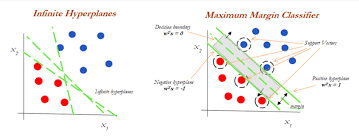
Support Vector Machine (SVM) is a powerful machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection tasks. SVMs can be used for a variety of tasks, such as text classification, image classification, spam detection, handwriting identification, gene expression analysis, face detection, and anomaly detection. **SVMs are adaptable and efficient in a variety of applications because they can manage high-dimensional data and nonlinear relationships**.

SVM algorithms are very effective as we try to find the maximum separating hyperplane between the different classes available in the target feature.

Ref link: <https://scikit-learn.org/stable/modules/svm.html>



**What is the Maximum Margin Separating Hyperplane**

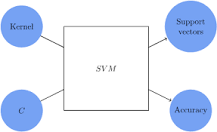
**Maximum Margin Separating Hyperplane** **(MMSH)**is a concept in machine learning that refers to a line (in 2D), a plane (in 3D), or a hyperplane (in higher dimensions) that **separates different classes of data points with the largest possible margin**. The **margin is the distance between the hyperplane and the closest data points from each clas**s, and the goal of MMSH is to find the hyperplane that maximizes this distance.



Code Ref: [SVM](http://localhost:8888/notebooks/Regularization%20techniques%2FSVM.ipynb)

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

How many support vectors are in SVM?



2

The SVM, in this example, uses 100% of the observations as support vectors. As it does so, it reaches maximum accuracy, whichever metric we want to use to assess it. The number of support vectors can however not be any lower than 2, and therefore this quantity does not appear problematic

What is Kernel Method?

A set of techniques known as kernel methods are used in machine learning to address classification, regression, and other prediction issues. They are built around the idea of kernels, which are functions that gauge how similar two data points are to one another in a high-dimensional feature space.

Kernel methods' fundamental premise is used to convert the input data into a high-dimensional feature space, which makes it simpler to distinguish between classes or generate predictions. Kernel methods employ a kernel function to implicitly map the data into the feature space, as opposed to manually computing the feature space.

Major Kernel Function in Support Vector Machine

In Support Vector Machines (SVMs), there are several types of kernel functions that can be used to map the input data into a higher-dimensional feature space. The choice of kernel function depends on the specific problem and the characteristics of the data.

**Here are some most commonly used kernel functions in SVMs:**

Linear Kernel

A linear kernel is a type of kernel function used in machine learning, including in SVMs (Support Vector Machines). It is the simplest and most commonly used kernel function, and it defines the dot product between the input vectors in the original feature space.

**The linear kernel can be defined as:**

1. K(x, y) = x .y

Where x and y are the input feature vectors. The dot product of the input vectors is a measure of their similarity or distance in the original feature space.

When using a linear kernel in an SVM, the decision boundary is a linear hyperplane that separates the different classes in the feature space. This linear boundary can be useful when the data is already separable by a linear decision boundary or when dealing with high-dimensional data, where the use of more complex kernel functions may lead to overfitting.

Polynomial Kernel

A particular kind of kernel function utilised in machine learning, such as in SVMs, is a polynomial kernel (Support Vector Machines). It is a nonlinear kernel function that employs polynomial functions to transfer the input data into a higher-dimensional feature space.

**One definition of the polynomial kernel is:**

Where x and y are the input feature vectors, c is a constant term, and d is the degree of the polynomial, K(x, y) = (x. y + c)d. The constant term is added to, and the dot product of the input vectors elevated to the degree of the polynomial.

The decision boundary of an SVM with a polynomial kernel might capture more intricate correlations between the input characteristics because it is a nonlinear hyperplane.

The degree of nonlinearity in the decision boundary is determined by the degree of the polynomial.

The polynomial kernel has the benefit of being able to detect both linear and nonlinear correlations in the data. It can be difficult to select the proper degree of the polynomial, though, as a larger degree can result in overfitting while a lower degree cannot adequately represent the underlying relationships in the data.

In general, the polynomial kernel is an effective tool for converting the input data into a higher-dimensional feature space in order to capture nonlinear correlations between the input characteristics.

Gaussian (RBF) Kernel

The Gaussian kernel, also known as the radial basis function (RBF) kernel, is a popular kernel function used in machine learning, particularly in SVMs (Support Vector Machines). It is a nonlinear kernel function that maps the input data into a higher-dimensional feature space using a Gaussian function.

**The Gaussian kernel can be defined as:**

1. K(x, y) = exp(-gamma \* ||x - y||^2)

Where x and y are the input feature vectors, gamma is a parameter that controls the width of the Gaussian function, and ||x - y||^2 is the squared Euclidean distance between the input vectors.

When using a Gaussian kernel in an SVM, the decision boundary is a nonlinear hyper plane that can capture complex nonlinear relationships between the input features. The width of the Gaussian function, controlled by the gamma parameter, determines the degree of nonlinearity in the decision boundary.

One advantage of the Gaussian kernel is its ability to capture complex relationships in the data without the need for explicit feature engineering. However, the choice of the gamma parameter can be challenging, as a smaller value may result in under fitting, while a larger value may result in over fitting.

Laplace Kernel

The Laplacian kernel, also known as the Laplace kernel or the exponential kernel, is a type of kernel function used in machine learning, including in SVMs (Support Vector Machines). It is a non-parametric kernel that can be used to measure the similarity or distance between two input feature vectors.

**The Laplacian kernel can be defined as:**

1. K(x, y) = exp(-gamma \* ||x - y||)

Where x and y are the input feature vectors, gamma is a parameter that controls the width of the Laplacian function, and ||x - y|| is the L1 norm or Manhattan distance between the input vectors.

When using a Laplacian kernel in an SVM, the decision boundary is a nonlinear hyperplane that can capture complex relationships between the input features. The width of the Laplacian function, controlled by the gamma parameter, determines the degree of nonlinearity in the decision boundary.

***Thank You***