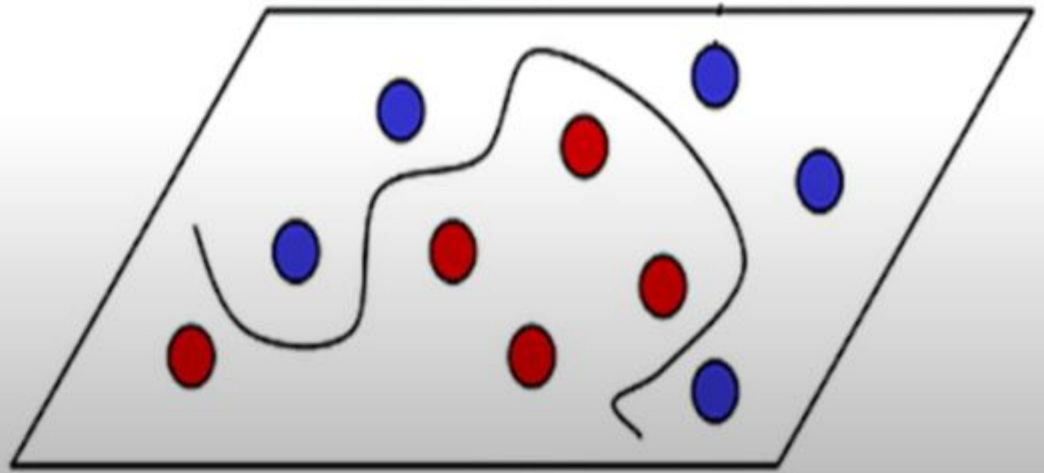
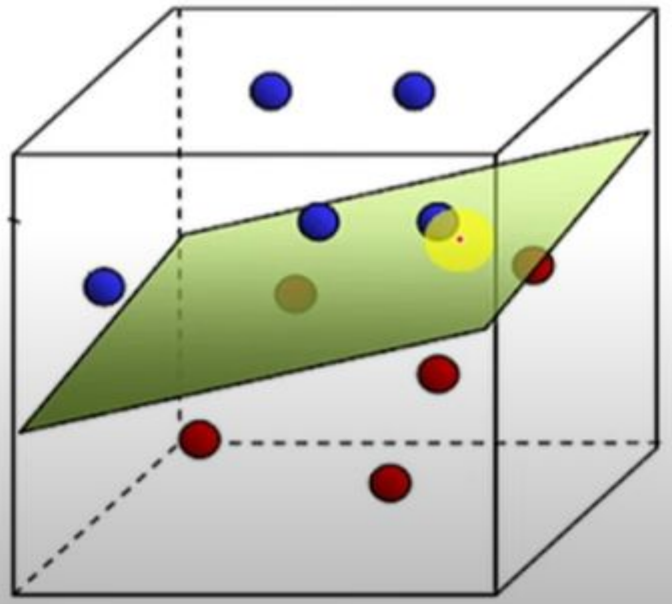
Kernel functions

