International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1765-1768 www.ijaem.net ISSN: 2395-5252

Machine Learning - A Review Paper

Rajveer Kaur, Suman Lata, Jaswinder Singh Brar

Date of Submission: 20-08-2022 Date of Acceptance: 31-08-2022

ABSTRACT: Everything has become sentient and capable of doing tasks like a human in this age of computer science. Numerous tools, strategies, and methods are suggested for that goal. A model for supervised learning that is used in statistics and computer science, support vector machines are used to analyse data and spot trends. Regression analysis and classification are the two main applications of SVM. Similarly, the k-nearest neighbour algorithm is a classification technique that uses training data to categorise data.
examples. In this study, we categorise the data and provide predictions (identify hidden patterns) for the target using the SVM and KNN algorithms. Here, we employ data mining, which is used to categorise text analysis in the future, to categorise and uncover the data patterns to forecast future disease using medical patients' nominal data..
Keywords: SVM, KNN, Patterns, Analysis, Classification.

I. INTRODUCTION:

The discipline of computer science known as knowledge discovery in databases is a relatively new one. The actual objective is to automatically or semi-automatically analyse vast amounts of data to uncover previously undetected patterns, such as clusters of data records (cluster analysis), atypical records (anomaly identification), and dependencies (association rule mining). Data mining allows for the finding of information from data produced by other domains, which is how this knowledge was developed. Data mining's fundamental objective is to extract knowledge from data and present it in a format that is understandable to humans [7]. [5]. Knowledge was necessary for numerous fields, including business, education, and industry, in order for them to advance and remain competitive.
knowledge for growth and with stand in the point where they reached this kind of data is helpful. The process of data mining consists of three stages:

SVM Static Vector Machine (SVM) Support Vector Machines are the most frequent and well-liked approach for machine learning tasks in classification and regression [16]. In this method, a set of training examples is provided, and each example bears a label designating which of the two categories it falls into. Then, a model that can predict whether a new example falls into one category or the other is built using the SVM algorithm.

Advantages of (SVM):

- • SVM typically does not experience overfitting and performs well when there is a distinct evidence of class separation. SVM works well in terms of memory performance and can be utilised when the total number of samples is fewer than the total number of dimensions.
- • SVM functions well and generalises to out-of-sample data. This is because SVM exhibits speedy performance on out-of-generalization sample data, as the kernel function is assessed and executed for each and every support vector in SVM for the classification of a single sample.
- • The ability to handle high-dimensional data is another significant benefit of the SVM algorithm, and this demonstrates

Disadvantages of (SVM):
: • For huge data sets, the support vector machine approach is unacceptable.
• It performs poorly when the target classes overlap and the data set contains more noise.
• The support vector machine will perform poorly when the number of attributes for each data point exceeds the number of training data specimens.
• There is no probabilistic explanation for the classification because the support vector classifier places data points above and below the classifying hyperplane.

Applications of support vector machine
1. Face observation - This technique is used to find faces in accordance with the model and classifier.
2. Text and hypertext arrangement - In this, the categorization technique is utilised to locate crucial information, or more precisely, information that is necessary for text arrangement.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1765-1768 www.ijaem.net ISSN: 2395-5252

3. Grouping of representations - This technique is also employed in grouping or you can say by comparing the item of information and taking an appropriate action.
4. Bioinformatics - It is also utilised in medical science for things like DNA research, lab work, and other things.
5. Handwriting recollection - This is used to recognise handwriting.
1. Protein fold and remote homology spotting - This technique is used to identify protein folds and to classify proteins according to their structural and functional properties based on the amino acid sequences they contain. It is one of the bioinformatics issues.
GPC, or generalised predictive control,
As the plants linear model is described, it is also employed for generalised predictive control (GPC), which depends on predictive control utilising a multilayer feed-forward network.

