SVM kernels and its type

Fig.3

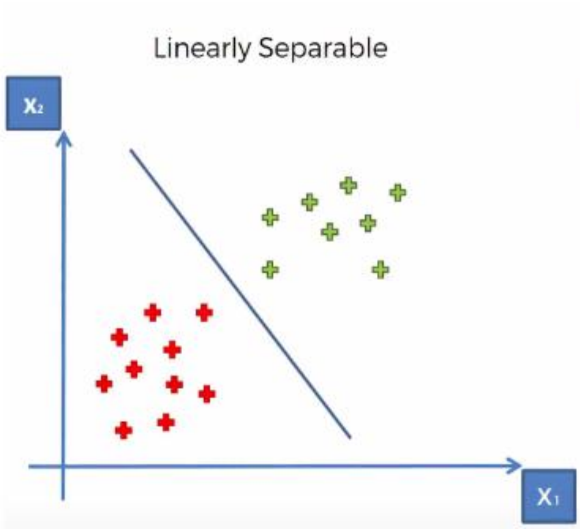
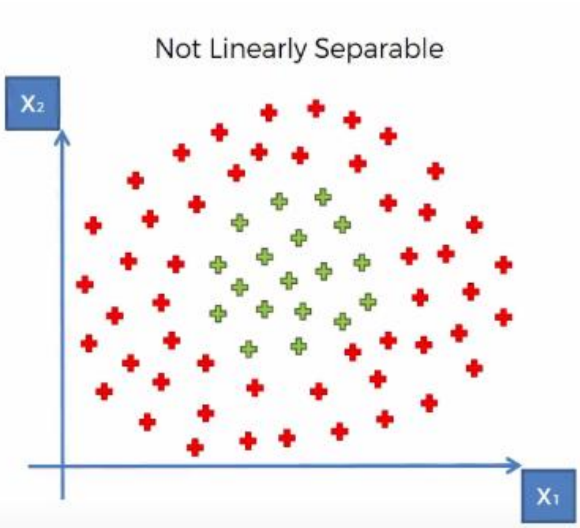
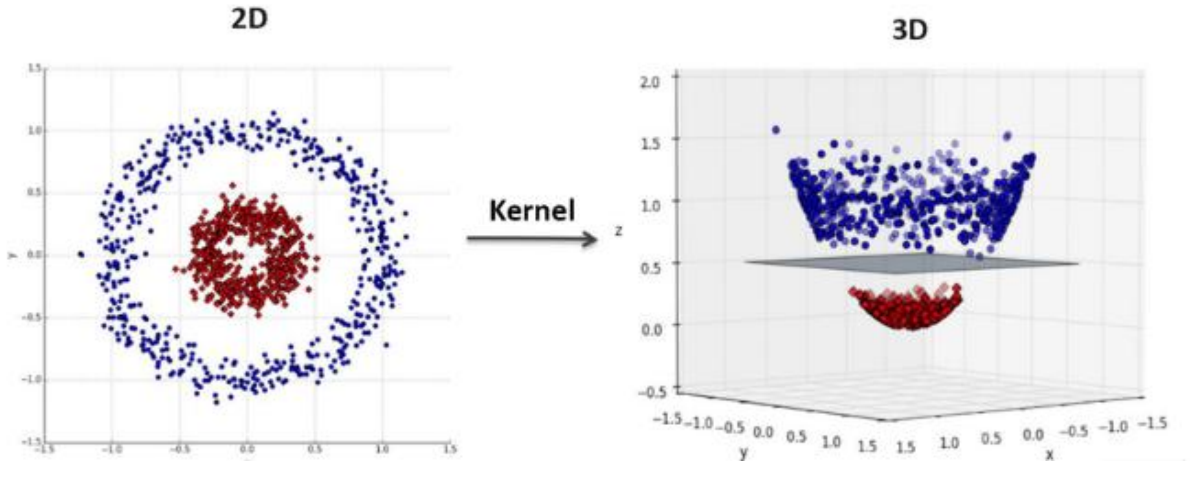
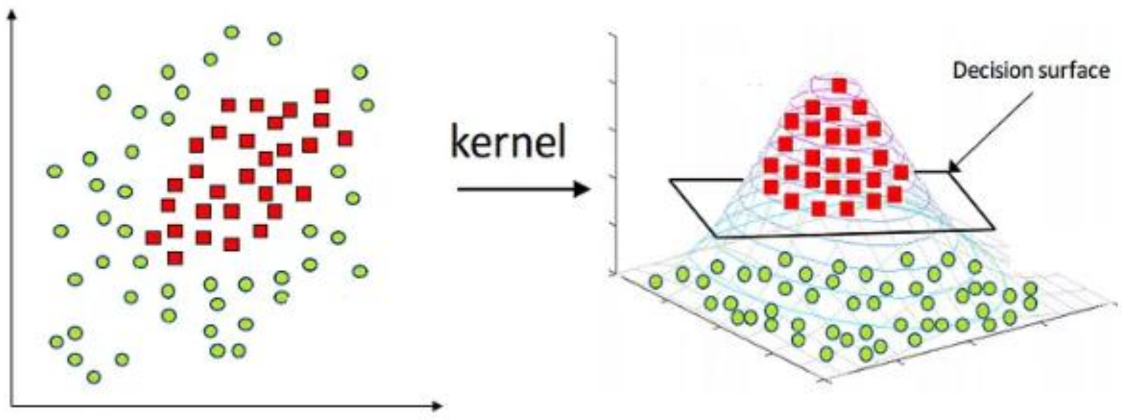


