Shrunga Ganesh - 23035807

Support Vector Machine (SVM)

Introduction to SVM:

A support vector machine (SVM) is a supervised machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space. SVM, which was first presented by Vladimir Vapnik and Corinna Cortes in 1992, seeks to determine the best hyperplane for separating data points of various classes with the greatest margin. The margin, which affects the decision boundary, is the separation between the hyperplane and the closest data points, also known as support vectors.

Key Milestones in SVM History

Year	Milestone
1960s-1970s	Development of VC Theory by Vapnik and Chervonenkis.
1992	Introduction of SVM by Vapnik and Cortes.
1995	Kernel trick formalized to handle non-linear data.
2001	LIBSVM released, facilitating practical SVM usage.
2000s	Widespread adoption for diverse ML tasks.
2010s	Adaptation for large-scale data and ensemble use.

What is SVM?

From "Support-Vector Networks" (1995) by Vapnik and Cortes - “The support-vector network is a new learning machine for two-group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high-dimensional feature space. In this feature space, a linear decision surface is constructed. Special properties of the decision surface ensure high generalization ability of the learning machine.”

Support Vector Machines (SVMs), sometimes referred to as support vector networks, are a family of incredibly potent models that may be applied to regression and classification problems. They employ method-based learning.

Finding the decision boundaries that divide observations with varying class memberships is their goal. To put it another way, a separating hyperplane officially defines SVM as a discriminative classifier.

Key Components of SVM

a) Support Vectors

- These are the data points that are closest to the hyperplane, or decision boundary.
- Because they affect the hyperplane's orientation and position, these points are crucial.

b) Hyperplane

- A line (in 2D) or a plane (in 3D) that divides the classes is called a hyperplane.
- For linear classification, this is a linear equation represented as wx+b=0.
- In higher dimensions, it generalizes to an nnn-dimensional space.

c) The margin

- The margin is the separation between each class's closest data points and the hyperplane.
- SVM seeks to optimize this margin while guaranteeing accurate training data categorization.

d) Kernel

- “A kernel function is a symmetric function that defines the similarity between two points in the input space. It corresponds to the inner product in some feature space and enables SVM to construct flexible, non-linear decision boundaries.” - From the book Pattern Recognition and Machine Learning (2006) by Christopher Bishop
- The radial basis function (RBF), sigmoid, linear, and polynomial are examples of common kernel functions.

e) Hard Margin

- A hard margin refers to the maximum-margin hyperplane that perfectly separates the data points of different classes without any misclassifications.

f) Soft Margin

- When data contains outliers or is not perfectly separable, SVM uses the soft margin technique. This method introduces a slack variable for each data point to allow some misclassifications while balancing between maximizing the margin and minimizing violations.

How does SVM works?

The SVM classifier, which is the main type of the SVM algorithm, is the best approach to comprehend it. The SVM classifier's concept is to create a hyper-lane in an N-dimensional space that separates data points into distinct classes. But since the hyperplane that offers the greatest margin between the two classes is taken into consideration, this hyper-pane is selected based on margin. Support Vectors are data points that are used to determine these margins. The data points that are close to the hyper-plane and aid in its orientation are known as support vectors.

The functioning of SVM classifier can be understood in the following ways-

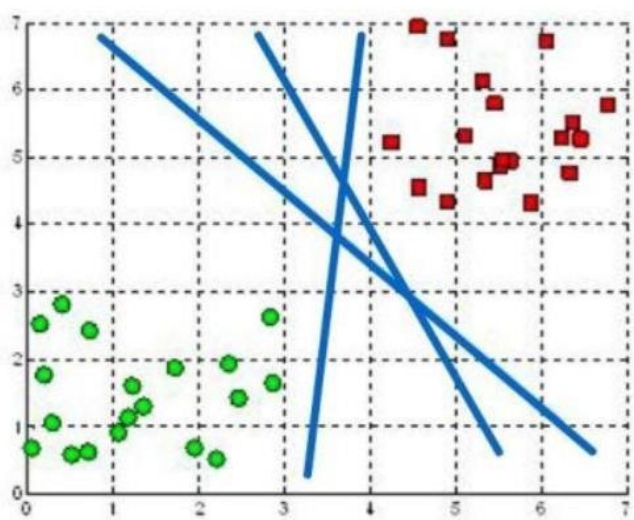
For binary classification tasks, the Support Vector Machine (SVM) algorithm gives the data points class labels. Positive classes are denoted by the symbol +1, whereas negative classes are denoted by the symbol −1. The mathematical formulation and decision-making process of SVM are based on these labels.

