SUPPORT VECTOR MACHINES

Machine Learning
Week 15

INTRODUCTION

Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92 (**Pittsburgh**).

Set of related supervised learning methods used for classification and regression

- Classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data

They belong to a family of generalized linear classifiers

Became popular after the technique gave accuracy comparable to sophisticated neural networks with elaborated features in a handwriting recognition task

SVM-ANN

SVM formulation uses the Structural Risk Minimization (SRM) principle, which has been shown to be superior, [4], to traditional Empirical Risk Minimization (ERM) principle, used by conventional neural networks

Convergence issue of neural network in case of multilayer neurons

We are given a set of training data $\{(\mathbf{x}1,y1)\ldots\ldots (\mathbf{x}l,yl)\}$ in $Rn \times R$ sampled space according to unknown probability distribution $P(\mathbf{x},y)$, and a loss function $V(y,f(\mathbf{x}))$ that measures the error

For a given $\mathbf{x}$, $f(\mathbf{x})$ is "predicted" instead of the actual value y.

The problem consists in finding a function f that minimizes the expectation of the error on new data that is:

Finding a function f that minimizes the expected error: $\int V(y,f(x))\, P(x, y)\, dx\, dy$

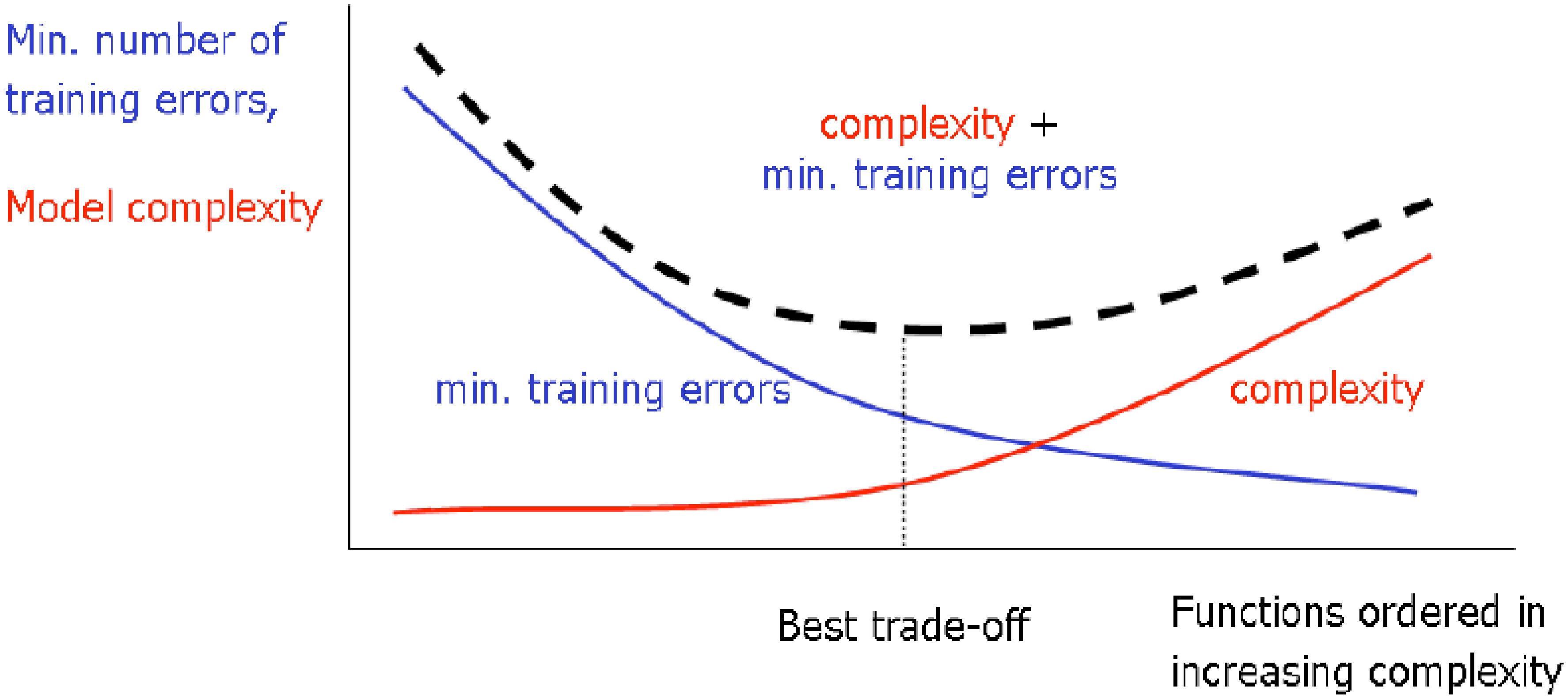
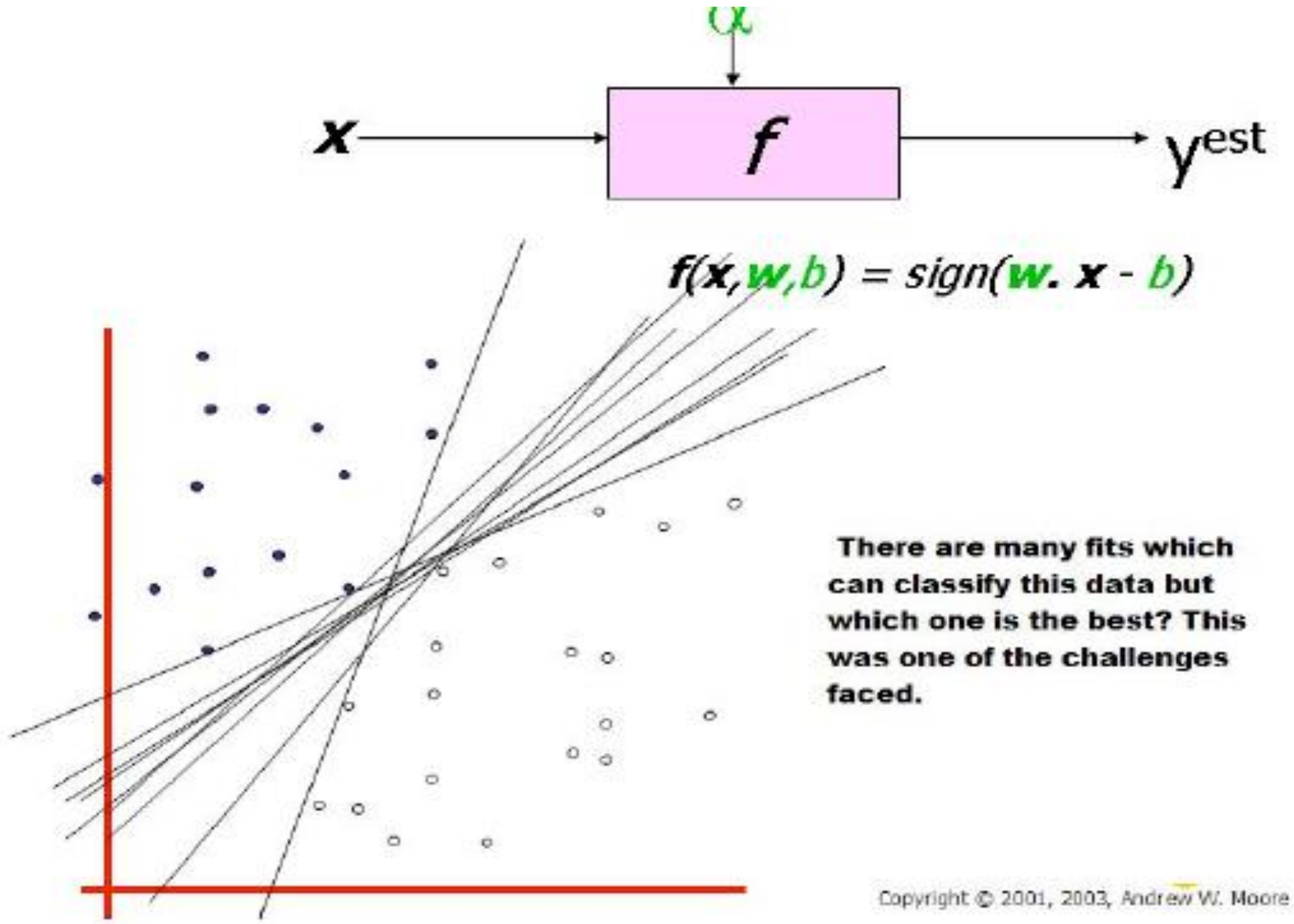
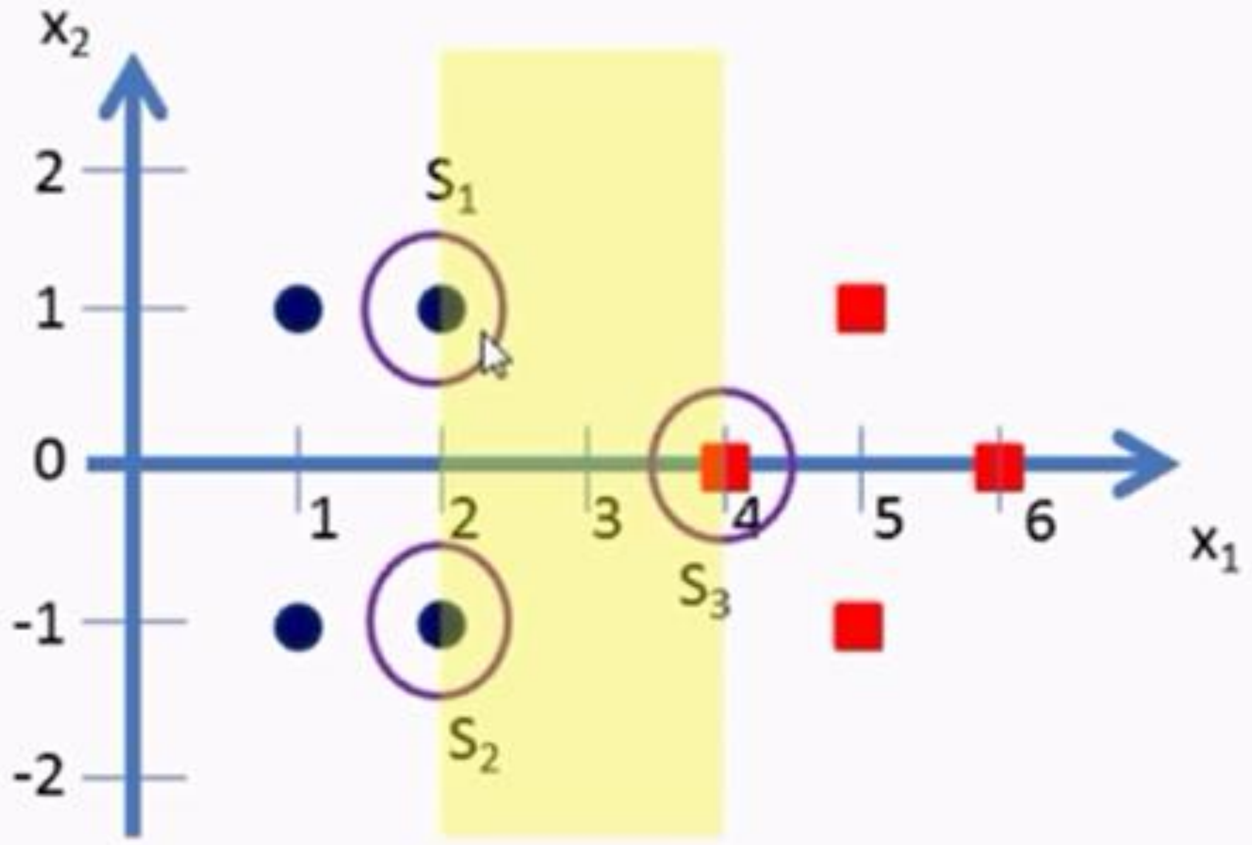