Fig.4



Support Vector Machines (SVMs) are a popular and powerful class of machine learning algorithms used for classification and regression tasks. One of the reasons for their flexibility and effectiveness is their ability to use different types of kernels. In this blog, we'll explore what SVM kernels are, how they work, and the most commonly used kernel functions.

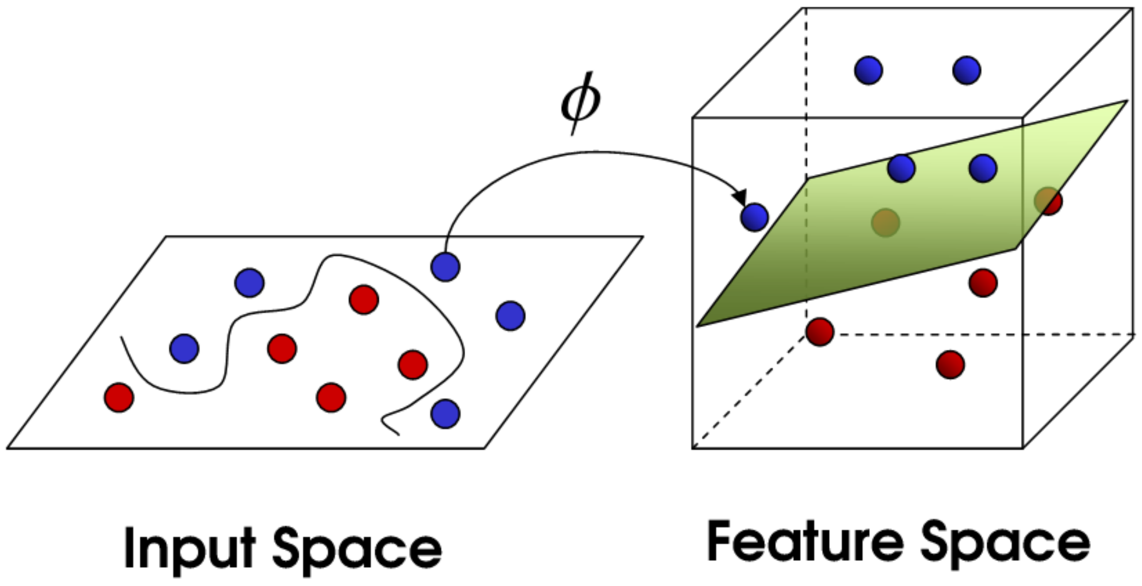




What is an SVM Kernel?

It is a mathematical function that helps organize data and make it easier to classify. The primary goal of an SVM is to find a hyperplane that best separates different classes of data points. However, in many real-world scenarios, the data is not linearly separable in the original feature space. Kernels help by implicitly mapping the original feature space into a higher-dimensional space where the data might be more easily separable.

