Cosmos Multidisciplinary Research E-Journal Online Available at www.cmrj.in
Recognized International Peer Reviewed Journal Impact Factor 4.94 ISSN No. 2456-1665

Kernel Methods in Machine Learning: A Case study

Mr. Dattatraya Bhikaji Raymal
Research Student, Department of Computer Science,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
dattatrayaraymal@gmail.com

Abstract

As neural networks started to gain some respect among researchers in the 1990s, thanks to this first success, a new approach to machine learning rose to fame and quickly sent neural nets back to oblivion: kernel methods. *Kernel methods* are a group of classification algorithms, the best known of which is the *support vector machine* (SVM). The modern formulation of an SVM was developed by Vladimir Vapnik and Corinna Cortes in the early 1990s at Bell Labs and published in 1995,2 although an older linear formulation was published by Vapnik and Alexey Chervonenkis as early as 1963.3 SVMs aim at solving classification problems by finding good *decision boundaries* (see figure 1.10) between two sets of points belonging to two different categories. A decision boundary can be thought of as a line or surface separating your training data into two spaces corresponding to two categories. To classify new data points, you just need to check which side of the decision boundary they fall on. SVMs proceed to find these boundaries in two steps:

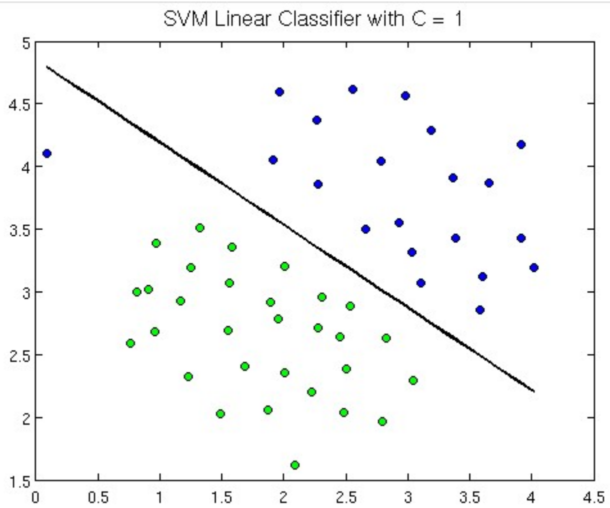
The data is mapped to a new high-dimensional representation where the decision boundary can be expressed as a hyperplane (if the data was twodimensional, as in figure 1.10, a hyperplane would be a straight line). A good decision boundary (a separation hyperplane) is computed by trying to maximize the distance between the hyperplane and the closest data points from each class, a step called *maximizing the margin*. This allows the boundary to generalize well to new samples outside of the training dataset.

Introduction:-

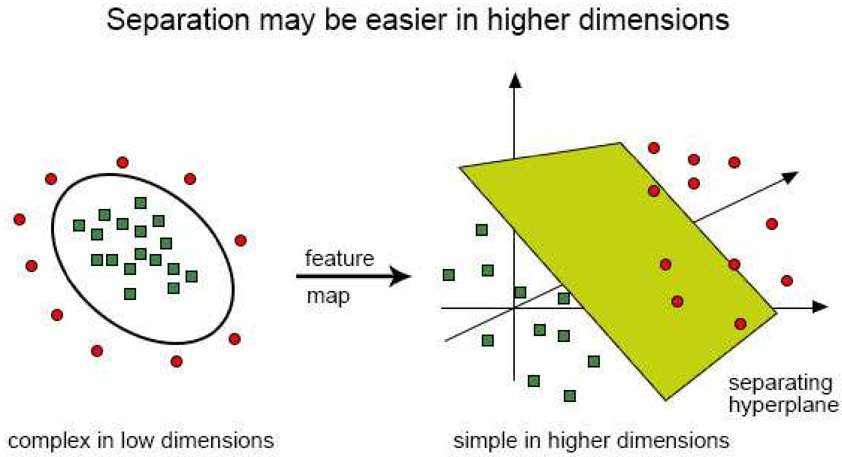
Kernels or kernel methods (also called Kernel functions) are sets of different types of algorithms that are being used for pattern analysis. They are used to solve a non-linear problem by using a linear classifier. Kernels Methods are employed in SVM (Support Vector Machines) which are used in classification and regression problems. The SVM uses what is called a "Kernel Trick" where the data is transformed and an optimal boundary is found for the possible outputs.

The Need for Kernel Method and its Working

Before we get into the working of the Kernel Methods, it is more important to understand support vector machines or the SVMs because kernels are implemented in SVM models. So, Support Vector Machines are supervised machine learning algorithms that are used in classification and regression problems such as classifying an apple to class fruit while classifying a Lion to the class animal. To demonstrate, below is what support vector machines look like:



Here we can see a hyperplane which is separating green dots from the blue ones. A hyperplane is one dimension less than the ambient plane. E.g. in the above figure, we have 2 dimension which represents the ambient space but the lone which divides or classifies the space is one dimension less than the ambient space and is called hyperplane.
But what if we have input like this:



It is very difficult to solve this classification using a linear classifier as there is no good linear line that should be able to classify the red and the green dots as the points are randomly distributed. Here comes the use of kernel function which takes the points to higher dimensions, solves the problem over there and returns the output. Think of this in this way, we can see that the green dots are enclosed in some perimeter area while the red one lies outside it, likewise, there could be other scenarios where green dots might be distributed in a trapezoid-shaped area.