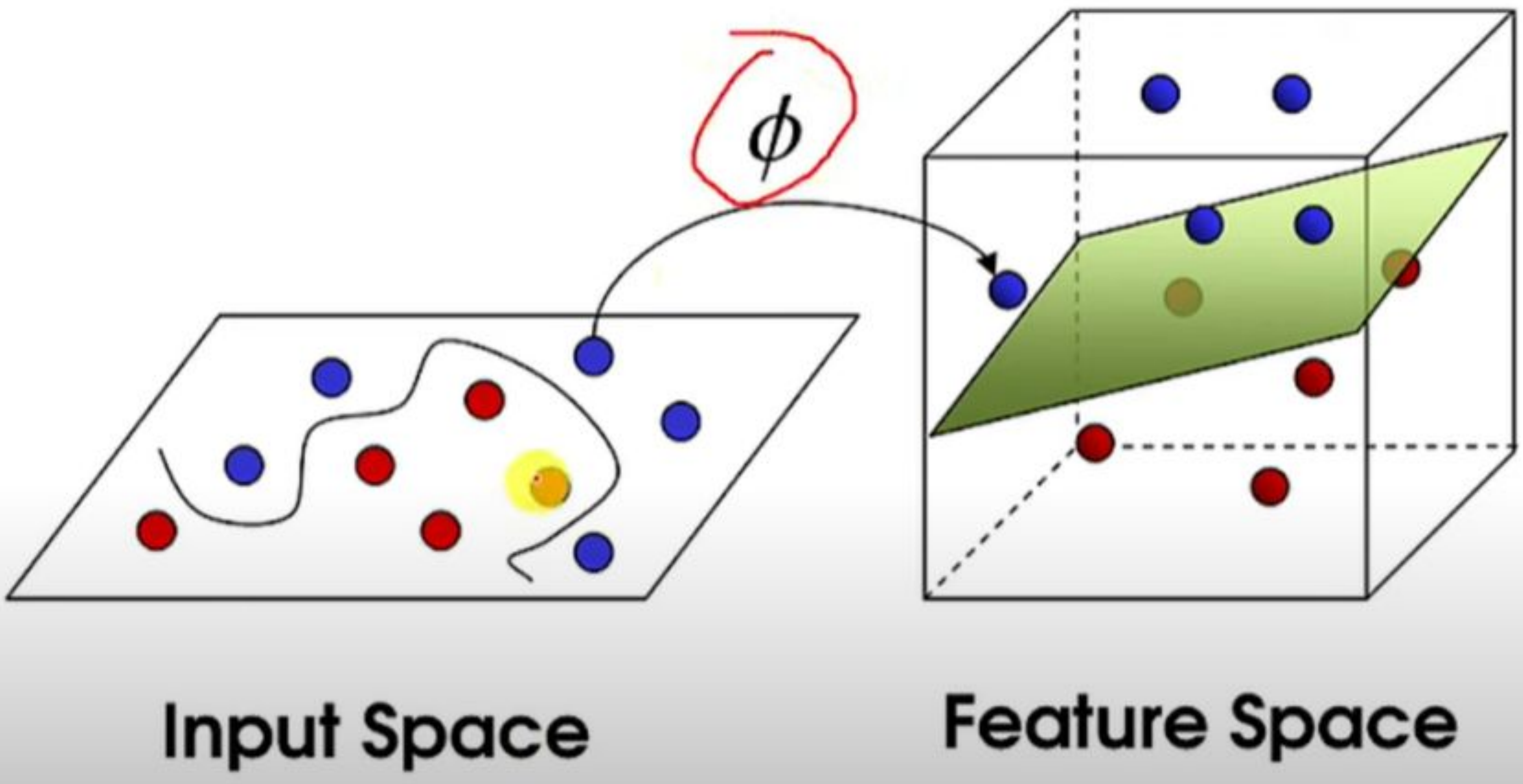
Kernel Trick in Support Vector Machine

- In machine learning applications, the data can be text, image, or video.
- So, there is a need to extract features from these data prior to classification.
- Hence, in the real world, many classification models are complex and mostly require non-linear hyperplanes.





Kernel Trick in Support Vector Machine



Kernel Trick in Support Vector Machine

- For example, one mapping function $\emptyset: \boldsymbol{R}^2 \rightarrow \boldsymbol{R}^3$ used to transform a 2D data to 3D data is given as follows:

$$\emptyset(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^2, \sqrt{2}\boldsymbol{xy}, \boldsymbol{y}^2)$$

- Consider a point (2, 3) in 2D space, if you apply above mapping function we can convert it into 3D space and it looks like this,
- Here $x = 2 \; and \; y = 3$,
- Hence point in 3D space is:

$$\left(\boldsymbol{2^2}, \sqrt{\boldsymbol{2}} * \boldsymbol{2} * \boldsymbol{3}, \boldsymbol{3^2}\right) = (\boldsymbol{4}, \boldsymbol{6}\sqrt{\boldsymbol{2}}, \boldsymbol{9})$$

