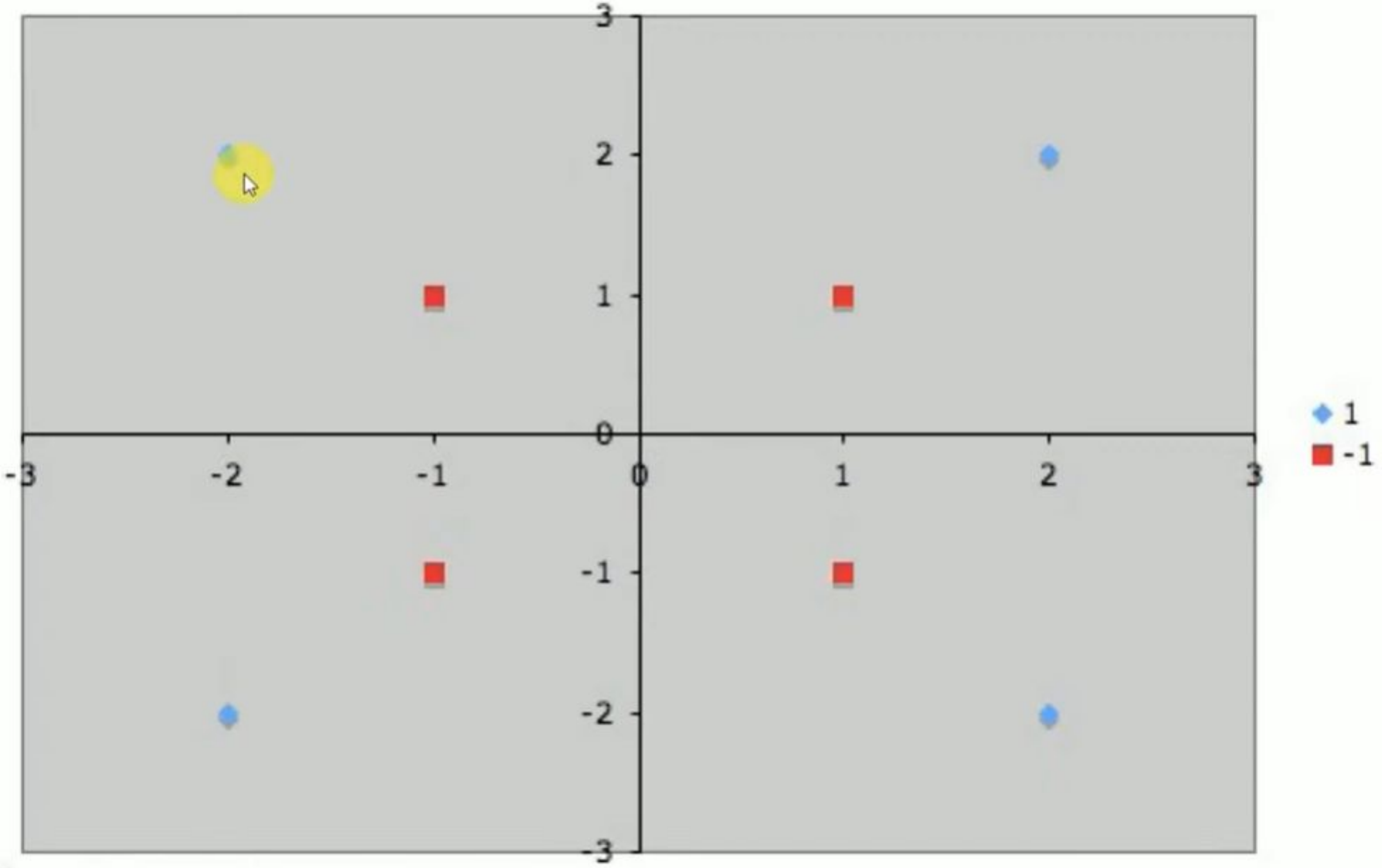
Support Vector Machine – Non Linear Example Solved