KNN The k-nearest neighbour algorithm: is a method for categorising objects based on closest training samples in the problem space in pattern recognition or classification. KNN is a form of instance-based learning, or lazy learning, in which all computation is postponed until after classification and the function is only locally approximated [3]. One of the simplest machine learning algorithms is the k-nearest neighbour algorithm, which classifies an item by a majority vote of its neighbours and assigns it to the class that is most prevalent among its k nearest neighbours (k is a positive integer, typically small). If k is equal to 1, the item is just put in the class of its closest neighbour.
To estimate continuous variables, the k-NN technique can also be modified. One such method uses the k-nearest multivariate neighbours' inverse distance weighted average. The way this algorithm works is as follows.
a) Calculate the Euclidean or Mahalanobis distance between the target plot and the sampling plots.
b) Sort samples according to determined distances.
Choose the k nearest neighbours based on RMSE calculated using the cross validation technique.
d) Create a weighted average based on inverse distance using the k-nearest multivariate neighbours.
Order samples while taking calculated distances into consideration. c) Based on the cross validation technique's RMSE results, select the k nearest neighbours that are heuristically ideal. d) Use the k-nearest multivariate neighbours to calculate an inverse distance weighted average.

Advantages of KNN Algorithm:
It is easy to put into action.
It can be more successful if the training data is vast and is robust to noisy training data.

Disadvantages of KNN Algorithm:
- K's value must always be determined, which can occasionally be difficult.
- by figuring out the separation between each data point for each training sample.

Applications of KNN
1. Text mining The computation cost is high because of
2. Agriculture
3. Finance
4. Medical
5. Facial recognition
6. Recommendation systems (Amazon, Hulu, Netflix, etc)

Working of Support Vector Algorithm:
SVM categorises data points even when they are not otherwise linearly separable by translating the data to a high-dimensional feature space. Once a separator between the categories is identified, the data are converted to enable the hyperplane representation of the separator.
Support Vector machines have a separating hyperplane called a "Support Vector Machine" and a few specific data points that we refer to as "Support Vectors." Therefore, SVM is essentially a frontier that best separates the classes. The data

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1765-1768 www.ijaem.net ISSN: 2395-5252

points in our data set that are closest to the dividing hyperplane are known as support vectors because their removal would change the position of the dividing hyperplane. As we can see, there are numerous hyperplanes..

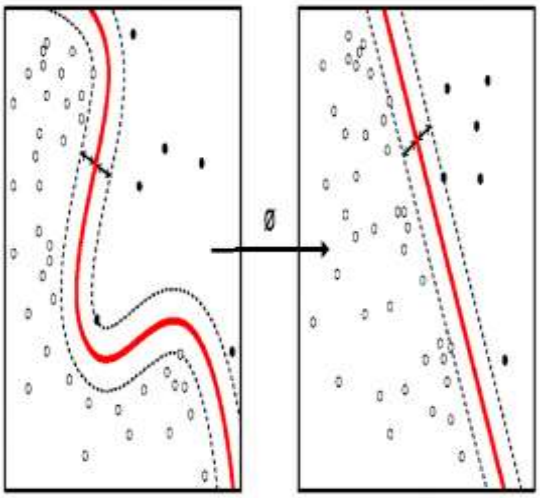


Figure 1 flow chart SVM

The mathematical function used for the transformation is known as the kernel function. SVM supports the following kernel types:
Linear
Polynomial
Radial basis function (RBF)
Sigmoid

How does K-NN work?

The foIllowing algorithm can be used to describe how K-NN works: Decide on the number of neighbours (K) (K).

• It converts any real value between 0 and 1 into another value.

Finding the Euclidean distance between K neighbours is step two.