Kernel Trick in Support Vector Machine

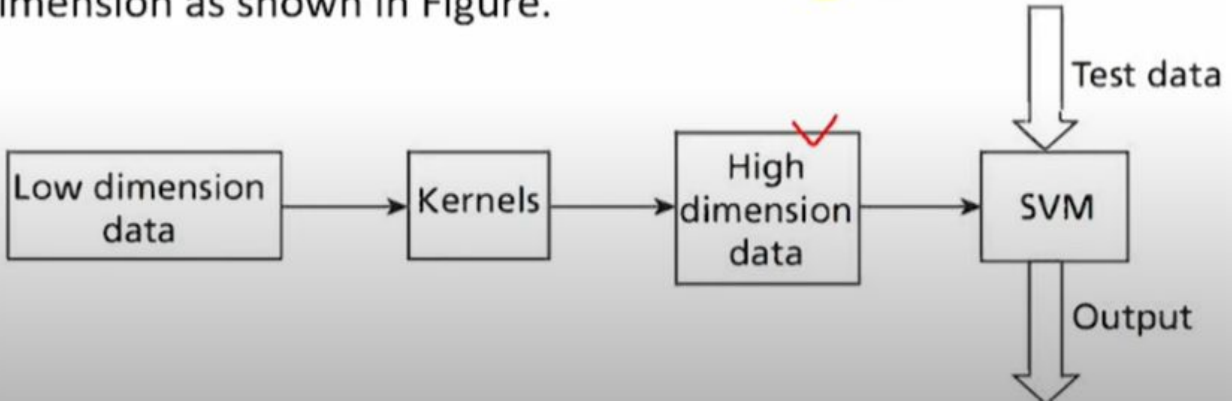
- While mapping functions play an important role, there are many disadvantages, as mapping involves more computations and learning costs.

D_1 D_2

- Also, the disadvantages of transformations are that there is no generalized thumb rule available describing what transformations should be applied and if the data is large, mapping process takes huge amount of time.
- In real applications, there might be many features in the data and applying transformations that involve many polynomial combinations of these features will lead to extremely high and impractical computational costs.
- In this context, only kernels are useful.

Kernel Trick in Support Vector Machine

- **What is a Kernel?**
- Kernels are a set of functions used to transform data from lower dimension to higher dimension and to manipulate data using dot product at higher dimensions.
- The use of kernels is to apply transformation to data and perform classification at the higher dimension as shown in Figure.



Kernel Trick in Support Vector Machine

- Kernel Trick for 2nd degree Polynomial Mapping

$\emptyset(x, y) = (x^2, \sqrt{2}xy, y^2)$

$$\phi(\mathbf{a})^T \cdot \phi(\mathbf{b}) = \begin{pmatrix} {a_1}^2 \\ \sqrt{2}\, a_1 a_2 \\ {a_2}^2 \end{pmatrix}^T \cdot \begin{pmatrix} {b_1}^2 \\ \sqrt{2}\, b_1 b_2 \\ {b_2}^2 \end{pmatrix} = {a_1}^2 {b_1}^2 + 2a_1 b_1 a_2 b_2 + {a_2}^2 {b_2}^2$$

$$= (a_1 b_1 + a_2 b_2)^2 = \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^T \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right)^2 = \left(\mathbf{a}^T \cdot \mathbf{b} \right)^2$$

Types of Kernels

- Linear Kernel
- Polynomial Kernel
- Exponential Kernel
- Homogeneous Kernel
- Inhomogeneous Kernel

Types of kernel

- Linear Kernel
- Polynomial Kernel
- Exponential Kernel
- Homogeneous Kernel

- Inhomogeneous Kernel
- Gaussian Kernel
- Sigmoid Kernel
- Radial-basis function kernel

Linear Kernel

- Linear kernels are of the type

$$k(x, y) = x^T . y$$

- where x and y are two vectors.
- Therefore $k(x, y) = \emptyset(x).\emptyset(y) = x^T . y$

Polynomial Kernel

- For inhomogeneous kernels, this is given as:

$$k(x, y) = (c + x^T y)^q$$

- Here c is a constant and q is the degree of the polynomial.
- If c is zero and degree is one, the polynomial kernel is reduced to a linear kernel.
- The value of degree q should be optimal as more degree may lead to overfitting.

Linear Kernel

- Consider two data points $x = \binom{1}{2}$ and $y = (2,3)$ with c =1. Apply linear, homogeneous and inhomogeneous kernels.
- **Solution:**
- The kernel is given by $k(x,y) = (x^T y)^q$
- If q = 1 the it is called linear kernel
- $k(x,y) = \left(\binom{1}{2}^T (2,3)\right)^1 = 8$

Homogeneous kernel

- Consider two data points $x = \binom{1}{2}$ and $y = (2, 3)$ with c =1. Apply linear, homogeneous and inhomogeneous kernels.
- **Solution:**
- The kernel is given by $k(x, y) = (x^T y)^q$
- If q = 2 the it is called homogeneous or quadratic kernel
- $k(x, y) = \left(\binom{1}{2}^T (2,3)\right)^2 = 8^2 = 64$