Suppose we are given the following positively labeled data points,

$$\left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix} \right\}$$

and the following negatively labeled data points,

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$$



3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
1
-1

Support Vector Machine – Non Linear Example Solved

- Our goal, again, is to discover a separating hyperplane that accurately discriminates the two classes.
- Of course, it is obvious that no such hyperplane exists in the input space
- Therefore, we must use a nonlinear SVM (that is, we need to convert data from one feature space to another.

$$\Phi_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

Support Vector Machine – Non Linear Example Solved

$$\Phi_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

Positive Examples

$$\left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix} \right\} \rightarrow$$

Support Vector Machine – Non Linear Example Solved

$$\Phi_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

Positive Examples

$$\left\{\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix}\right\} \rightarrow \left\{\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 10 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 10 \end{pmatrix}\right\}$$

Negative Examples

Support Vector Machine – Non Linear Example Solved

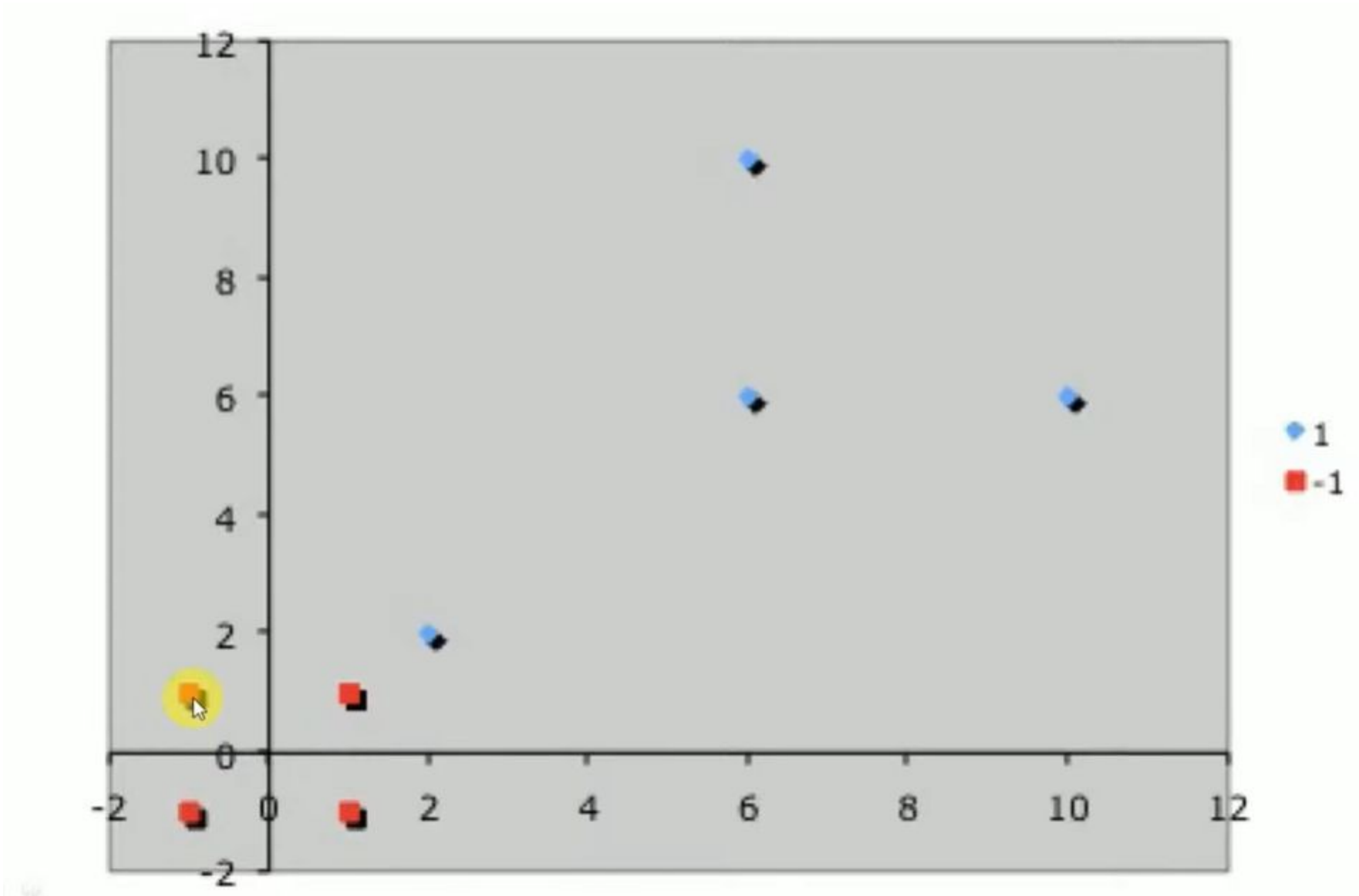
$$\Phi_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