The original feature space is the space defined by the input features of the data. Each dimension in this space represents one feature. For example, if your data has two features (say, height and weight), the original feature space will be two-dimensional.



Why Use Kernels?

1. **Non-Linearity Handling:** Kernels allow SVMs to handle non-linearly separable data by transforming the feature space. This is achieved without explicitly performing the transformation, which can be computationally expensive.

2. **Flexibility:** Different kernels can be used depending on the nature of the data and the problem at hand, allowing SVMs to adapt to a variety of tasks.

3. **Feature Extraction:** Kernels can implicitly perform feature extraction by projecting data into a space where it becomes linearly separable.

When data is not linearly separable in the original feature space, SVM uses a method called the **kernel trick** to map the data to a higher-dimensional feature space.

Higher-Dimensional Feature Space: By applying a kernel function, the data is transformed into a new, higher-dimensional space where the data may become linearly separable. In this new feature space, SVM can find a linear hyperplane that effectively separates the classes, even though the data appeared non-linear in the original space.

The key idea of SVMs is that we don't need to explicitly compute the mapping to the higher-dimensional feature space. Instead, the **kernel function** computes the similarity between data points in the higher-dimensional space without having to directly compute the coordinates of each point in that space. This allows SVMs to handle complex, non-linear relationships between features while maintaining computational efficiency

Commonly Used SVM Kernels

- Common kernel functions for SVM
  - linear $k(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1 \cdot \mathbf{x}_2$
  - polynomial $k(\mathbf{x}_1, \mathbf{x}_2) = (\gamma\, \mathbf{x}_1 \cdot \mathbf{x}_2 + c)^d$
  - Gaussian or radial basis $k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2\right)$
  - sigmoid $k(\mathbf{x}_1, \mathbf{x}_2) = \tanh(\gamma\, \mathbf{x}_1 \cdot \mathbf{x}_2 + c)$

Here x1 and x2 are data points and gamma. d is the degree of the polynomial

1. C: Inverse of the strength of regularization.

Behavior: As the value of 'c' increases the model gets overfits.

As the value of 'c' decreases the model underfits.

2. γ : Gamma (used only for RBF kernel)

Behavior: As the value of ' γ' increases the model gets overfits.

As the value of ' γ' decreases the model underfits.

Linear Kernel

1. The linear kernel is the simplest and most straightforward kernel function.
2. This kernel is used when the data is already linearly separable. It effectively means that no transformation is applied to the data.
3. **Advantages:**

- Simple and fast to compute.
- Effective for linearly separable data.

4. Disadvantages:

- Not suitable for complex, non-linear data.

Polynomial Kernel

1. The polynomial kernel allows for more complex decision boundaries by adding polynomial features to the data. It is defined as:
2. This kernel can capture interactions between features up to a certain degree.
3. **Advantages:**

- Can model interactions between features.
- Suitable for non-linearly separable data.

4. Disadvantages:

- Computationally more expensive than the linear kernel.
- Risk of overfitting with high-degree polynomials.

Radial Basis Function (RBF) Kernel

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

1. The RBF kernel, also known as the Gaussian kernel, is a popular choice due to its flexibility. It is defined as:
2. This kernel can handle very complex and non-linear relationships.
3. **Advantages:**

- Can handle a wide range of data distributions.
- Effective in high-dimensional spaces.

4. Disadvantages:

- Requires careful tuning of the σ parameter.

- Can be computationally expensive with large datasets.

Sigmoid Kernel

1. **Advantages:**

- Can be used to model relationships similar to those found in neural networks.
- Simple to implement.

1. **Disadvantages:**

- Less commonly used compared to other kernels.
- Can be less effective for certain types of data.

Choosing the Right Kernel

Selecting the appropriate kernel for your SVM model depends on several factors:

- **Data Complexity:** For linearly separable data, the linear kernel is sufficient. For more complex data, consider polynomial or RBF kernels.
- **Computational Resources:** RBF and polynomial kernels are computationally more intensive than the linear kernel. Ensure that your computational resources can handle the increased complexity.
- **Model Performance:** Experiment with different kernels and use cross-validation to determine which kernel yields the best performance for your specific problem.