Inhomogeneous kernel

- Consider two data points $x = \binom{1}{2}$ and $y = (2,3)$ with c =1. Apply linear, homogeneous and inhomogeneous kernels.
- **Solution:**
- The kernel is given by $k(x,y) = (x^T y)^q$
- If q = 2 and c =1 the it is called inhomogeneous kernel
- $k(x,y) = (c + x^T y)^q = \left(1 + \binom{1}{2}^T (2,3)\right)^2 = (1+8)^2 = 81$

Gaussian Kernel

- Radial Basis Functions (RBFs) or Gaussian kernels are extremely useful in SVM.
- The RBF function is shown as below:

$$k(x, y) = e^{\frac{-(x-y)^2}{2\sigma^2}}$$

- Here, y is an important parameter. If y is small, then the RBF is similar to linear SVM and if y is large, then the kernel is influenced by more support vectors.
- The RBF performs the dot product in R^{∞}, and therefore, it is highly effective in separating the classes and is often used.

- Consider two data points x = (1, 2) and y = (2, 3) with $\sigma = 1$.
- Apply RBF kernel and find the value of RBF kernel for these points.

$$k(x, y) = e^{\frac{-(x-y)^2}{2\sigma^2}}$$

- Substitute the value of x and y in RBF kernel.
- The squared distance between the points (1, 2) and (2, 3) is given as:

$$(1 - 2)^2 + (2 - 3)^2 = 2$$

- If $\sigma = 1$, then $k(x, y) = e^{\left\{-\frac{2}{2}\right\}} = e^{-1} = 0.3679.$

Sigmoid Kernel

- The sigmoid kernel is given as:

$$k(x_i, y_i) = \tanh(k x_i y_j - \sigma)$$

Why do we need kernel trick

Kernel Trick

- Kernel trick means replacing the dot product in mapping functions with a kernel function.

$$\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{y}) = \emptyset(\boldsymbol{x}) \cdot \emptyset(\boldsymbol{y})$$

- Similar to mapping functions, kernels help in mapping data from input space to higher-dimensional feature space with least computations.
- Performing the kernel operation is much easier compared to mapping functions.
- This is illustrated in the following numerical example.

- Consider two data points (1, 2) and (3, 4)
- Apply a polynomial kernel $k(x, y) = (x^T y)^2$ and show that it is equivalent to mapping function $\emptyset = (x^2, y^2, \sqrt{2}\, xy)$
- **Solution:**
- The mapping function is given as $\emptyset = (x^2, y^2, \sqrt{2}\, xy)$
- Let us apply the mapping function first for the first data point (1, 2) using Eq. given as:
- $\emptyset = (x^2, y^2, \sqrt{2}\, xy)$
- $First\ data\ point\ \emptyset(1,2) = \left(1^2, 2^2, \sqrt{2} * 1 * 2\right) = \left(1, 4, 2\sqrt{2}\right)$
- $Second\ data\ point\ \emptyset(3,4) = \left(3^2, 4^2, \sqrt{2} * 3 * 4\right) = \left(9, 16, 12\sqrt{2}\right)$

- Consider two data points (1, 2) and (3, 4)
- Apply a polynomial kernel $k(x, y) = (x^T y)^2$ and show that it is equivalent to mapping function $\emptyset = (x^2, y^2, \sqrt{2}\,xy)$
- **Solution:**
- $First\ data\ point\ \emptyset(1,2) = \left(1^2, 2^2, \sqrt{2} * 1 * 2\right) = \left(1, 4, 2\sqrt{2}\right)$
- $Second\ data\ point\ \emptyset(3,4) = \left(3^2, 4^2, \sqrt{2} * 3 * 4\right) = \left(9, 16, 12\sqrt{2}\right)$
- $\emptyset(1,2).\emptyset(3,4) = (1,\ 4,\ 2\sqrt{2}) \cdot (9, 16, 12\sqrt{2}) = (1 \times 9 + 4 \times 16 + 24(2) = 121$

- It can be seen the computation involves many multiplication operations.
- The operation can be computed quickly using kernel functions.
- Now, using polynomial kernel function, $k(x, y) = (x^T y)^2$, it can be computed as:

$$\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 3 & 4 \end{pmatrix}\right)^2 = 11^2 = 121$$