Positive Examples

$$\left\{\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix}\right\} \rightarrow \left\{\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 10 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \begin{pmatrix} 6 \\ 10 \end{pmatrix}\right\}$$

Negative Examples

$$\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\} \rightarrow \left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$$



12
10
8
6
4
2
0
-2
-2
0
2
4
6
8
10
12
1
-1

Support Vector Machine – Non Linear Example Solved

- Now we can easily identify the support vectors,

$$\left\{ s_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, s_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$$

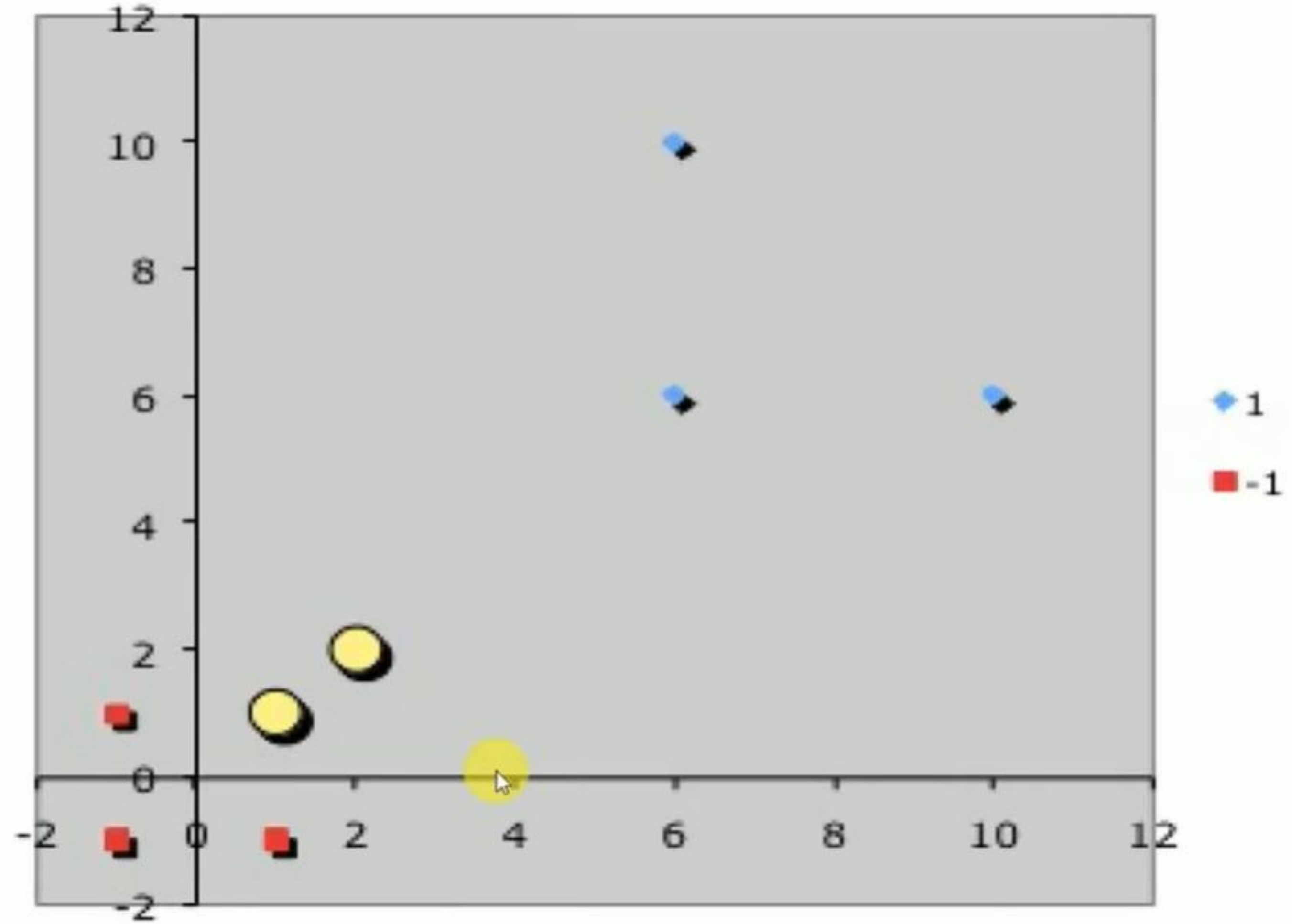
Support Vector Machine – Non Linear Example Solved