Figure 1: Number of Epochs Vs Complexity. [8][9][11]

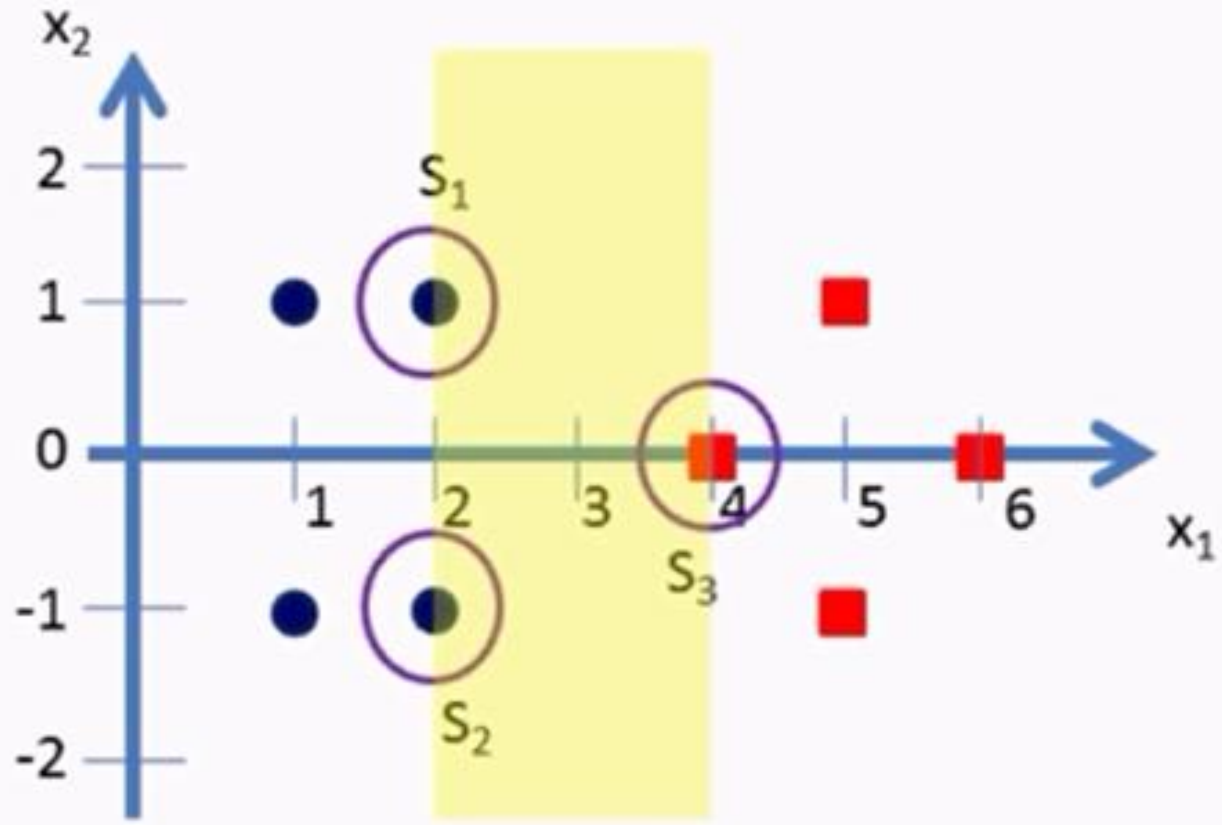
Now let us look at another example where we plot the data and try to classify it and we see that there are many hyper planes which can classify it. But which one is better?



- Here we select 3 Support Vectors to start with.
- They are S_1, S_2 and S_3.



- Here we select 3 Support Vectors to start with.
- They are S_1, S_2 and S_3.



$$S_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

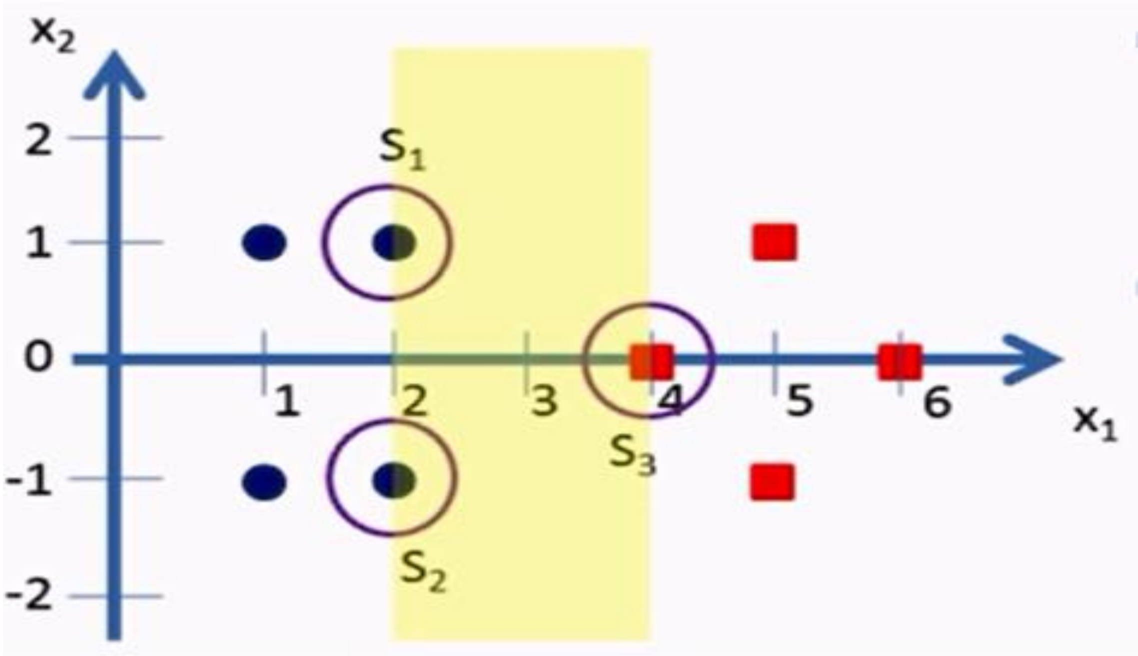
$$S_3 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

- Here we will use vectors augmented with a 1 as a bias input, and for clarity we will differentiate these with an over-tilde. That is:

$$S_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad \widetilde{S_1} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \qquad \widetilde{S_2} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

$$S_3 = \begin{pmatrix} 4 \\ 0 \end{pmatrix} \qquad \widetilde{S_3} = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$



- Now we need to find 3 parameters α_1, α_2, and α_3 based on the following 3 linear equations:

$$\alpha_1 \widetilde{S_1}.\widetilde{S_1} + \alpha_2 \widetilde{S_2}.\widetilde{S_1} + \alpha_3 \widetilde{S_3}.\widetilde{S_1} = -1 \;\; (-ve\ class)$$

$$\alpha_1 \widetilde{S_1}.\widetilde{S_2} + \alpha_2 \widetilde{S_2}.\widetilde{S_2} + \alpha_3 \widetilde{S_3}.\widetilde{S_2} = -1 \; (-ve\ class)$$

$$\alpha_1 \widetilde{S_1}.\widetilde{S_3} + \alpha_2 \widetilde{S_2}.\widetilde{S_3} + \alpha_3 \widetilde{S_3}.\widetilde{S_3} = +1 \;\; (+ve\ class)$$

- Now we need to find 3 parameters α_1, α_2, and α_3 based on the following 3 linear equations:

$$\alpha_1 \widetilde{S_1} . \widetilde{S_1} + \alpha_2 \widetilde{S_2} . \widetilde{S_1} + \alpha_3 \widetilde{S_3} . \widetilde{S_1} = -1 \ (-ve\ class)$$

$$\alpha_1 \widetilde{S_1} . \widetilde{S_2} + \alpha_2 \widetilde{S_2} . \widetilde{S_2} + \alpha_3 \widetilde{S_3} . \widetilde{S_2} = -1 \ (-ve\ class)$$

$$\alpha_1 \widetilde{S_1} . \widetilde{S_3} + \alpha_2 \widetilde{S_2} . \widetilde{S_3} + \alpha_3 \widetilde{S_3} . \widetilde{S_3} = +1 \ (+ve\ class)$$

Let's substitute the values for $\widetilde{S_1}$, $\widetilde{S_2}$ and $\widetilde{S_3}$ in the above equations.

$$\widetilde{S_1} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \qquad \widetilde{S_2} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \qquad \widetilde{S_3} = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

$$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} . \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} . \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} . \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} . \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} . \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} . \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = +1$$

$$\alpha_1 \begin{pmatrix}2\\1\\1\end{pmatrix}.\begin{pmatrix}2\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}2\\-1\\1\end{pmatrix}.\begin{pmatrix}2\\1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}4\\0\\1\end{pmatrix}.\begin{pmatrix}2\\1\\1\end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix}2\\1\\1\end{pmatrix}.\begin{pmatrix}2\\-1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}2\\-1\\1\end{pmatrix}.\begin{pmatrix}2\\-1\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}4\\0\\1\end{pmatrix}.\begin{pmatrix}2\\-1\\1\end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix}2\\1\\1\end{pmatrix}.\begin{pmatrix}4\\0\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}2\\-1\\1\end{pmatrix}.\begin{pmatrix}4\\0\\1\end{pmatrix} + \alpha_3 \begin{pmatrix}4\\0\\1\end{pmatrix}.\begin{pmatrix}4\\0\\1\end{pmatrix} = +1$$

- After simplification we get:

$$6\alpha_1 + 4\alpha_2 + 9\alpha_3 = -1$$

$$4\alpha_1 + 6\alpha_2 + 9\alpha_3 = -1$$

$$9\alpha_1 + 9\alpha_2 + 17\alpha_3 = +1$$

- Simplifying the above 3 simultaneous equations we get: α_1=α_2= -3.25 and α_3= 3.5.