Data points that belong to the positive class (+1) are located on one side of the decision boundary. Data points in the negative class (−1) are located on the other side of the boundary.
By identifying which side of the hyperplane a data point is on, the SVM algorithm makes predictions about the classifications:

- If $w \cdot x + b > 0$, the data point is classified as +1
- If $w \cdot x + b < 0$, the data point is classified as -1

By correctly classifying as many points as feasible, SVM aims to place the hyperplane in a way that optimizes the margin between the nearest points of these two classes.

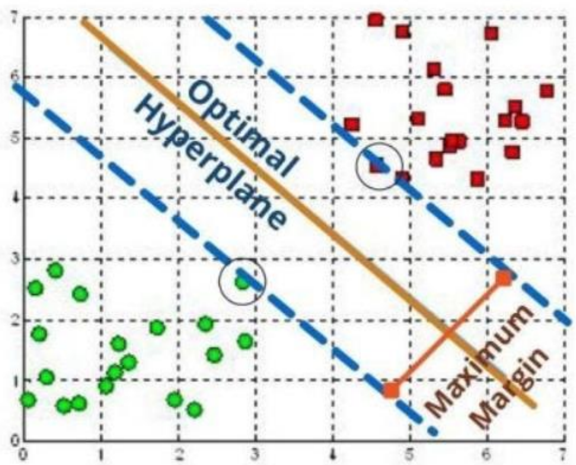
Possibilities



The business problem is transformed into a mathematical equation including unknowns by all machine learning techniques. The task is then transformed into an optimization problem in order to identify these unknowns. In the case of the SVM classifier, a loss function called the hinge loss function is utilized and adjusted to find the maximum margin because optimization issues always seek to maximize or minimize something when searching for and adjusting for the unknowns.

$$\min \frac{1}{2}||w||^2 + C \sum \text{hinge loss}$$

Optimal



The classifier's ability to predict class labels is assessed by the loss function, also known as the cost function, in SVM. The cost is 0 if there are no incorrectly classified points. On the other hand, a penalty (loss) is applied depending on the hinge loss when projections are off or points fall inside the margin.

The problem occurs when excessive margin maximization increases misclassification, particularly in data that is noisy or overlaps. A regularization parameter (C) is included to strike a balance between maximizing the margin and decreasing the classification error. This parameter ensures an ideal balance for improved generalization by regulating the relative priority of minimizing loss against maximizing the margin.

Weights are optimized by computing the gradients utilizing sophisticated calculus ideas, such as partial derivatives, as is the case with the majority of optimization issues.

The gradients are updated only by using the regularization parameter when there is no error in the classification while the loss function is also used when misclassification happens.

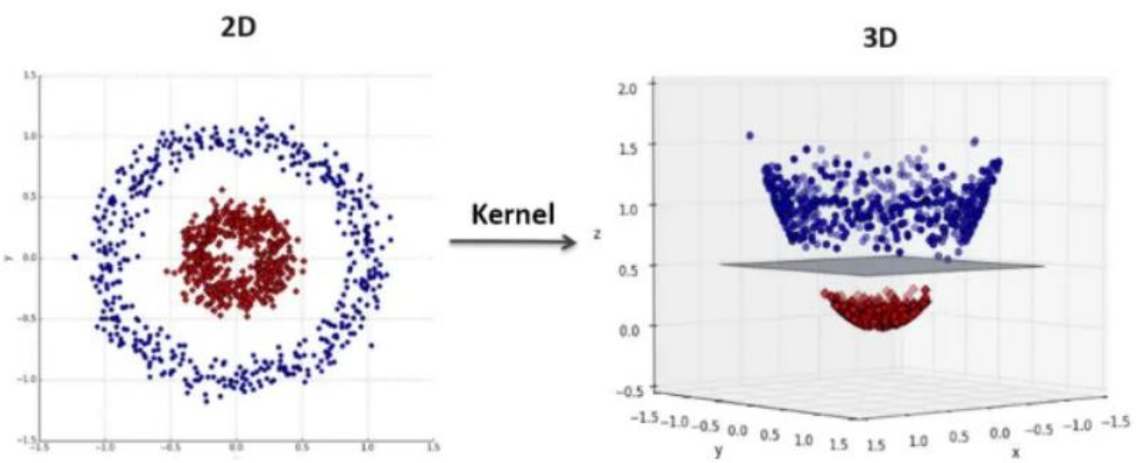
The gradients are updated only by using the regularization parameter when there is no error in the classification, while the loss function is also used when misclassification happens.