- Now we can easily identify the support vectors,

$$\left\{ s_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, s_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}$$

- Each vector is augmented with a 1 as a bias input

$$\tilde{s_1} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \tilde{s_2} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$


12
10
8
6
4
2
0
-2
-2
0
2
4
6
8
10
12
1
-1

Support Vector Machine – Non Linear Example Solved

$$\alpha_1 \tilde{s_1} \cdot \tilde{s_1} + \alpha_2 \tilde{s_2} \cdot \tilde{s_1} = -1$$
$$\alpha_1 \tilde{s_1} \cdot \tilde{s_2} + \alpha_2 \tilde{s_2} \cdot \tilde{s_2} = +1$$

$$\alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = 1$$

Support Vector Machine – Non Linear Example Solved

$$\alpha_1 \tilde{s_1} \cdot \tilde{s_1} + \alpha_2 \tilde{s_2} \cdot \tilde{s_1} = -1$$
$$\alpha_1 \tilde{s_1} \cdot \tilde{s_2} + \alpha_2 \tilde{s_2} \cdot \tilde{s_2} = +1$$

$$\alpha_1(1+1+1) + \alpha_2(2+2+1) = -1$$
$$\alpha_1(2+2+1) + \alpha_2(4+4+1) = 1$$

$$\alpha_1 \begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}2\\2\\1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} = -1$$

$$3\alpha_1 + 5\alpha_2 = -1$$
$$5\alpha_1 + 9\alpha_2 = 1$$

$$\alpha_1 \begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}2\\2\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}2\\2\\1\end{pmatrix}\begin{pmatrix}2\\2\\1\end{pmatrix} = 1$$