HYPERPLANE

$\alpha_1 = \alpha_2 = -3.25$ and $\alpha_3 = 3.5$

$$\widetilde{S_1} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \quad \widetilde{S_2} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad \widetilde{S_3} = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

- The hyper plane that discriminates the positive class from the negative class is give by:

$$\widetilde{w} = \sum_i \alpha_i \widetilde{S_i}$$

- Substituting the values we get:

$$\widetilde{w} = \alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

$$\widetilde{w} = (-3.25).\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + (-3.25).\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + (3.5).\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}$$

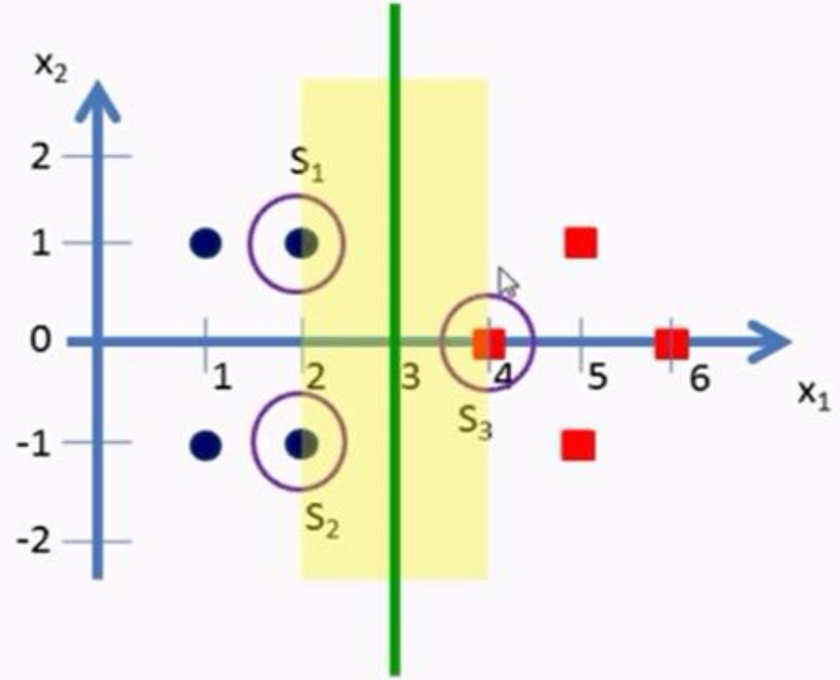
HOW TO INTERPRET THE HYPERPLANE

$$\tilde{w} = (-3.25).\begin{pmatrix}2\\1\\1\end{pmatrix} + (-3.25).\begin{pmatrix}2\\-1\\1\end{pmatrix} + (3.5).\begin{pmatrix}4\\0\\1\end{pmatrix} = \begin{pmatrix}1\\0\\-3\end{pmatrix}$$

- Our vectors are augmented with a bias.
- Hence we can equate the entry in $\tilde{w}$ as the hyper plane with an offset b.
- Therefore the separating hyper plane equation $y = wx + b$ with $w = \begin{pmatrix}1\\0\end{pmatrix}$ and offset $b = -3$.

HOW TO PLOT THE HYPERPLANE?

- $y = wx + b$ with $w = \begin{pmatrix}1\\0\end{pmatrix}$ and offset $b = -3$.



Here,
(1 0) is a straight line with gradient of 90 degrees
Offset -3 will pass through +3
Offset +2 will pass through -2

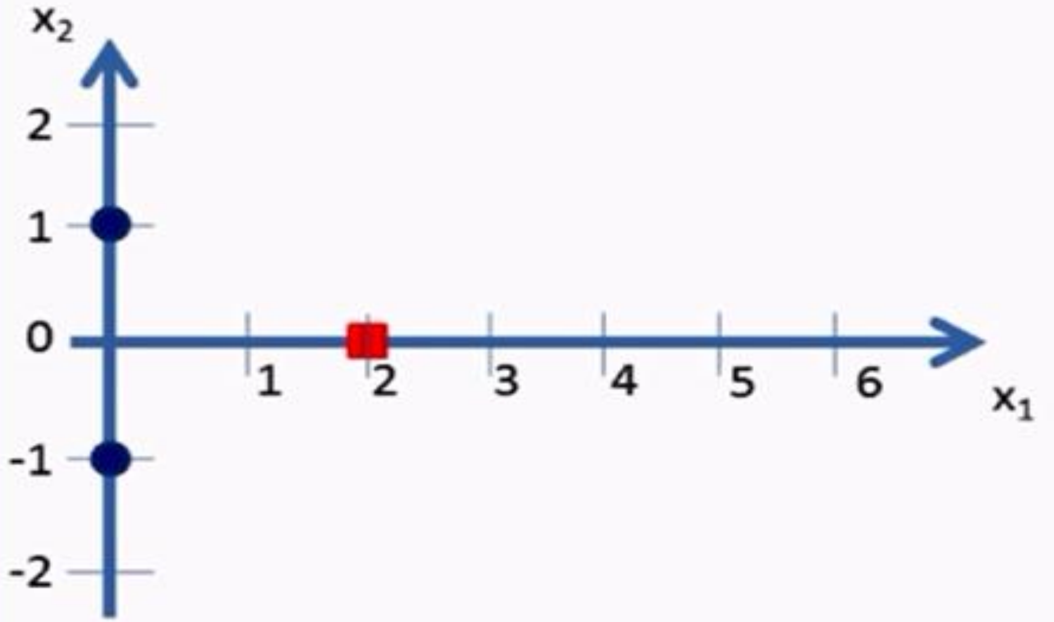
This is the expected decision surface of LSVM

PROGRAMMING EXAMPLE 1

Here,
Y defines the positive
 and negative class

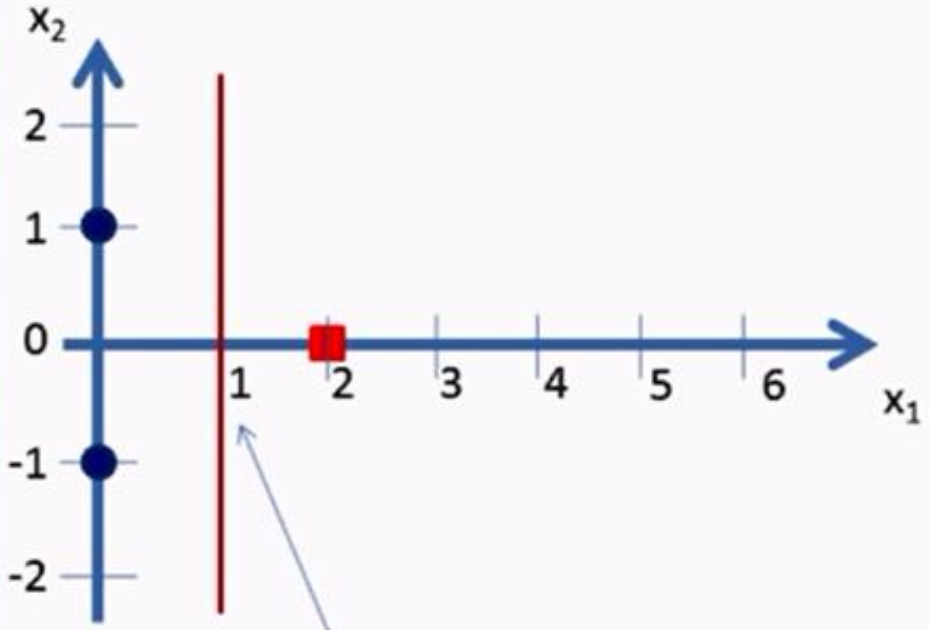
P,q,r are alphas

```
% 3 support vector version

s1 = [ 0 -1 1 ];
s2 = [ 0 1 1 ];
s3 = [ 2 0 1 ];

A = [ sum(s1.*s1) sum(s2.*s1) sum(s3.*s1) ;
sum(s1.*s2) sum(s2.*s2) sum(s3.*s2) ;
sum(s1.*s3) sum(s2.*s3) sum(s3.*s3) ]
Y = [ -1 -1 +1 ]
X = Y/A

p = X(1)
q = X(2)
r = X(3)

W = [ p*s1 + q*s2 + r*s3 ]
```



PROGRAMMING EXAMPLE 1

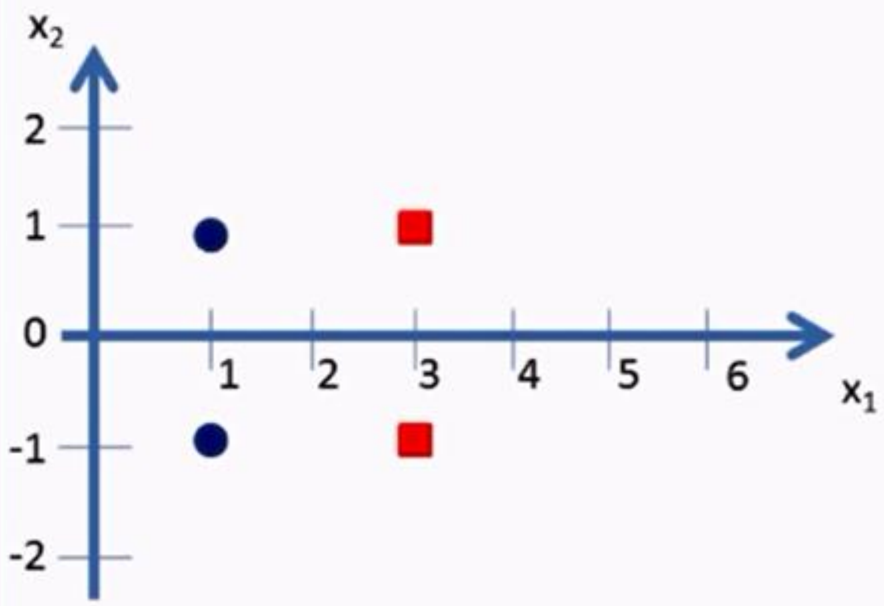
```
% 3 support vector version

s1 = [ 0 -1 1 ];
s2 = [ 0 1 1 ];
s3 = [ 2 0 1 ];

A = [ sum(s1.*s1) sum(s2.*s1) sum(s3.*s1) ;
sum(s1.*s2) sum(s2.*s2) sum(s3.*s2) ;
sum(s1.*s3) sum(s2.*s3) sum(s3.*s3) ]
Y = [ -1 -1 +1 ]
X = Y/A

p = X(1)
q = X(2)
r = X(3)

W = [ p*s1 + q*s2 + r*s3 ]
```



When you run you should get: $\tilde{w}$ = [1 0 -1]. This is a vertical line passing through x1=1.