Kernels

One of the main reasons the Support Vector Machine (SVM) method is such a strong and adaptable machine learning tool is because it uses kernels. As previously mentioned, the SVM method automatically generates a linear hyperplane to divide classes. Although this is effective for linearly separable data, linear classifiers are insufficient for real-world situations since they are frequently non-linear.

The idea of kernel transformation comes in quite handy in this situation. SVM may implicitly transfer data from a low-dimensional input space to a higher-dimensional feature space thanks to a kernel function. Even complicated non-linear data can be linearly separated in this altered space. The kernel approach allows SVM to classify non-linear data while maintaining computing efficiency by skipping the explicit computation of the transformation.

Different linear and non-linear problems are solved using different kinds of kernels. For example, the RBF (Gaussian) kernel efficiently models non-linear relationships, the polynomial kernel captures more intricate interactions, and the linear kernel manages straightforward linear correlations. A crucial component of model design is selecting the appropriate kernel for a given issue, which necessitates meticulously adjusting kernel parameters as hyperparameters to maximize performance. Because of its adaptability, SVM can handle a variety of jobs and datasets.



Types of SVM

a) Linear SVM

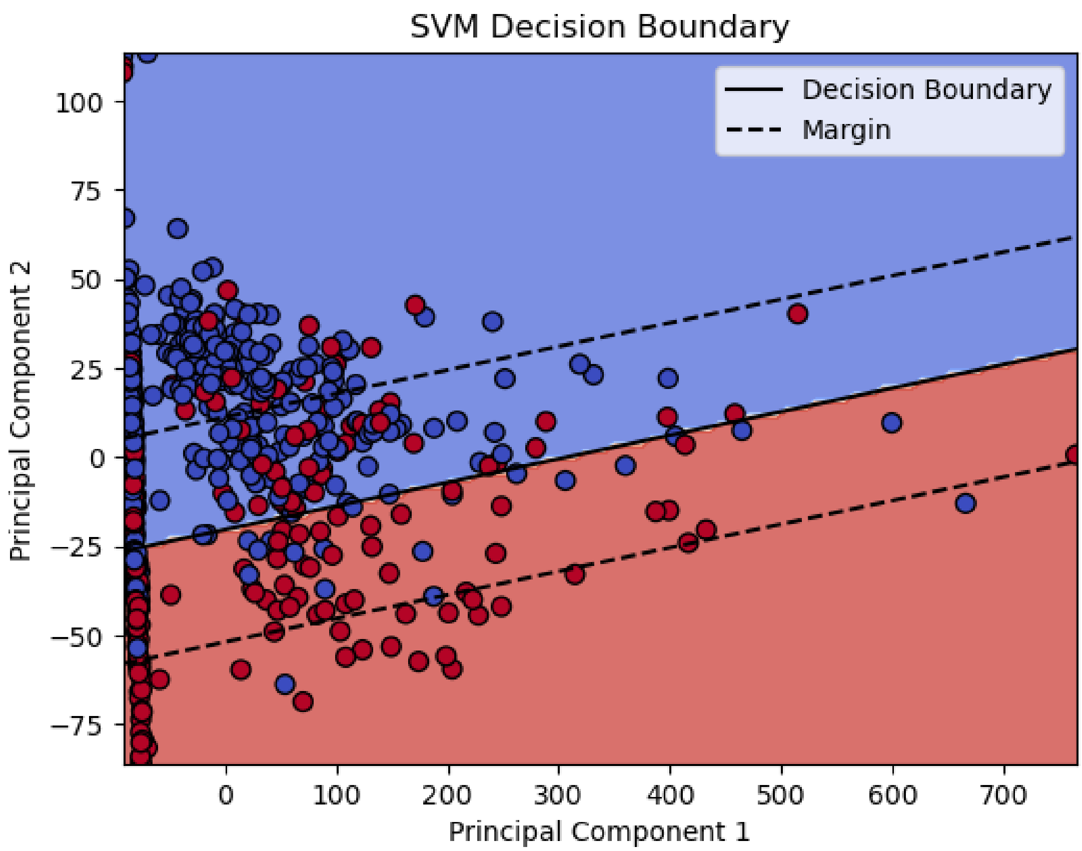
- Used when a flat hyperplane or a straight line may divide the data.
- As an illustration, separate dogs and cats according to height and weight.

