



Supervised Learning Algorithm in Classification : A Comparison of SVM and KNN

¹Dr. V. Maniraj

Associate Professor, Research Supervisor, Head of the Department,
Department of Computer Science, A.V.V.M Sri Pushpam College(Autonomous), Poondi,
Thanjavur(Dt), Affiliated To Bharathidasam University, Thiruchirappalli, Tamilnadu,

²G.Varja M.Sc Computer Science

Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi,
Thanjavur(Dt), Affiliated To Bharathidasan University, Thiruchirappalli, Tamilnadu,

ABSTRACT

Machine Learning is one of the fastest growing areas of computer science, with far-reaching applications. In these paper, Supervised Learning is one of the tasks most frequently carried out by the Intelligent systems. Supervised Learning Algorithms as well as determines the most efficient classification algorithm based on the data set, the number of enhances and features. The main goal of this paper the overall concept of SVM and KNN algorithms using Supervised Learning Techniques.

Keywords: Supervised Learning, Distance, Algorithms, classification, Prediction.

INTRODUCTION

Machine learning is a most important part of today world. It is a subset of Artificial Intelligence(AI). There are many applications of Machine Learning, in computer technology like healthcare, Finance, Retail, Entertainment, Transportation etc... Machine Learning has become one of the mainways of Information Technology and with that, central, usually hidden, part of our life. currently, Machine Learning Algorithms have been developed, updated and improved and ability to automatically apply a variety of complex mathematical calculation to a big data, which calculates the result much faster.

The basic process of machine learning using Five methods:

- I. **Input data:** The past or previous data information collect and learn from machines like text,images,categorical, audio files.
- II. **Algorithms:** The input data represented a particular problems occur to machine learn from data into a set of mathematical rules. They can be supervised, unsupervised, reinforcement.
- III. **Trained data:** The machine learned from previous data and feed data machine learning models and adjust its parameters to make accurate decisions.
- IV. **Prediction:** Once a model has been trained; it can make predictions based on new data. V. **Output:** Finally, the machine identify a data and executed the correct data.



SUPERVISED LEARNING

In Supervised learning is a type of machine learning. These model is trained on labeled data. The task of machine learning algorithms to find patterns and construct the mathematical models. Supervised Algorithms are predetermined. The motivation of supervised learning, learned to input data and corresponding of the output, so it can make accurate predictions of new data. Supervised Learning generates a function that Maps inputs to desired outputs.

There are two main models of supervised learning:

- i. Classification models
- ii. Regression models

Classification: Classification or classifiers is a method of Machine learning, It referring to a problem classify into predict to the label. It can be carried out structured or unstructured data. Two classifiers:

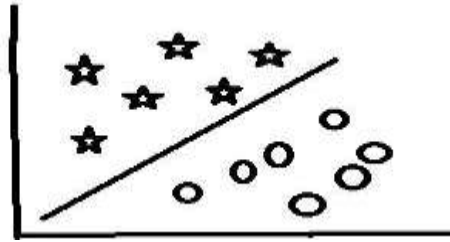
Binary classification: It consists of classification is divided into two classes, 'Yes' or 'No'. For example, 'spam' or 'not spam' email service provider.

Multiclass classification: The multiclass classification does not have the principle of normal and abnormal outcomes, unlike binary classification tasks.

Regression: If the problem can be overloaded, many parameters find the relationship between variable so that prediction called regression. It is robust to outliers. For Example, Company Profits, Whether Conditions, etc...

SUPPORT VECTOR MACHINE(SVM):

Support Vector Machine(SVM) is one of the most important algorithms for supervised learning. It is used for both classification and regression models. SVM chooses the extreme points/vectors that helps in creating the hyperplane. These extreme cases are called as Support Vectors. There are many advantages are High Dimensional Performance, non linear capability, Memory efficiency.



The Simple Diagram for Support Vector Machine Works to identified a hyperplane (line) which maximises the separation between the ‘circle’ and ‘star’.

There are two types:

- I. Linear SVM: It consists of a dataset can be classified into two classes by using a straight line to single and separable data.
- II. Non-linear SVM: It consists of a dataset cannot be classified into using by straight line. The Non-linear SVM easily used for 2-Dimensional changed in 3-Dimensional that is called best hyperplane.

K-Nearest Neighbor(KNN):

K-Nearest Neighbor or KNN is one type of supervised Learning Algorithm. To understand the algorithm a simple dataset in the machine learning technology. KNN algorithm is mostly used for classification problems. KNN is a non-parametric algorithm. It is also called a lazy learner algorithm because it is not identify the training set immediately. The advantages are simple, robust, more effective. KNN is used for distance metrics:

- I. Euclidean Distance
- II. Manhattan Distance
- III. Minkowskhi Distance

For Example, KNN Algorithm using find out the Customer Review?

KNN is a labeled data so it is understanding to very easily. It is very simple customer dataset. Each of the customer rate on the product details review:

Customer having Great quality product is classified as ‘Excellent’

Customer having Good Quality Product is classified as ‘Nice or Good’

Customer having Poor Quality Product is classified as ‘Terrible’

CUSTOMER NAME	AGE	GENDER	REVIEW
John	21	M	Terrible



Jack	30	M	Nice
Aadvika	15	F	Excellent
Sara	22	F	Nice
Smith	26	M	Poor
Mark	18	M	Excellent
Manasa	24	F	Good
Nitara	35	F	Terrible
Ajay	41	M	Good
Sachin	23	M	?

Already, Machine trained the data and after classification models using predict the data. Then, find the K value factor. Below the datasets are 9 data points. So, the k value factor is 3. $d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}$ to find the distance between any two points.

Find the Distance in Customer Review.

CUSTOMER NAME	Distance
John	29.7
Jack	36.4
Aadvika	26.9
Sara	31.1
Smith	33.2
Mark	27.8
Manasa	32.5
Nitara	41.0
Ajay	45.6
Sachin	?

So, K factor is 3 for Sachin. The KNN algorithm using closest value is 26.9,27.8,29.7.

CUSTOMER NAME	DISTANCE	REVIEW
Aadvika	26.9	Excellent
Mark	27.8	Excellent
John	29.7	Terrible

Finally, Get the Output for Majority Voting is Excellent. So, Sachin rating the Review is Excellent.

CONCLUSION:

In these model comparative performances of Supervised Learning different between Support



Vector Machine and K-Nearest Neighbor Algorithms. An extensive or effective list of information and statistical and mathematical measures for a dataset. After a better understanding of the strengths and limitations of each methods, to solve the problem should be investigated. The KNN algorithms that high level of correlation among to the output model more efficient. By comparing this two algorithms that the K-Nearest Neighbors(KNN) is simple and easier to understand than Support Vector Machine(SVM).

REFERENCE:

1. Padraig Cuningham; Sarah Jane Delany (2007) k-Nearest Neighbour classifiers.
2. Tom Mitchell, McGraw Hill (2015) Machine Learning.
3. Sisodia D, Sisodia DS Prediction of Diabetes using classification algorithms. Procedia Comput Sci. 2018;132:1578-85.
4. Shai Shalev-Shwartz and Shai Ben-David (2014). Understanding Machine Learning From Theory To Algorithms.
5. Amanpreet Singh, Narina Thakur, Aakanksha Sharma (2016). A review of supervised machine learning algorithms. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on.
6. Boukerche A, Wang J. Machine learning-based traffic prediction models for intelligent transportation systems. ComputNetw. 2020;181.
7. Essien A, Ptrounias I, Sampaio P, Sampaio S. A deep-learning model for urban traffic flow prediction with traffic events mined from twitter. In: World Wide Web, 2020: 1-24.