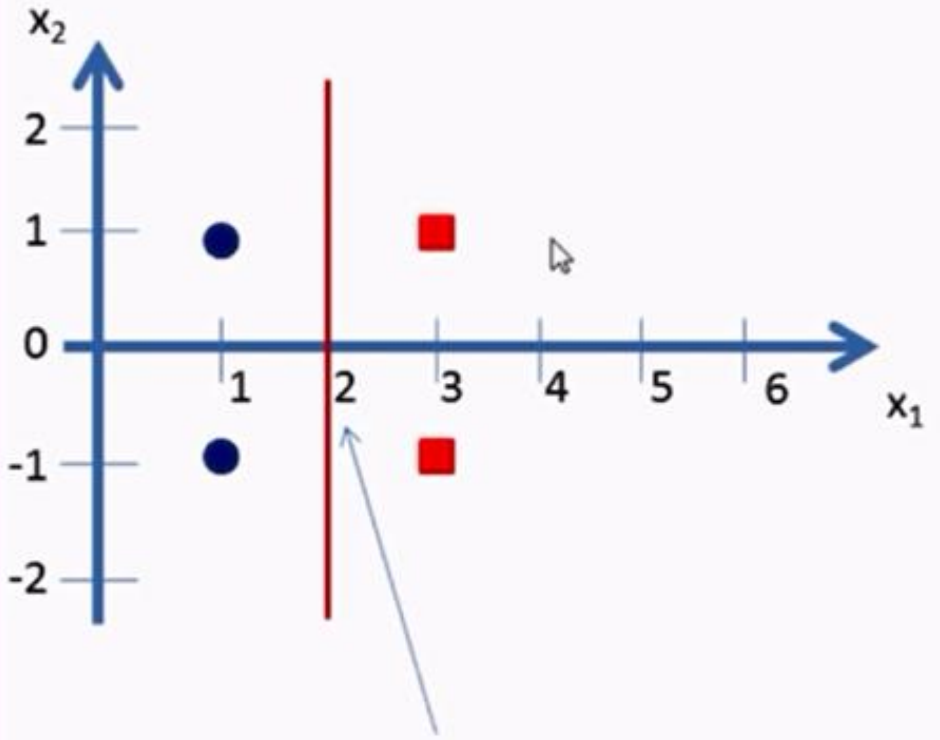
PROGRAMMING EXAMPLE 2

```
% 4 support vector version

s1 = [ 1 1 1 ];
s2 = [ 1 -1 1 ];
s3 = [ 3 -1 1 ];
s4 = [ 3 1 0 ];

A = [ sum(s1.*s1) sum(s2.*s1) sum(s3.*s1) sum(s4.*s1);
   sum(s1.*s2) sum(s2.*s2) sum(s3.*s2) sum(s4.*s1);
   sum(s1.*s3) sum(s2.*s3) sum(s3.*s3) sum(s4.*s3);
   sum(s1.*s4) sum(s2.*s4) sum(s3.*s4) sum(s4.*s4);]
Y = [ -1 -1 +1 +1 ]
X = Y/A

p = X(1)
q = X(2)
r = X(3)
s = X(4)

W = [ p*s1 + q*s2 + r*s3 + s*s4 ]
```



PROGRAMMING EXAMPLE 2

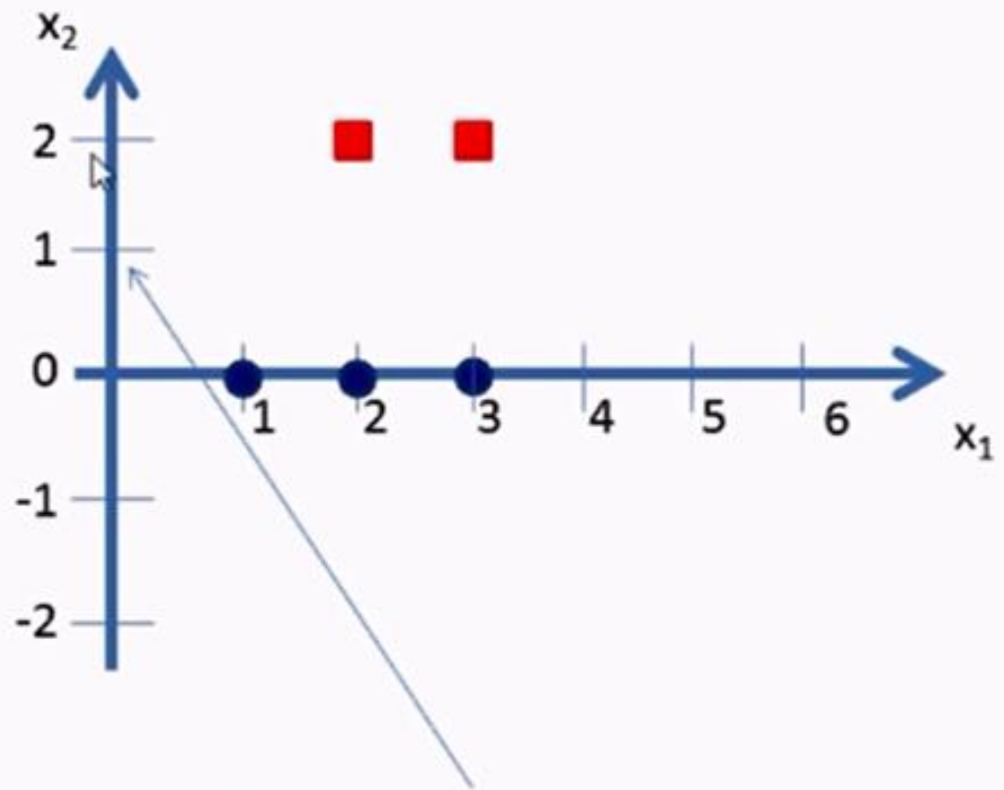
```
•   % 4 support vector version
•
•   s1 = [ 1 1 1 ];
•   s2 = [ 1 -1 1 ];
•   s3 = [ 3 -1 1 ];
•   s4 = [ 3 1 0 ];
•
•   A = [ sum(s1.*s1) sum(s2.*s1) sum(s3.*s1) sum(s4.*s1);
•      sum(s1.*s2) sum(s2.*s2) sum(s3.*s2) sum(s4.*s1);
•      sum(s1.*s3) sum(s2.*s3) sum(s3.*s3) sum(s4.*s3);
•      sum(s1.*s4) sum(s2.*s4) sum(s3.*s4) sum(s4.*s4);]
•   Y = [ -1 -1 +1 +1 ]
•   X = Y/A
•
•   p = X(1)
•   q = X(2)
•   r = X(3)
•   s = X(4)
•
•   W = [ p*s1 + q*s2 + r*s3 + s*s4 ]
```



When you run you should get: $\tilde{w}$ = [1 0 -2]. This is a vertical line passing through x1=2.

PROGRAMMING EXAMPLE 3

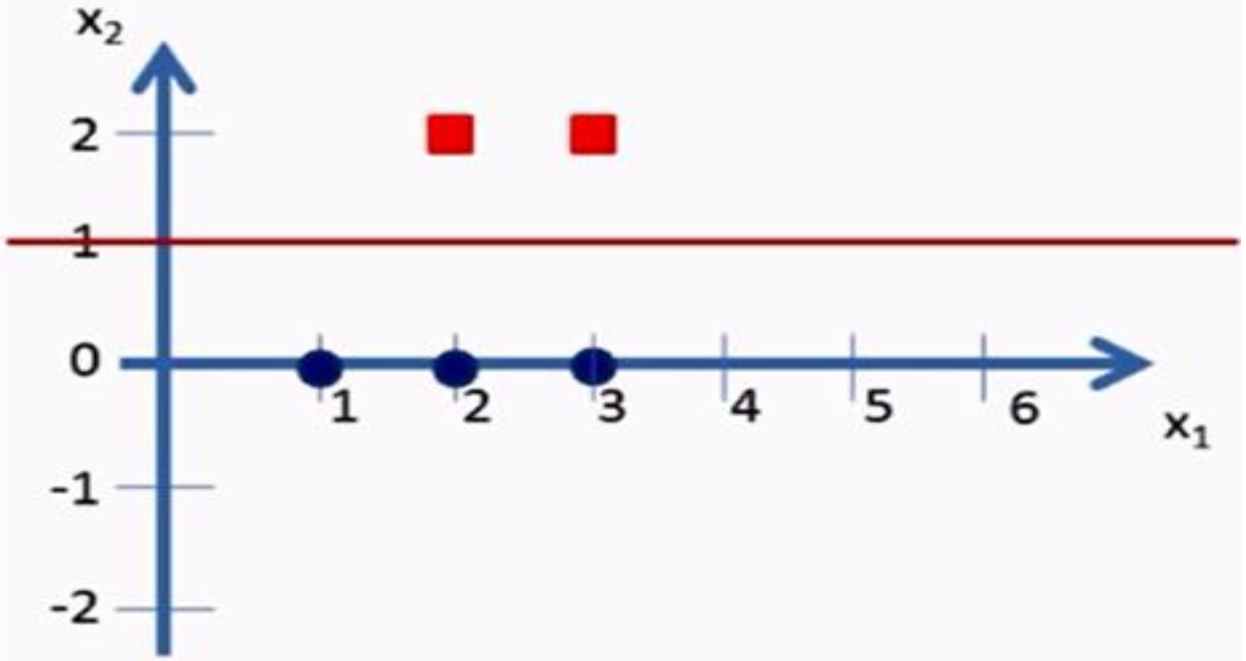
- % 5 support vector version
-
- s1 = [1 0 1];
- s2 = [2 0 1];
- s3 = [3 0 1];
- s4 = [2 2 1];
- s5 = [3 2 1];
-
- A = [sum(s1.*s1) sum(s2.*s1) sum(s3.*s1) sum(s4.*s1) sum(s5.*s1);
- sum(s1.*s2) sum(s2.*s2) sum(s3.*s2) sum(s4.*s2) sum(s5.*s2);
- sum(s1.*s3) sum(s2.*s3) sum(s3.*s3) sum(s4.*s3) sum(s5.*s3);
- sum(s1.*s4) sum(s2.*s4) sum(s3.*s4) sum(s4.*s4) sum(s5.*s4);
- sum(s1.*s5) sum(s2.*s5) sum(s3.*s5) sum(s4.*s5) sum(s5.*s5)]
- Y = [-1 -1 -1 +1 +1]
- X = Y/A
-
- p = X(1)
- q = X(2)
- r = X(3)
- s = X(4)
- t = X(5)
-
- W = [p*s1 + q*s2 + r*s3 + s*s4 + t*s5]



When you run you should get: $\widetilde{w}$ = [0 1 -1]. This is a horizontal line passing through x2=1.