So what we do is to convert the two-dimensional plane which was first classified by one-dimensional hyperplane ("or a straight line") to the three-dimensional area and here our classifier i.e. hyperplane will not be a straight line but a two-dimensional plane which will cut the area. In order to get a mathematical understanding of kernel, let us understand the Lili Jiang's equation of kernel which is:

$K(x, y)=<f(x), f(y)>$ where,
K is the kernel function,
X and Y are the dimensional inputs,
f is the map from n-dimensional to m-dimensional space and,
$< x, y >$ is the dot product.

Illustration with the help of an example.
Let us say that we have two points, x= (2, 3, 4) and y= (3, 4, 5)
As we have seen, K(x, y) = < f(x), f(y) >.
Let us first calculate < f(x), f(y) >
f(x)=(x1x1, x1x2, x1x3, x2x1, x2x2, x2x3, x3x1, x3x2, x3x3)
f(y)=(y1y1, y1y2, y1y3, y2y1, y2y2, y2y3, y3y1, y3y2, y3y3)
so,
f(2, 3, 4)=(4, 6, 8, 6, 9, 12, 8, 12, 16)and
f(3 ,4, 5)=(9, 12, 15, 12, 16, 20, 15, 20, 25)
so the dot product,
f (x). f (y) = f(2,3,4) . f(3,4,5)=
(36 + 72 + 120 + 72 +144 + 240 + 120 + 240 + 400)= 1444
And,
K(x, y) = (2*3 + 3*4 + 4*5) ^2=(6 + 12 + 20)^2=38*38=1444.
This as we find out, f(x).f(y) and K(x, y) give us the same result, but the former method required a lot of calculations(because of projecting 3 dimensions into 9 dimensions) while using the kernel, it was much easier.
Types of Kernel and methods in SVM
Let us see some of the kernel function or the types that are being used in SVM:

1. Liner Kernel

Let us say that we have two vectors with name x1 and Y1, then the linear kernel is defined by the dot product of these two vectors:
K(x1, x2) = x1 . x2

2. Polynomial Kernel

A polynomial kernel is defined by the following equation:
K(x1, x2) = (x1 . x2 + 1)d,
Where,
d is the degree of the polynomial and x1 and x2 are vectors

3. Gaussian Kernel

This kernel is an example of a radial basis function kernel. Below is the equation for this:

$$k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma\|\mathbf{x_i} - \mathbf{x_j}\|^2)$$

The given sigma plays a very important role in the performance of the Gaussian kernel and should neither be overestimated and nor be underestimated, it should be carefully tuned according to the problem.

4. Exponential Kernel

This is in close relation with the previous kernel i.e. the Gaussian kernel with the only difference is – the square of the norm is removed.
The function of the exponential function is:

$$k(x, y) = \exp\left(-\frac{||x - y||}{2\sigma^2}\right)$$

This is also a radial basis kernel function.

5. Laplacian Kernel

This type of kernel is less prone for changes and is totally equal to previously discussed exponential function kernel, the equation of Laplacian kernel is given as:

$$k(x, y) = \exp\left(-\frac{||x - y||}{\sigma}\right)$$

6. Hyperbolic or the Sigmoid Kernel

This kernel is used in neural network areas of machine learning. The activation function for the sigmoid kernel is the bipolar sigmoid function. The equation for the hyperbolic kernel function is:

$$k(x, y) = \tanh(\alpha x^T y + c)$$

This kernel is very much used and popular among support vector machines.

7. Anova radial basis kernel

This kernel is known to perform very well in multidimensional regression problems just like the Gaussian and Laplacian kernels. This also comes under the category of radial basis kernel. The equation for Anova kernel is :

$$k(x, y) = \sum_{k=1}^{n} \exp(-\sigma(x^k - y^k)^2)^d$$

There are a lot more types of Kernel Method and we have discussed the mostly used kernels. It purely depends on the type of problem which will decide the kernel function to be used.

Conclusion

In this section, we have seen the definition of the kernel and how it works. We tried to explain with the help of diagrams about the working of kernels. We have then tried to give a simple illustration using math about the kernel function. In the final part, we have seen different types of kernel functions that are widely used today.

References :-

- https://www.educba.com/kernel-methods/
- Deep learning with python by Francois chollet ,and the publication of Manning Publications Co. 20 Baldwin RoadPO Box 761Shelter Island, NY 11964 ISBN 9781617294433
- Vladimir Vapnik and Corinna Cortes, "Support-Vector Networks," Machine Learning 20, no. 3 (1995): 273–297.
- Vladimir Vapnik and Alexey Chervonenkis, "A Note on One Class of Perceptrons," Automation and Remote Control 25 (1964).
- Master In Data Science: A Practical Understanding Of Data Science: Decision Trees And Random Forests by Merna Laperouse (Author) ISBN-13: 979-8749666083 Publisher : Independently published (May 6, 2021)
- Big Data Analytics for Internet of Things by Tausifa Jan Saleem (Editor), Mohammad Ahsan Chishti (Editor) ISBN-13: 978-1119740759 , ISBN-10: 1119740754 Publisher : Wiley; 1st edition (April 20, 2021)
- Reverse Clustering: Formulation, Interpretation and Case Studies by Jan W. Owsiński (Author), Jarosław Stańczak (Author), Karol Opara (Author), Sławomir Zadrożny (Author), Janusz Kacprzyk (Author) ISBN-10 : 3030693589 Publisher : Springer; 1st ed. 2021 edition (March 4, 2021)
- Data Clustering: Algorithms and Applications by Charu C. Aggarwal (Editor), Chandan K. Reddy (Editor) ASIN : B00EYROAQU ISBN-13 978-1466558212 Publication date 21 August 2013 Publisher : Chapman and Hall/CRC; 1st edition (21 August 2013)