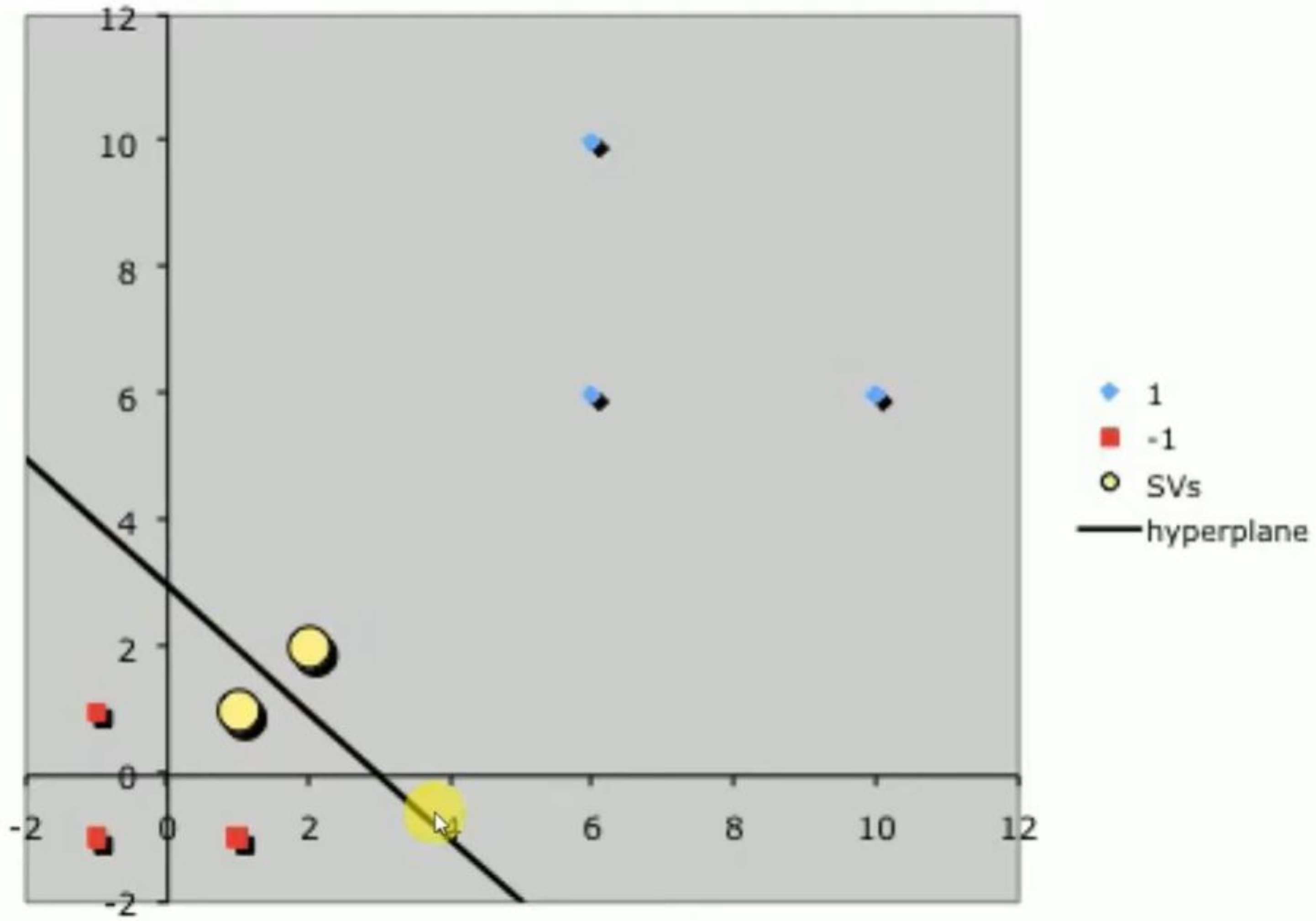
$$\alpha_1 = -7$$
$$\alpha_2 = 4$$

Support Vector Machine – Non Linear Example Solved

$$\tilde{w} = \sum_i \alpha_i \tilde{s}_i = -7 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 4 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \\ -3 \end{pmatrix}$$

- Finally, remembering that our vectors are augmented with a bias.
- We can equate the last entry in $\widetilde{w}$ as the hyperplane offset b and write the separating
- Hyperplane equation $\boldsymbol{y} = \boldsymbol{wx} + \boldsymbol{b}$
- with $\boldsymbol{w} = \binom{1}{1}$ and $\boldsymbol{b} = -\mathbf{3}$.



12
10
8
6
4
2
0
-2
-2
0
2
6
8
10
12
1
-1
SVs
hyperplane

Kernel Trick in Support Vector Machine Solved Example

Kernel Trick

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

$$k(x, y) = \emptyset(x) \cdot \emptyset(y)$$

- 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.