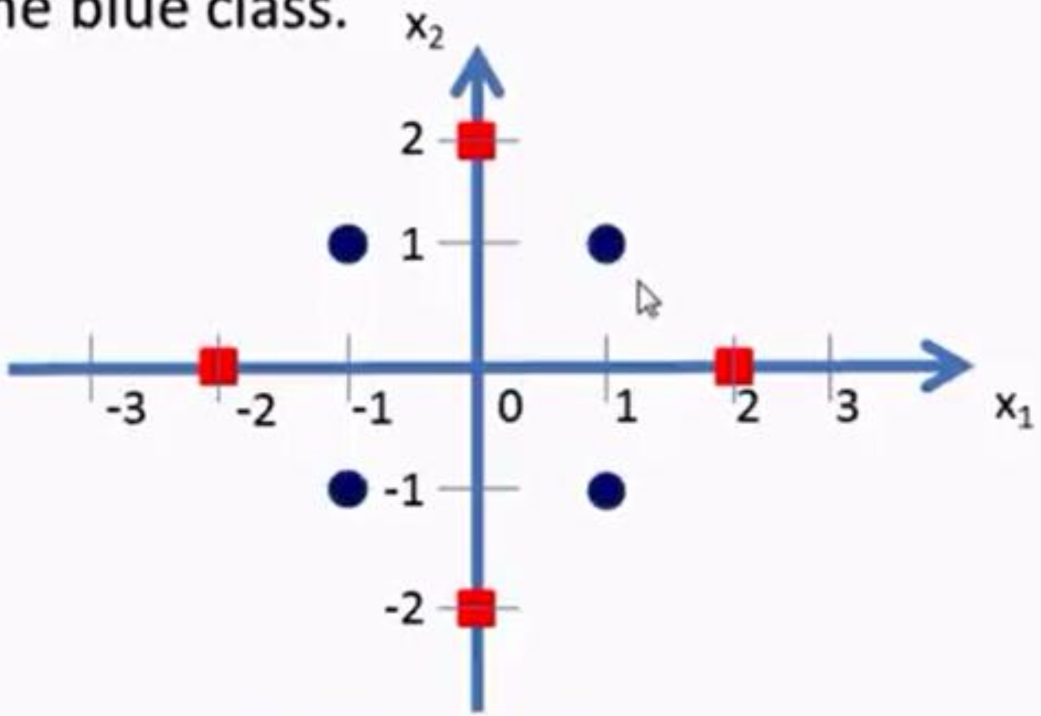
CLASSIFY UNKNOWN VECTORS

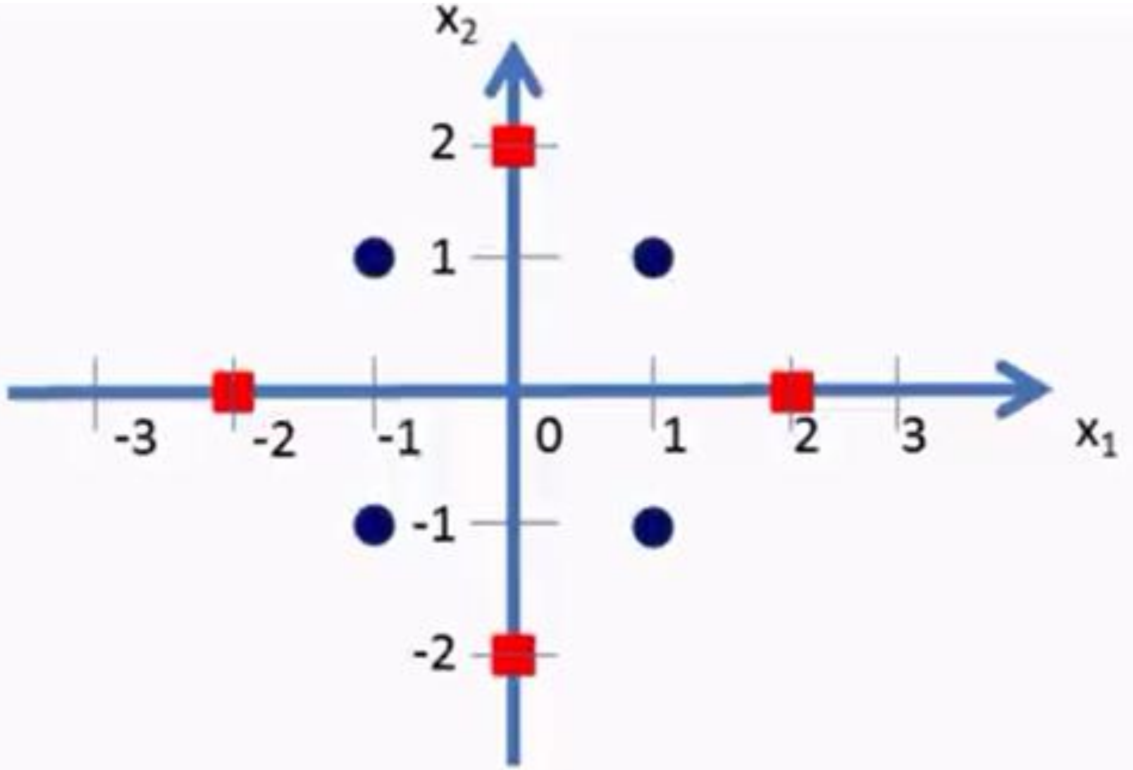
- Let's take the 5 support vector version
- $\tilde{w} = [\,0 \;\; 1 \;\; 1\,]$. This is a horizontal line passing through x2=1.
- Let's classify the point (x1,x2)=(4,2).
- $w.x = \binom{0}{1}.\binom{4}{2} = 2 > 1$
- Hence this point belongs to the red class
- Let's classify the point (x1,x2)=(2,-2).
- $w.x = \binom{0}{1}.\binom{2}{-2} = -2 < 1$
- Hence this point belongs to the blue class
- We can do the same for any new point.



NONLINEAR SUPPORT VECTOR MACHINES

Obviously there is no clear separating hyperplane between the red class and the blue class.





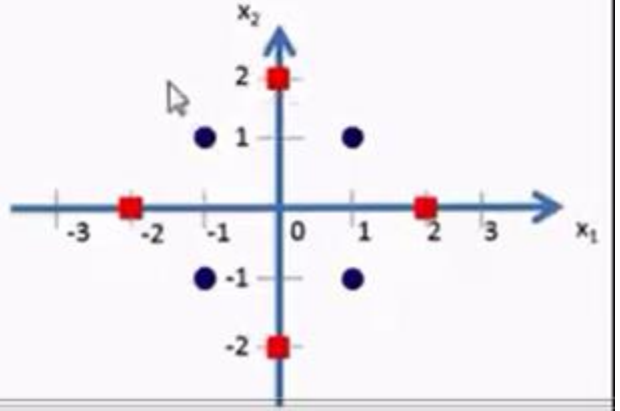
- Blue class vectors are: $\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}-1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}$
- Red class vectors are: $\begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}0\\2\end{pmatrix}, \begin{pmatrix}-2\\0\end{pmatrix}, \begin{pmatrix}0\\-2\end{pmatrix}$

NON – LINEAR SVM (FUNCTION)

We need to find a non-linear mapping function phi which can tranform these data into a new feature space where a separating hyperplane can be found

Let us consider the following mapping function.

$$\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{cases}\begin{pmatrix}6 - x_1 + (x_1 - x_2)^2\\6 - x_2 + (x_1 - x_2)^2\end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} \geq 2\\ \begin{pmatrix}x_1\\x_2\end{pmatrix} & \text{otherwise}\end{cases}$$



The points outside the blue class will be fransformed into a different feature space

FOR THE BLUE CLASS

Now let us transform the blue and red calss vectors using the non-linear mapping function Φ .

$$\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{cases}\begin{pmatrix}6 - x_1 + (x_1 - x_2)^2\\6 - x_2 + (x_1 - x_2)^2\end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} \geq 2\\ \begin{pmatrix}x_1\\x_2\end{pmatrix} & \text{otherwise}\end{cases}$$

Blue class vectors are: $\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix}, \begin{pmatrix}-1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}$ no change since $\sqrt{x_1^2 + x_2^2} < 2$ for all the vectors

FOR THE RED CLASS

$$\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{cases}\begin{pmatrix}6 - x_1 + (x_1 - x_2)^2\\6 - x_2 + (x_1 - x_2)^2\end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} \geq 2\\ \begin{pmatrix}x_1\\x_2\end{pmatrix} & \text{otherwise}\end{cases}$$

