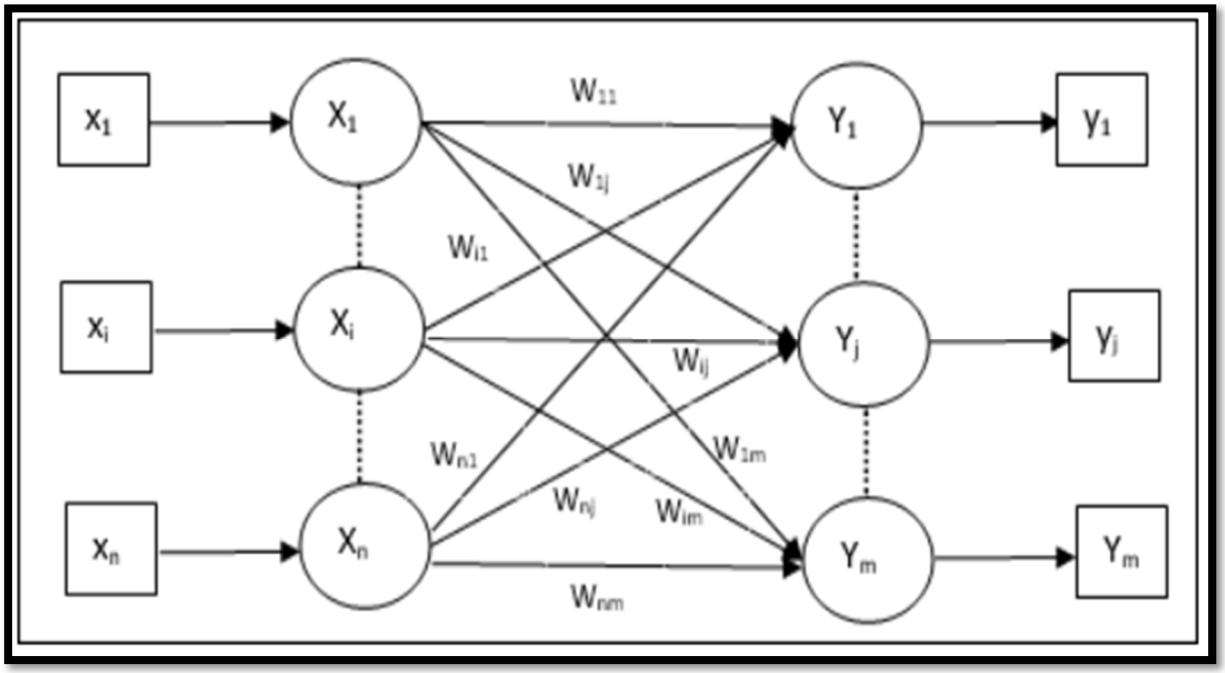


Support vectors are the datapoints closest to the hyperplane . As we can see in the above diagram, Hyperplane is a decision plane or space which is divided between a set of objects having different classes. Margin may be defined as the gap between two lines on the closet data points of different classes.

In the SVM model we take the kernel as poly and random state as 0 and after it is done implementing we get an accuracy score of 0.75 which also could vary if we decided to use a different dataset and a different part of the data.

It is a process of classifying the patterns where each output unit represents a class. As it uses supervised learning, the network will be given a set of training patterns with known classification along with an initial distribution of the output class. After completing the training process, LVQ will classify an input vector by assigning it to the same class as that of the output unit.

Architecture:



X= Training Vector (X1 , X2 , …….. Xn)

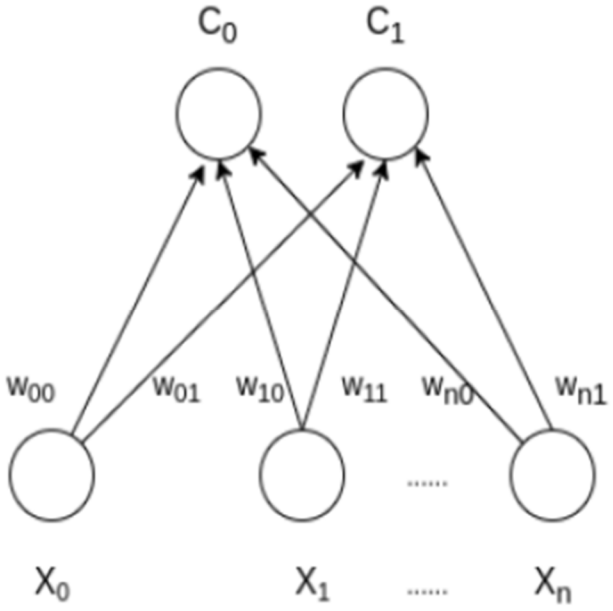
T= Class for Training Vector X

Wj= weight of Jth output unit

Cj= class associated with the Jth output unit

In the LVQ model we take the learning rate as 0.6 and epoch as 1 and random weight array and after it is done implementing we get an accuracy score of 0.88 which also could vary if we decided to use a different dataset and a different part of the data.

Self Organizing Map (or Kohonen Map or SOM) is a type of Artificial Neural Network which is also inspired by biological models of neural systems form the 1970's. It follows an unsupervised learning approach and trained its network through a competitive learning algorithm. SOM is used for clustering and mapping (or dimensionality reduction) techniques to map multidimensional data onto lower-dimensional which allows people to reduce complex problems for easy interpretation.



Steps involved are :

- Weight initialization
- For 1 to N number of epochs
- Select a training example
- Compute the winning vector
- Update the winning vector
- Repeat steps 3, 4, 5 for all training examples.
- Clustering the test sample

In the SOM model we take the learning rate as 0.6 and epoch as 1 and random weight array and after it is done implementing we get an accuracy score of 0.85 which also could vary if we decided to use a different dataset and a different part of the data.

GITHUB LINK:

https://github.com/Anshal55/Int246

Conclusion

Upon implementing all the given models on the provided dataset we can conclude that the accuracies as result obtained by Perceptron Model , SVM , LVQ and SOM are 0.72 , 0.75 , 0.88 and 0.85 which are subject to change if changes are made to the process or dataset.

References:

Sheng-Hsien Lee Department of Business Administration, School of Management, Yu Da University, No. 168, Hsueh-fu Rd., Miao-li County, 36143, Taiwan. E-mail: stephen@ydu.edu.tw.

Theodoras Evgeniou , Massimiliano Pontil , Support Vector Machines :Theory and Applications.

C .V Karthik

Explorable , https://explorable.com/example-of-a-research-paper .

Atsushi Sato & Keiji Yamada Information Technology Research Laboratories, NEC Corporation.

Umut Asan , Secil Ercan , Am Introduction to Self Organizing Maps.

Pooja Rana , Google Classroom Material.