Kernel Trick in Support Vector Machine Solved 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:
- $\emptyset = (x^2, y^2, \sqrt{2}\,xy)$
- $First\ data\ point\ \emptyset(1, 2) = \left(1^2, 2^2, \sqrt{2} * 1 * 2\right) = \left(1, 4, 2\sqrt{2}\right)$
- $Second\ data\ point\ \emptyset(3, 4) = \left(3^2, 4^2, \sqrt{2} * 3 * 4\right) = \left(9, 16, 12\sqrt{2}\right)$

Kernel Trick in Support Vector Machine Solved 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:**
- *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* $\emptyset(3,4) = \left(3^2, 4^2, \sqrt{2} * 3 * 4\right) = \left(9, 16, 12\sqrt{2}\right)$
- $\emptyset(1,2).\emptyset(3,4) = (1,\ 4,\ 2\sqrt{2}) \cdot (9, 16, 12\sqrt{2}) = (1 \times 9 + 4 \times 16 + 24(2) = 121$

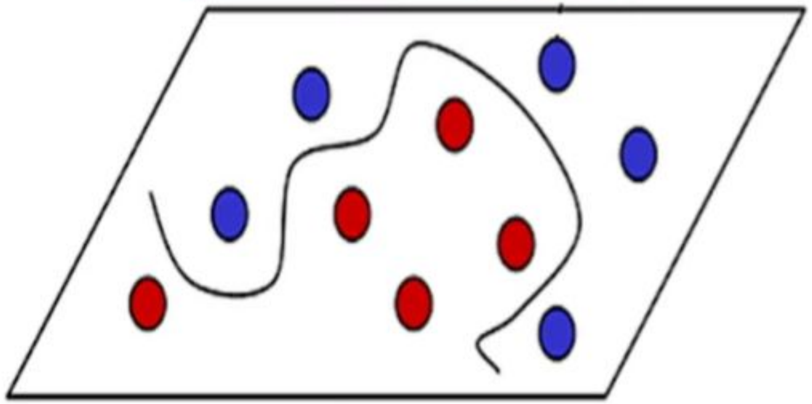
Kernel Trick in Support Vector Machine Solved Example

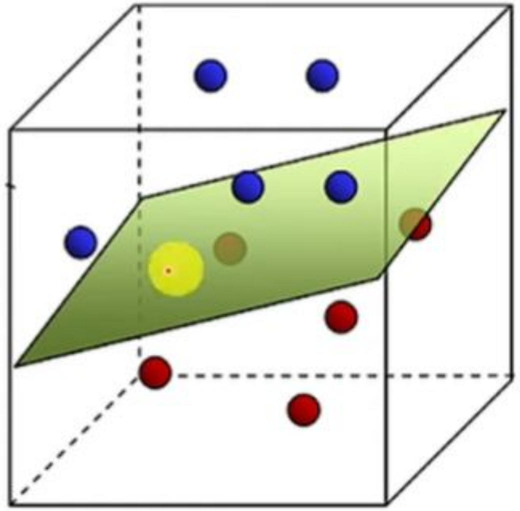
- It can be seen the computation involves many multiplication operations.
- The operation can be computed quickly using kernel functions.
- Now, using polynomial kernel function, $k(x, y) = (x^T y)^2$, it can be computed as:

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

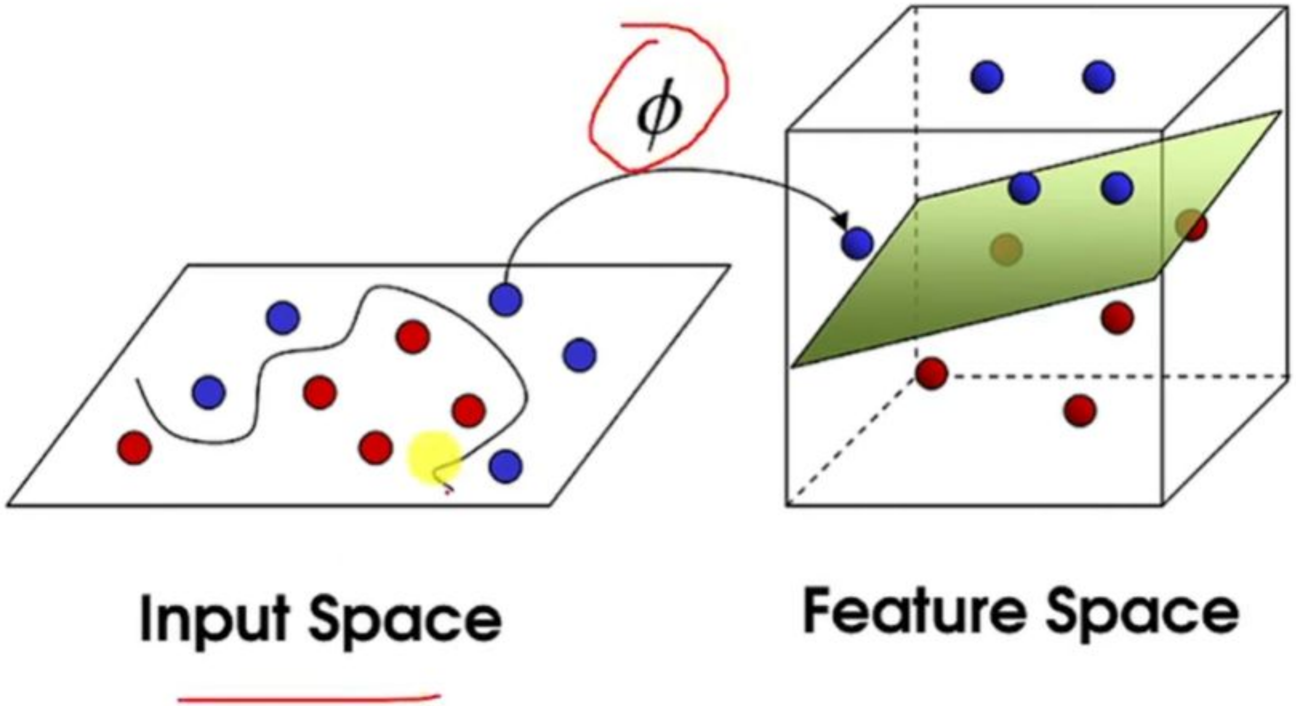
Kernel Trick in Support Vector Machine

- 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:

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

- 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:

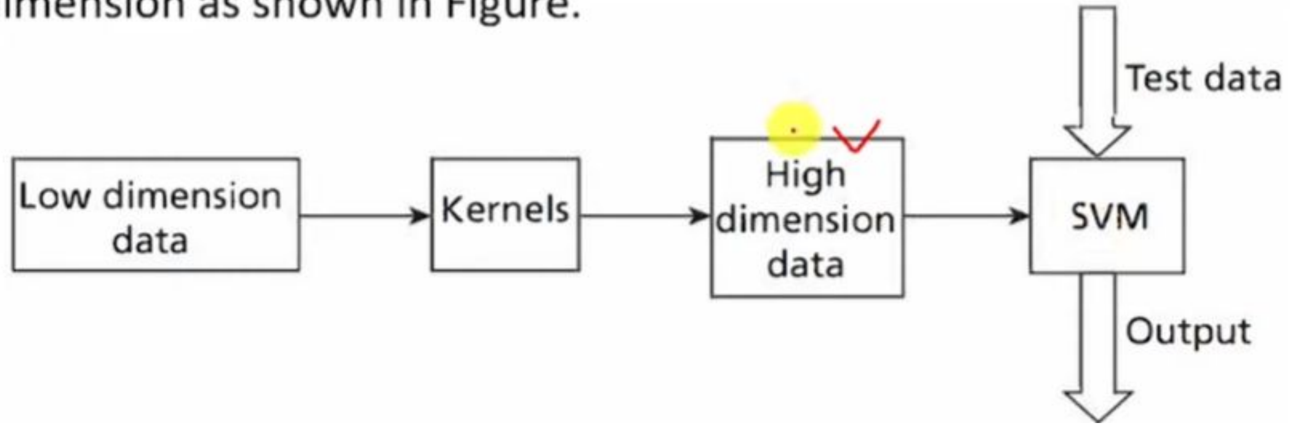
$$(\mathbf{2}^2, \sqrt{\mathbf{2}} * \mathbf{2} * \mathbf{3}, \mathbf{3}^2) = (\mathbf{4}, \mathbf{6}\sqrt{\mathbf{2}}, \mathbf{9})$$