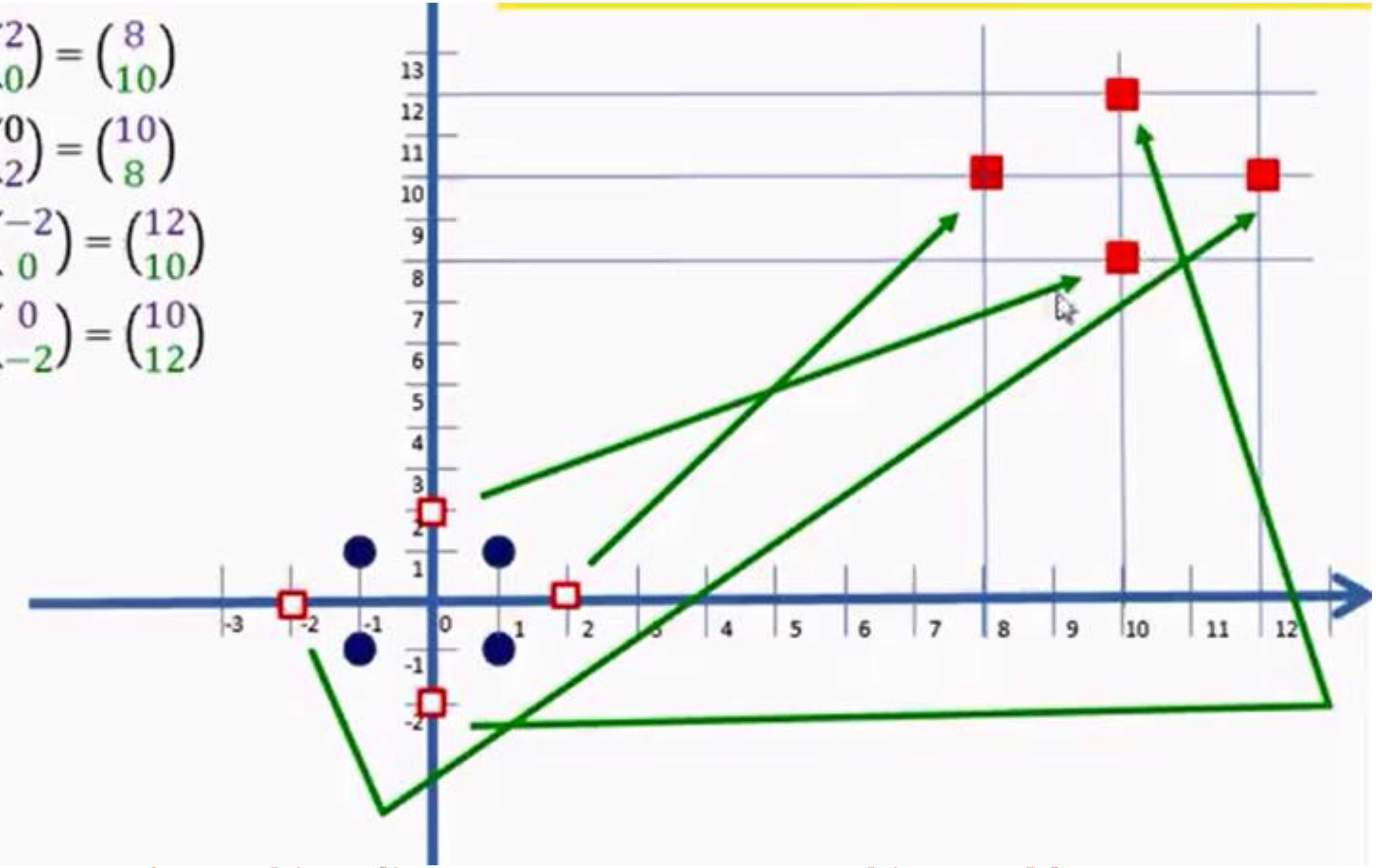
Let us take Red class vectors : $\begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}0\\2\end{pmatrix}, \begin{pmatrix}-2\\0\end{pmatrix}, \begin{pmatrix}0\\-2\end{pmatrix}$

FOR THE RED CLASS

- $\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{cases}\begin{pmatrix}6 - x_1 + (x_1 - x_2)^2\\6 - x_2 + (x_1 - x_2)^2\end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} \geq 2\\ \begin{pmatrix}x_1\\x_2\end{pmatrix} & \text{otherwise}\end{cases}$
- Let us take Red class vectors : $\begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}0\\2\end{pmatrix}, \begin{pmatrix}-2\\0\end{pmatrix}, \begin{pmatrix}0\\-2\end{pmatrix}$
- $\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \Phi\begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}6-2+(2-0)^2\\6-0+(2-0)^2\end{pmatrix} = \begin{pmatrix}8\\10\end{pmatrix}$
- $\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \Phi\begin{pmatrix}0\\2\end{pmatrix} = \begin{pmatrix}6-0+(0-2)^2\\6-2+(0-2)^2\end{pmatrix} = \begin{pmatrix}10\\8\end{pmatrix}$
- $\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \Phi\begin{pmatrix}-2\\0\end{pmatrix} = \begin{pmatrix}6+2+(-2-0)^2\\6-0+(-2-0)^2\end{pmatrix} = \begin{pmatrix}12\\10\end{pmatrix}$
- $\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \Phi\begin{pmatrix}0\\-2\end{pmatrix} = \begin{pmatrix}6-0+(0+2)^2\\6+2+(0+2)^2\end{pmatrix} = \begin{pmatrix}10\\12\end{pmatrix}$

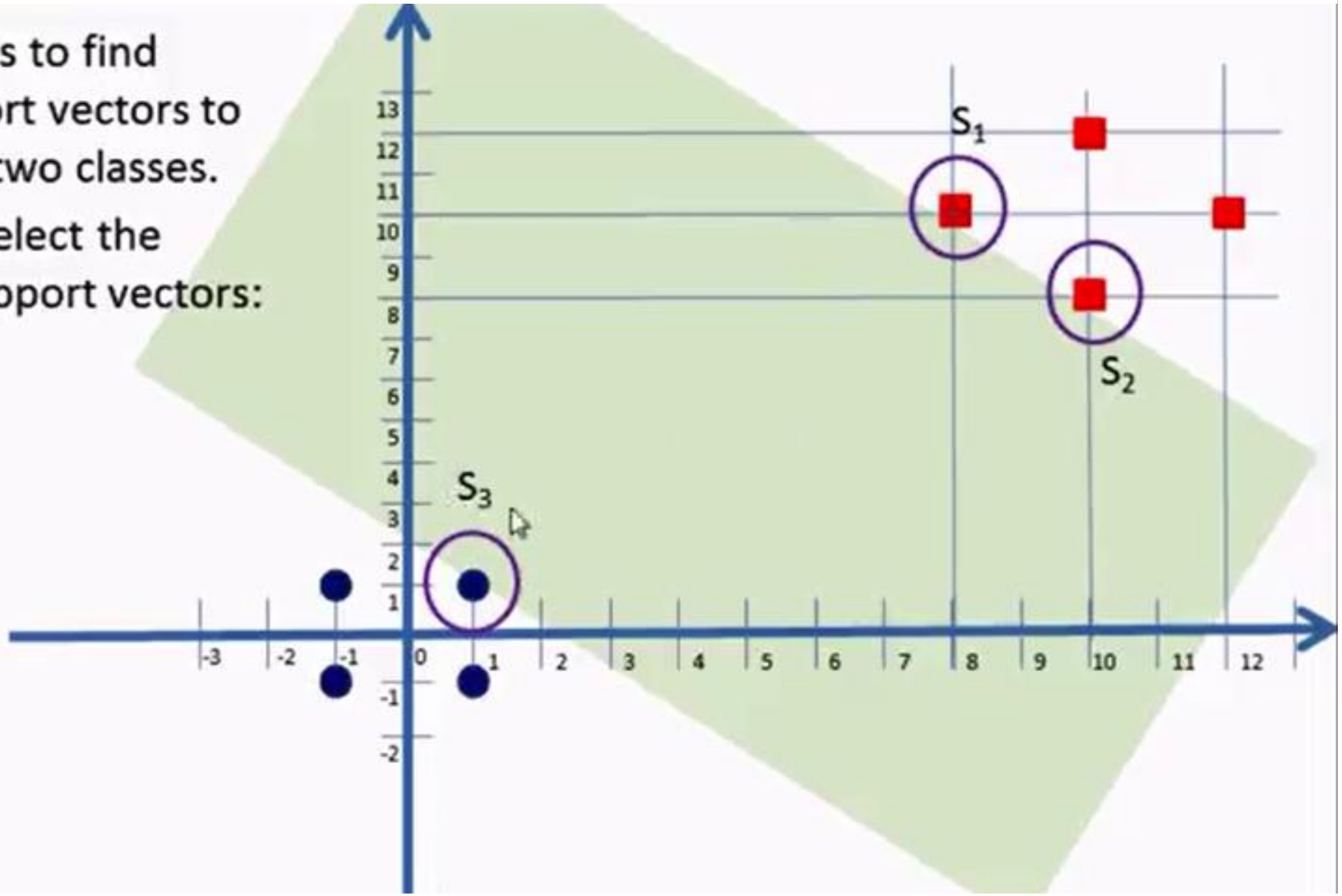
TRANSFORMATION OF THE RED CLASS

- $\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \Phi\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 10 \end{pmatrix}$
- $\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \Phi\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 10 \\ 8 \end{pmatrix}$
- $\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \Phi\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \end{pmatrix}$
- $\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \Phi\begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \end{pmatrix}$



Transformed into linear support vector machine problem

- Now our task is to find suitable support vectors to classify these two classes.
- Here we will select the following 3 support vectors:
- $S_1 = \binom{8}{10}$,
- $S_2 = \binom{10}{8}$,
- and $S_3 = \binom{1}{1}$



Here we will use vectors augmented with a 1 as a bias input, and for clarity we will differentiate these with an over-tilde. That is:

$$S_1 = \begin{pmatrix} 8 \\ 10 \end{pmatrix} \qquad \widetilde{S_1} = \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 10 \\ 8 \end{pmatrix} \qquad \widetilde{S_2} = \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix}$$

$$S_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \widetilde{S_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Now we need to find 3 parameters α_1, α_2, and α_3 based on the following 3 linear equations:

$$\alpha_1 \widetilde{S_1}.\widetilde{S_1} + \alpha_2 \widetilde{S_2}.\widetilde{S_1} + \alpha_3 \widetilde{S_3}.\widetilde{S_1} = +1 \ (+ve\ class)$$

$$\alpha_1 \widetilde{S_1}.\widetilde{S_2} + \alpha_2 \widetilde{S_2}.\widetilde{S_2} + \alpha_3 \widetilde{S_3}.\widetilde{S_2} = +1 \ (+ve\ class)$$

$$\alpha_1 \widetilde{S_1}.\widetilde{S_3} + \alpha_2 \widetilde{S_2}.\widetilde{S_3} + \alpha_3 \widetilde{S_3}.\widetilde{S_3} = -1 \ (-ve\ class)$$

$$\alpha_1 \tilde{S_1}.\tilde{S_1} + \alpha_2 \tilde{S_2}.\tilde{S_1} + \alpha_3 \tilde{S_3}.\tilde{S_1} = +1 \quad (+ve\ class)$$

$$\alpha_1 \tilde{S_1}.\tilde{S_2} + \alpha_2 \tilde{S_2}.\tilde{S_2} + \alpha_3 \tilde{S_3}.\tilde{S_2} = +1 \quad (+ve\ class)$$

$$\alpha_1 \tilde{S_1}.\tilde{S_3} + \alpha_2 \tilde{S_2}.\tilde{S_3} + \alpha_3 \tilde{S_3}.\tilde{S_3} = -1 \quad (-ve\ class)$$