b) Non-Linear SVM

- SVM employs a method known as the kernel trick to convert data into a higher-dimensional space where it can be linearly separated when it isn't.

SVM performed on the dataset

1. Import and load the dataset
2. Label Encoding the Target Variable: The LabelEncoder is applied to the outcome column in the dataset.The fit_transform function assigns numerical labels (e.g., "True" becomes 1 and "False" becomes 0).
3. Splitting Features (X) and Target Variable (Y): The features (independent variables) are stored in X by selecting all columns except the last one (dataset.iloc[:, :-1]).
4. The target variable (outcome, dependent variable) is stored in Y as the last column (dataset.iloc[:, -1]).
5. Train-Test Split: Train_test_split is used to separate the dataset into training and testing sections, with test_size=0.2 indicates that 80% of the data is used for training and 20% is set aside for testing.Random_state = 0. fixes the random seed, ensuring reproducibility.
6. Apply PCA : The algorithm reduces the features of the dataset to two principle components using principle Component Analysis (PCA). It converts both training and test data into 2D for viewing and computational efficiency, and it fits PCA on training data (X_train) to capture the most variation.
7. Plot to show the SVM decision boundary :
   The dimensions produced by Principal Component Analysis (PCA) are Principal Component 1 (PC1) and Principal Component 2 (PC2). The initial features (such as insulin, glucose, etc.) that capture the most variance in the dataset are represented by these components, which are linear combinations. In particular, the data's greatest variance is explained by PC1.The second largest variance is explained by PC2, which is orthogonal to PC1.
   The SVM classifier divides the two classes (blue and red points, which represent diabetic and non-diabetic receptively) by the solid black line, which serves as the decision boundary.
   The support vectors are the data points that are closest to the margins. The decision boundary's direction and location are directly impacted by these points.
   The margins, which are equally spaced from the decision boundary, are shown by the dashed black lines.
   These margins show the greatest separation between the decision boundary and each class's nearest data points, or support vectors. To improve generalization, SVM seeks to maximize this margin.



8. Accuracy score

```
]: from sklearn.metrics import accuracy_score
   ac=accuracy_score(Y_test, y_pred)*100
   ac

]: 81.81818181818183
```

The SVM classifier accurately predicted the outcomes (diabetes or non-diabetes) for roughly 81.81% of the test data, according to the output 81.81% accuracy. Eighty-one percent of the test samples were accurately identified as either diabetes or non-diabetic. The model was unable to assign those samples to the appropriate class, as evidenced by the remaining ~18.19% of predictions being wrong.

This accuracy is comparatively good for many machine learning applications, indicating that the SVM has successfully learned the data patterns.

In conclusion, the SVM classifier does a good job, but more advancements may be possible if we can figure out why misclassifications happen (e.g., feature overlap, outliers, or inadequate data).