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.
- Kernels are used to compute the value without transforming the data.

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**

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

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

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

Kernel Trick in Support Vector Machine

Types of Kernels

- Linear Kernel
- Polynomial Kernel
- Exponential Kernel
- Homogeneous Kernel
- Inhomogeneous Kernel
- Gaussian Kernel
- Hyperbolic or the Sigmoid Kernel
- Radial-basis function kernel
- Etc...

Types of Kernel in Support Vector Machine

Linear Kernel

- Linear kernels are of the type

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

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

Types of Kernel in Support Vector Machine

Linear Kernel

- Linear kernels are of the type

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

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

Types of Kernel in Support Vector Machine

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.

Types of Kernel in Support Vector Machine

- Consider two data points $x = \begin{pmatrix}1\\2\end{pmatrix}$ 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
- $k(x,y) = \left(\begin{pmatrix}1\\2\end{pmatrix}^T \cdot (2,3)\right)^1 = 8$

Types of Kernel in Support Vector Machine

- Consider two data points $x = \begin{pmatrix}1\\2\end{pmatrix}$ 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
- $k(x,y) = \left(\begin{pmatrix}1\\2\end{pmatrix}^T (2,3)\right)^2 = 8^2 = 64$

Types of Kernel in Support Vector Machine

- 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
- $k(x, y) = (\underline{c} + x^T y)^q = \left(1 + \underline{\binom{1}{2}^T (2,3)}\right)^2 = \underline{(1 + 8)}^2 = 81$

Types of Kernel in Support Vector Machine

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.

Types of Kernel in Support Vector Machine

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

Types of Kernel in Support Vector Machine

Sigmoid Kernel

- The sigmoid kernel is given as:

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

How to Select Best Hyperplane in SVM

- The hyperplane function for two variables is $\boldsymbol{b + a_1 x_1 + a_2 x_2}$.
- If two hyperplanes given are

  $\boldsymbol{5 + 2x_1 + 5x_2}$ for **classifier 1** and

  $\boldsymbol{5 + 20x_1 + 50x_2}$, for **classifier 2**.
- Find the distance error function and pick a good classifier constructed using these hyperplanes?

How to Select Best Hyperplane in SVM

- Distance error function of SVM is given by: $\sqrt{a_1^2 + a_2^2}$
- The weight vector for the first equation omitting the intercept is (2, 5).
- $\boldsymbol{Distance\ Error\ 1} = \sqrt{2^2 + 5^2} = 5.39$
- The weight vector for the first equation omitting the intercept is (20, 50).
- $\boldsymbol{Distance\ Error\ 2} = \sqrt{20^2 + 50^2} = 53.85$

- **Hyperplane for classifier 1**
- $5 + 2x_1 + 5x_2$
- **Hyperplane for classifier 2**
- $5 + 20x_1 + 50x_2$

How to Select Best Hyperplane in SVM

- $Distance\ Error\ 1 = \sqrt{2^2 + 5^2} = 5.39$
- $Distance\ Error\ 2 = \sqrt{20^2 + 50^2} = 53.85$

- For classifier 1 → $\frac{2}{||w||} = \frac{2}{5.39} = 0.37$
- For Classifier 2 → $\frac{2}{||w||} = \frac{2}{53.85} = 0.037$

- **Hyperplane for classifier 1**
- $5 + 2x_1 + 5x_2$
- **Hyperplane for classifier 2**
- $5 + 20x_1 + 50x_2$