Let's substitute the values for $\tilde{S}_1$, $\tilde{S}_2$ and $\tilde{S}_3$ in the above equations.

$$\tilde{S_1} = \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} \quad \tilde{S_2} = \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} \quad \tilde{S_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} . \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} . \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} = +1$$

$$\alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} . \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} . \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} = +1$$

$$\alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$$

SIMULTANEOUS EQUATIONS

$$\alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} . \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} . \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} = +1$$

$$\alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} . \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} . \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} = +1$$

$$\alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$$

After multiplication we get:

$$165\,\alpha_1 + 161\,\alpha_2 + 19\,\alpha_3 = +1$$

$$161\,\alpha_1 + 165\,\alpha_2 + 19\,\alpha_3 = +1$$

$$19\,\alpha_1 + 19\,\alpha_2 + 3\,\alpha_3 = -1$$

SIMULTANEOUS EQUATIONS

$$\alpha_1\begin{pmatrix}8\\10\\1\end{pmatrix}.\begin{pmatrix}8\\10\\1\end{pmatrix}+\alpha_2\begin{pmatrix}10\\8\\1\end{pmatrix}.\begin{pmatrix}8\\10\\1\end{pmatrix}+\alpha_3\begin{pmatrix}1\\1\\1\end{pmatrix}.\begin{pmatrix}8\\10\\1\end{pmatrix}=+1$$

$$\alpha_1\begin{pmatrix}8\\10\\1\end{pmatrix}.\begin{pmatrix}10\\8\\1\end{pmatrix}+\alpha_2\begin{pmatrix}10\\8\\1\end{pmatrix}.\begin{pmatrix}10\\8\\1\end{pmatrix}+\alpha_3\begin{pmatrix}1\\1\\1\end{pmatrix}.\begin{pmatrix}10\\8\\1\end{pmatrix}=+1$$

$$\alpha_1\begin{pmatrix}8\\10\\1\end{pmatrix}.\begin{pmatrix}1\\1\\1\end{pmatrix}+\alpha_2\begin{pmatrix}10\\8\\1\end{pmatrix}.\begin{pmatrix}1\\1\\1\end{pmatrix}+\alpha_3\begin{pmatrix}1\\1\\1\end{pmatrix}.\begin{pmatrix}1\\1\\1\end{pmatrix}=-1$$

After multiplication we get:

$$165\,\alpha_1+161\,\alpha_2+19\,\alpha_3=+1$$

$$161\,\alpha_1+165\,\alpha_2+19\,\alpha_3=+1$$

$$19\,\alpha_1+19\,\alpha_2+3\,\alpha_3=-1$$

Simplifying the above 3 simultaneous equations we get: α_1=α_2= 0.859 and α_3= -1.4219.

DISCRIMINATING HYPERPLANE

The hyper plane that discriminates the positive class from the negative class is given by:

$$\tilde{w} = \sum_i \alpha_i \tilde{S}_i$$

Substituting the values we get:

$$\tilde{w} = \alpha_1 \begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\tilde{w} = (0.0859).\begin{pmatrix} 8 \\ 10 \\ 1 \end{pmatrix} + (0.0859).\begin{pmatrix} 10 \\ 8 \\ 1 \end{pmatrix} + (-1.4219).\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.1243 \\ 0.1243 \\ -1.2501 \end{pmatrix}$$

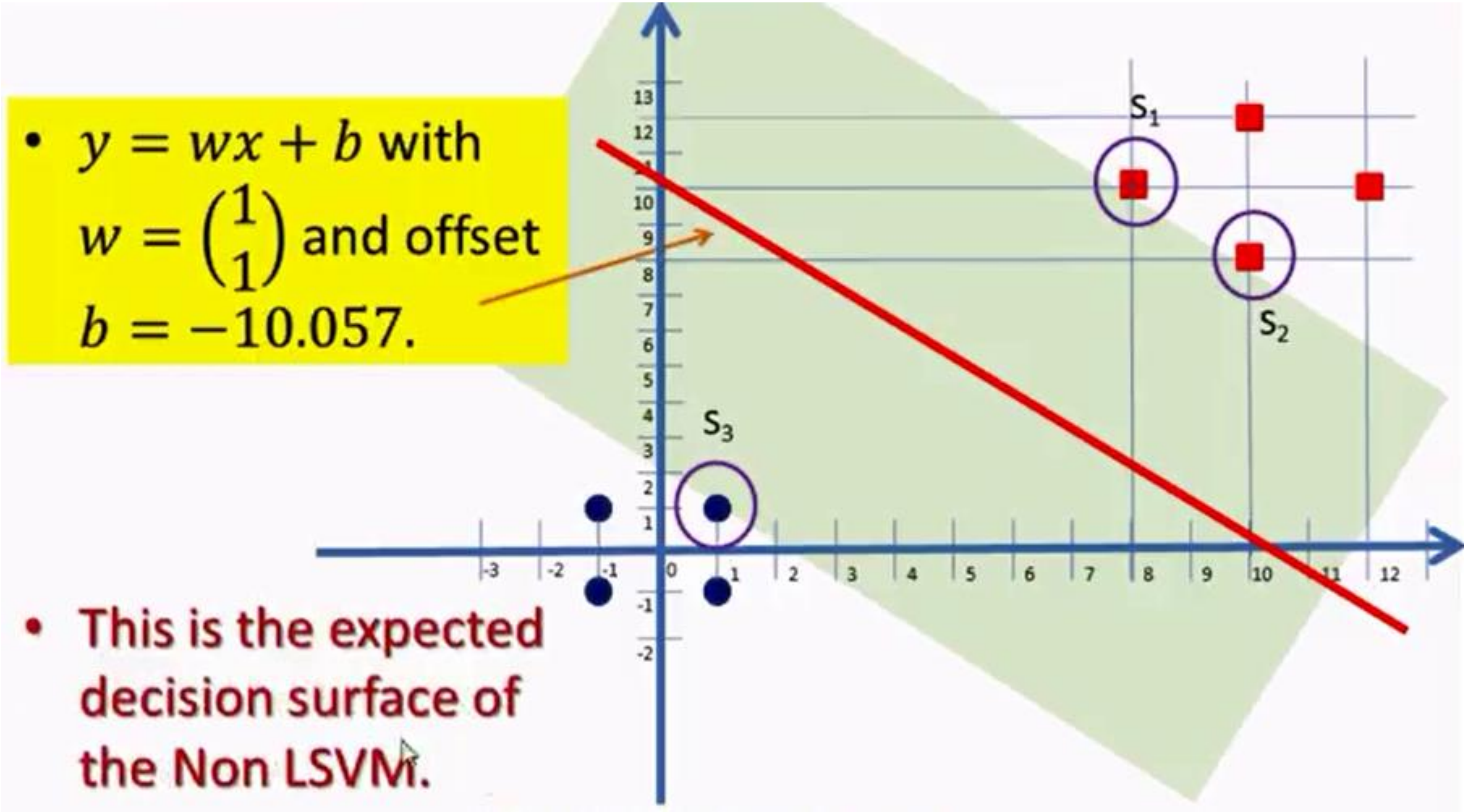
EQUATION FOR DISCRIMINATING HYPERPLANE

- Our vectors are augmented with a bias.
- Hence we can equate the entry in $\tilde{w}$ as the hyper plane with an offset b.
- Therefore the separating hyper plane equation

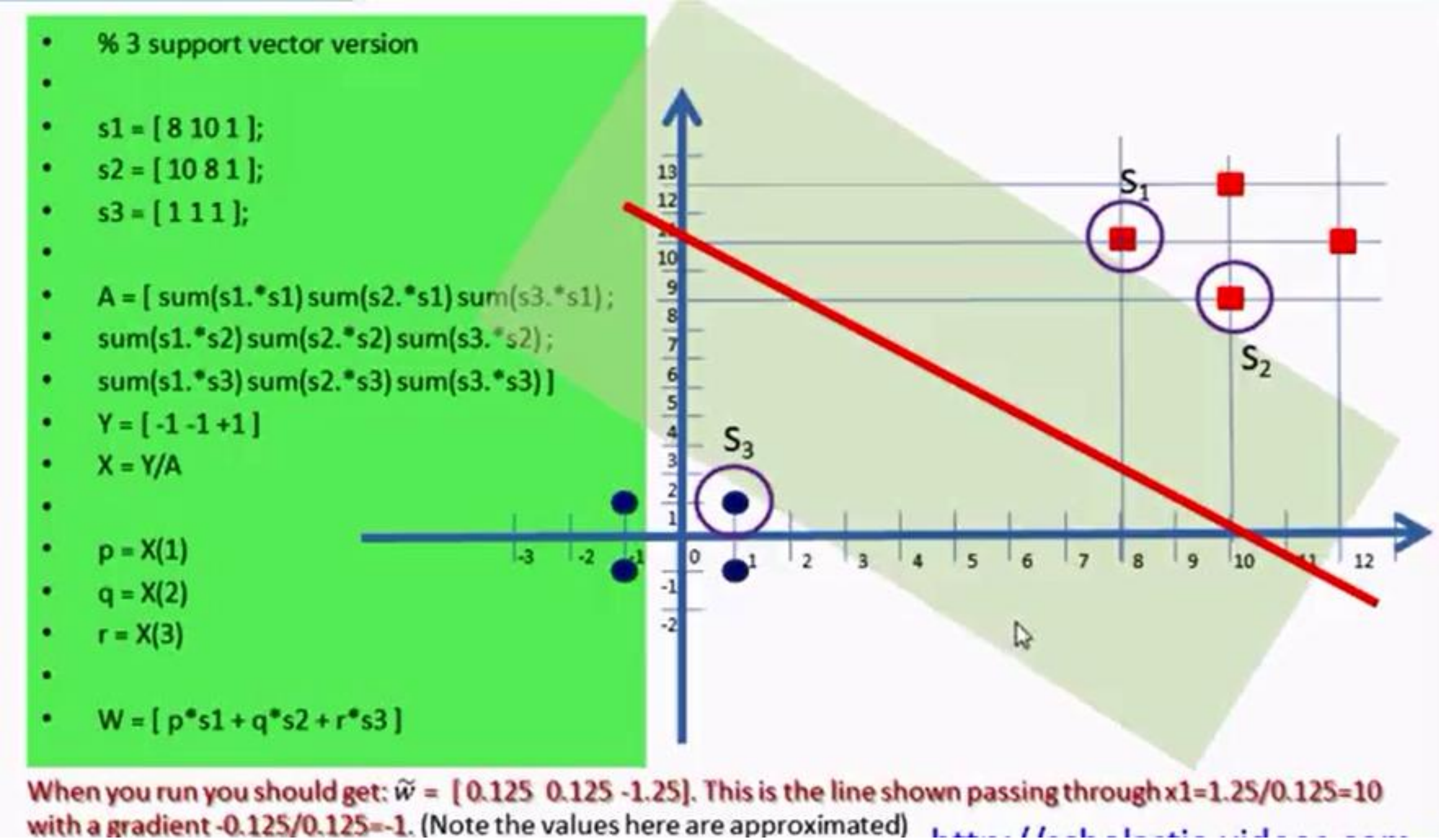
$y = wx + b$ with $w = \begin{pmatrix} 0.1243/0.1243 \\ 0.1243/0.1243 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