Please find the code and data set use above in the github link : https://github.com/shrungalatha/Machine-Learning-and-Neural-Networks.git

Advantages of SVM

Support Vector Machines (SVM) are effective machine learning techniques for applications involving regression and classification. Here are a few main benefits:

a) **Effective in High-Dimensional Spaces**: SVMs are appropriate for tasks like text classification and bioinformatics, where high-dimensional datasets are frequently encountered, even when the number of features exceeds the number of data points.
b) **Memory Efficiency**: Support vector machines (SVMs) focus on a subset of training points called support vectors, which drastically lowers computational costs. This makes SVMs memory efficient.
c) **Versatility with Kernels:** SVM uses kernel functions to handle both linear and non-linear issues. SVMs can establish decision boundaries that more effectively divide complex datasets by employing kernels (such as polynomial or radial basis functions) to translate data into higher dimensions.
d) **Robust to Overfitting**: SVMs are dependable for many tasks, especially when dealing with noisy data, because they are less likely to overfit when properly regularized, especially in high-dimensional environments.
e) **Good Performance with Clear Margins of Separation**: SVM performs best when there is a distinct margin between classes, which results in good generalization and classification accuracy.

Because of these benefits, SVM is the preferred technique for a wide range of real-world machine learning applications, such as image identification and medical diagnosis.

Limitations of SVM

Despite its advantages, Support Vector Machines (SVM) come with several limitations that can impact their practical application:

a) **Computational Complexity:** SVMs can be costly to run, particularly when dealing with big datasets. It is less appropriate for real-time applications or for working with big volumes of data because the training time increases dramatically as the dataset size increases.
b) **Kernel selection and hyperparameter sensitivity**: SVM necessitates careful adjustment of parameters such as the kernel type, regularization parameter (C), and kernel parameters (e.g., gamma in the RBF kernel). Poor model performance might result from inappropriate decisions.
c) **Problem with Noisy Data**: SVMs may have trouble processing datasets with overlapping classes or a lot of noise. The model's decision boundary may be negatively impacted by outliers and noisy data points, which could result in overfitting or incorrect categorization.

d) **Poor Performance on Large Datasets**: Compared to alternative methods like decision trees or random forests, SVMs are inefficient due to their quadratic complexity in terms of both time and memory, which prevents them from scaling to large datasets.

Applications of SVM

Because Support Vector Machines (SVMs) are so good at addressing classification, regression, and outlier identification issues, they are used in many different industries. Here are some significant instances of real-world applications:

a) Text Classification

- Spam Filtering**:** SVM is frequently used in email systems to differentiate between authentic and junk communications. For instance, SVM and other machine learning techniques are used by Gmail to filter spam.
- Document Classification : Document classification is the process of grouping documents into predetermined groups, such as news stories classified as technology, sports, or politics
- SVM is used in the Reuters news categorization dataset to obtain excellent text classification accuracy.

b) Image Recognition:

- Face Detection : By recognizing facial features and differentiating them from backgrounds, SVMs are used to identify faces in photos
- Object Classification: Identifying items in photos, such as cars, animals, or pedestrians in autonomous driving systems, is known as object classification.
- SVM is used in the Viola-Jones face identification method to categorize image patches into face and non-facial regions

c) Bioinformatics:

- Protein Classification: SVM predicts the activities of proteins by classifying them according to their structures or sequences.
- Gene Expression Analysis: By using gene expression patterns to categorize tumor samples as either benign or malignant, gene expression analysis is used in the diagnosis of cancer
- SVM has been used to diagnose breast cancer and leukemia with excellent accuracy using microarray datasets.

d) Finance

- Fraud Detection: By categorizing irregularities in payment or banking systems, SVM assists in detecting fraudulent transactions.
- Stock Price Prediction: This method divides historical data into patterns of rising or falling prices in order to forecast stock movements.

e) Healthcare

- Disease Diagnosis: SVM is used to categorize diseases from patient records. For example, it can be used to differentiate between different types of cancer or to diagnose diabetes from health data (such as the Pima dataset).

- Real-world Example: MRI imaging data has been utilized to classify Alzheimer's disease using support vector machines (SVMs).

f) Industry and Engineering:

- Fault Detection: By categorizing vibration data, SVM finds mechanical flaws in systems like turbines or engines.
- Quality control: Used in production to identify faulty items.