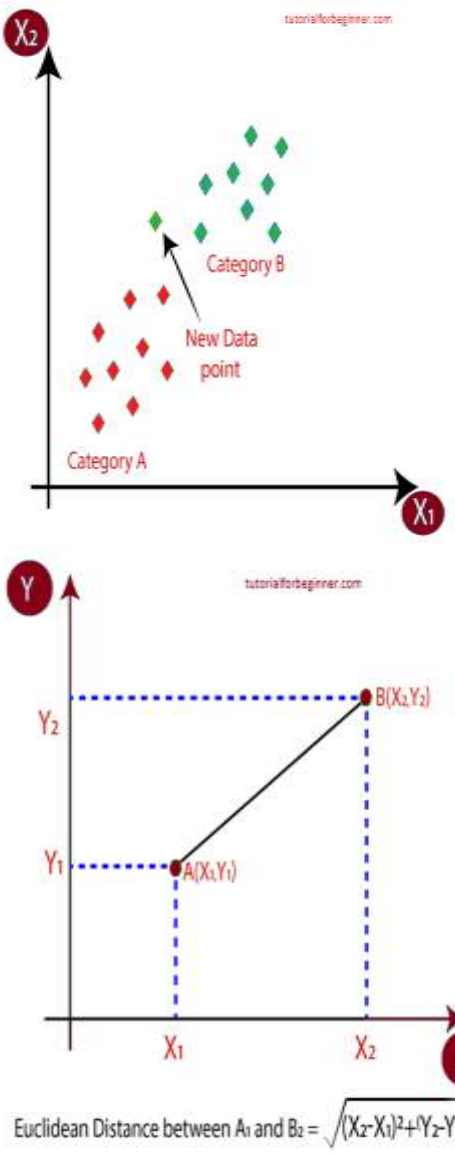
Step 3: Locate the K nearest neighbours using the Euclidean distance that was calculated.

Step 4: Among these k neighbours, count the number of data points in each category.

The fifth step is to assign the fresh data points to the category with the most neighbours.

• Step 6: Our model is finished.

Let's say we have a new data point that needs to be placed in the appropriate category. Consider the following illustration:

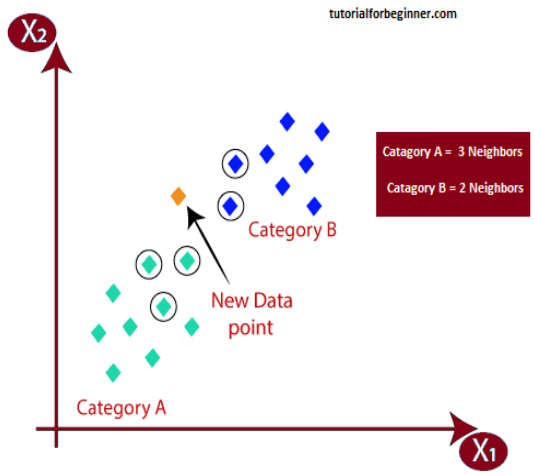


First, we'll decide on the number of neighbors, thus we'll go with k=5.

Next, a calculation of the Euclidean distance between the data points will be performed. In geometry class, we studied about the distance between two points known as the Euclidean distance. The following formula can be used to compute it:

hree people were closest to each other in category A and two people were closest to each other in category B. Take the example below into consideration.: The Euclidean distance was calculated to determine who was closest to whom. T

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 8 Aug. 2022, pp: 1765-1768 www.ijaem.net ISSN: 2395-5252



As can be seen, the three closest neighbors are all from category A, hence this new data point must also be from that category.

II. CONCLUSION:

After putting it into practise, we discovered that K-NN is a very good classifier, but when we use it to classify textual data (also known as nominal data), its performance parameters change depending on the size of the dataset. K-NN performs poorly as the size of the data set grows; it fits tiny data sets the best. SVM is a sophisticated classifier, and we use a leaner kernel in this case. We discovered that the number of training cycles has a far greater influence on all performance measures than the size of the dataset. The best classifier for our text mining is this one (contain mining). SVM will be used in the future for text analysis and data analysis on the web. Their website application includes

REFERENCE:

[1]. B. Silver, "Netman: A learning network traffic controller," in Proc
[2]. ide.geeksforgeeks.org.
[3]. Arbor Networks – Annual report(2015), http://www.arbornetworks.com/resources/annual-security-report [accessed on Januar
[4]. https.//data-flair.training
[5]. Lata, S., and R. Kumar. "A Hybrid Approach for ECG Signal Analysis." Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2018, 2018, doi:10.1109/ICACCCN.2018.8748858.
[6]. Lata, Suman, and Rakesh Kumar. "Disease Classification Using ECG Signals Based on R-Peak Analysis with ABC and ANN." International Journal of Electronics, Communications, and Measurement Engineering, vol. 8, no. 2, July 2019, pp. 67–86, doi:10.4018/IJECME.2019070105.
[7]. Lata, Suman, and Dheerendra Singh. "A Hybrid Approach for Cloud Load Balancing." In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 548-552. IEEE, 2022.