and an offset $b = -\frac{1.2501}{0.1243} = -10.057.$

HYPERPLANE SEPARATING THE TWO SURFACE



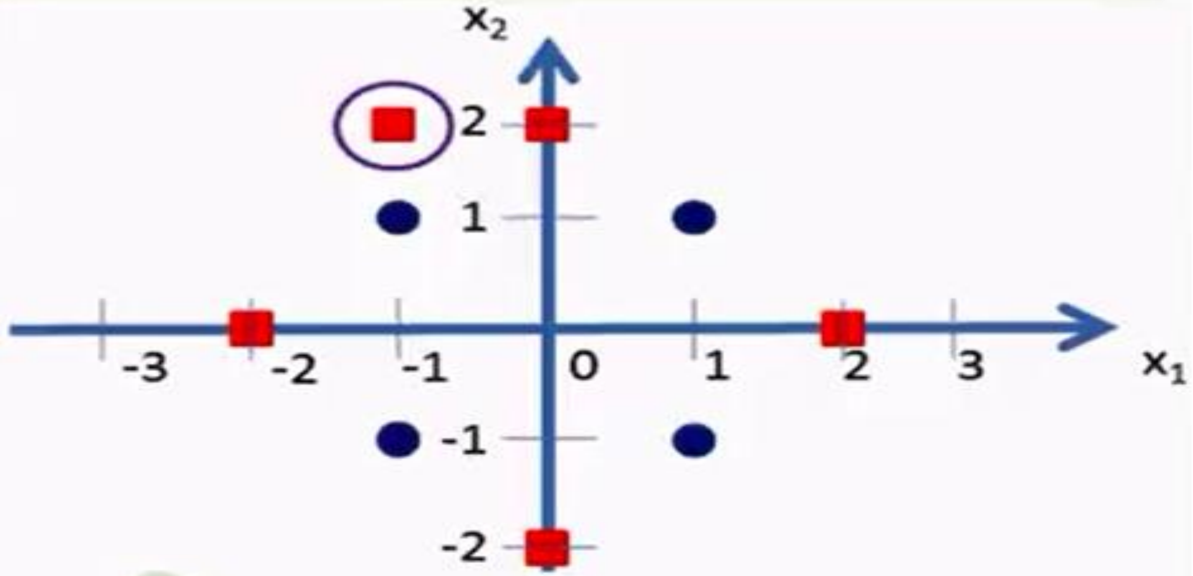
PROGRAMMING EXAMPLE 1

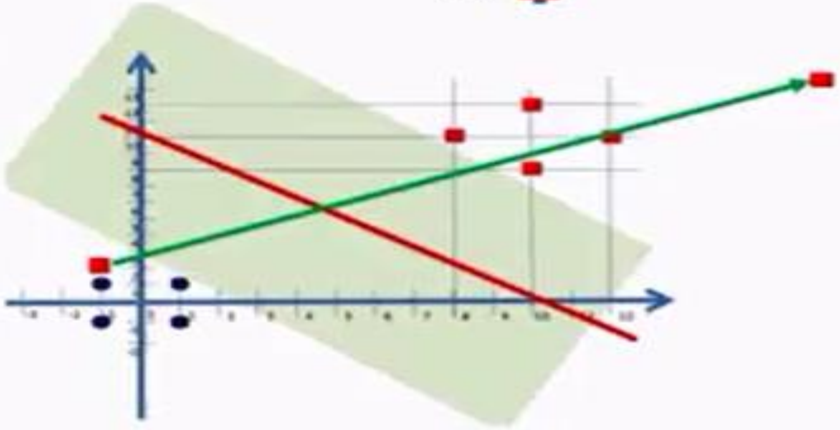


NLSVM - EXAMPLE2

$$\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{cases}\begin{pmatrix}6 - x_1 + (x_1 - x_2)^2\\6 - x_2 + (x_1 - x_2)^2\end{pmatrix} & \text{if } \sqrt{x_1^2 + x_2^2} \geq 2\\ \begin{pmatrix}x_1\\x_2\end{pmatrix} & \text{otherwise}\end{cases}$$

- Let's consider a classification example here.
- Let's classify the point $(x_1,x_2)=(-1,2)$.
- $\Phi\begin{pmatrix}x_1\\x_2\end{pmatrix} = \Phi\begin{pmatrix}-1\\2\end{pmatrix} = \begin{pmatrix}6+1+(-1-2)^2\\6-2+(-1-2)^2\end{pmatrix} = \begin{pmatrix}16\\13\end{pmatrix}$
- $w.\Phi(x) = \begin{pmatrix}1\\1\end{pmatrix}.\begin{pmatrix}16\\13\end{pmatrix} = 29 > 10$
- Hence this point belongs to the red class.





From above illustration, there are many linear classifiers (hyper planes) that separate the data. However only one of these achieves maximum separation. The reason we need it is because if we use a hyper plane to classify, it might end up closer to one set of datasets compared to others and we do not want this to happen and thus we see that the concept of maximum margin classifier or hyper plane as an apparent solution. The next illustration

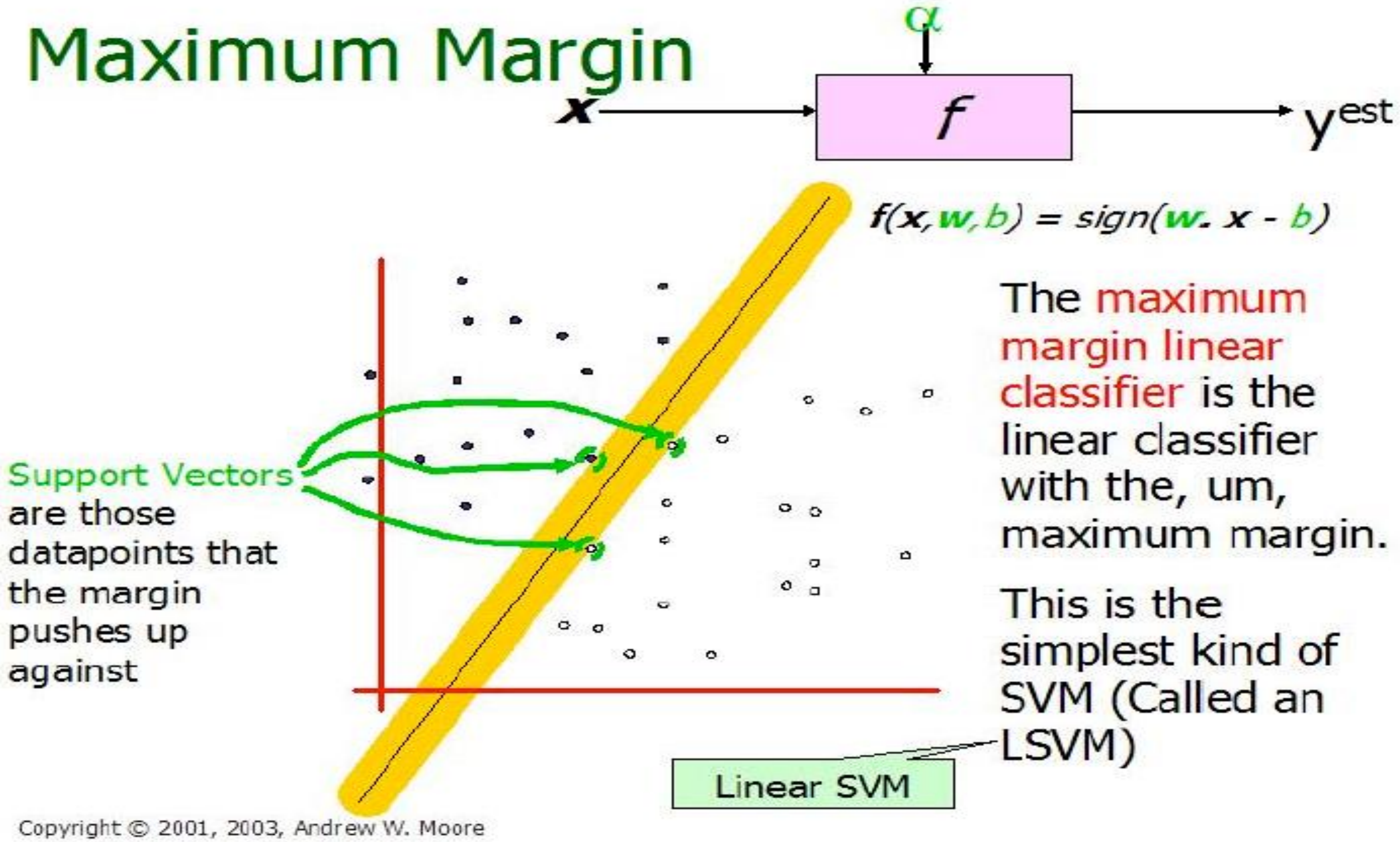


Figure 4: Illustration of Linear SVM. (Taken from Andrew W. Moore slides 2003) [2]. Note the legend is not described as they are sample plotting to make understand the concepts involved.