Support Vector Machines (SVM)

Mitul Shah (CE - 104)
Mohit Shah (CE - 105)

Outline

- Introduction
- Motivation
- How it works?
- Problems
- Kernel function
- Applications
- Advantages
- Disadvantages
- References

Introduction

- A support vector machine (SVM) is a computer algorithm that learns by example to assign labels to objects.
- In simple words we can say that SVM is a classification technique.
- It is a kind of supervised learning method that analyze data and recognizes pattern.

Introduction

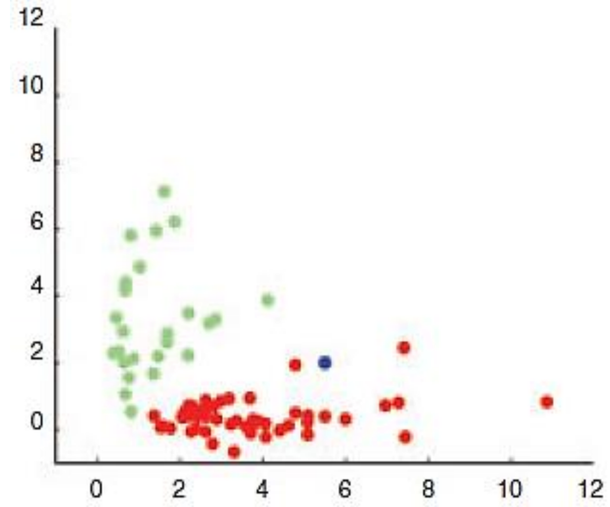
- We have to use training set to first train our machine then after it can predict the next unseen pattern weather it belongs to one or other class.
- Some applications are like an SVM can learn to recognize handwritten digits by examining a large collection of scanned images of handwritten zeroes, ones and so forth.

Motivation

- SVM performs well in practical applications.
- Can be used in text classification/recognition.
- When using with images it gives better accuracy compare to neural network.
- Also can be used in regression.

How it works?

- Consider the following figure.
- Here green and red dots represent different classes and blue is the one for which we have to identify to which it belongs to.

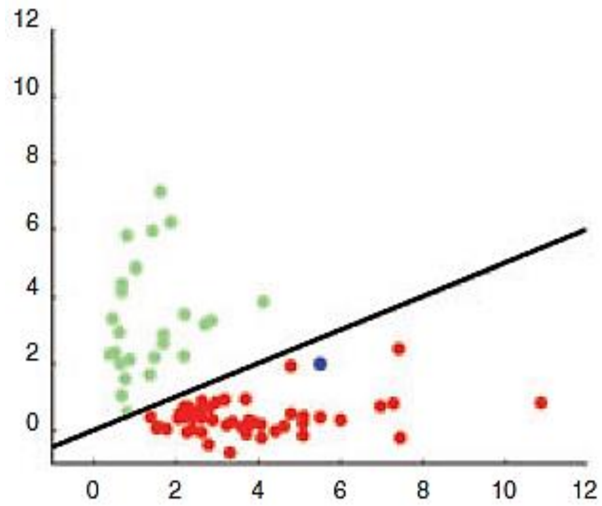


How it works?

- Human eye is very good at pattern recognition. By just looking once the figure we can find different clustering.
- But for machines we can use equation like if Y=2*X then it belongs to class 1(Green) and vice versa.

How it works?

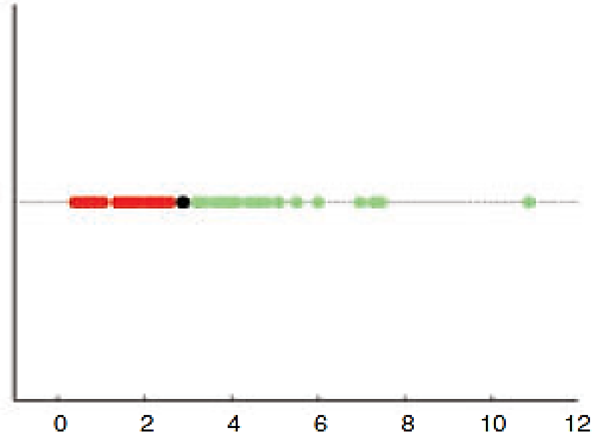
- Here we can draw a line which separates this two classes.
- This works for only for 2D.
- For 3D and 1D next slide shows an exp. of each.

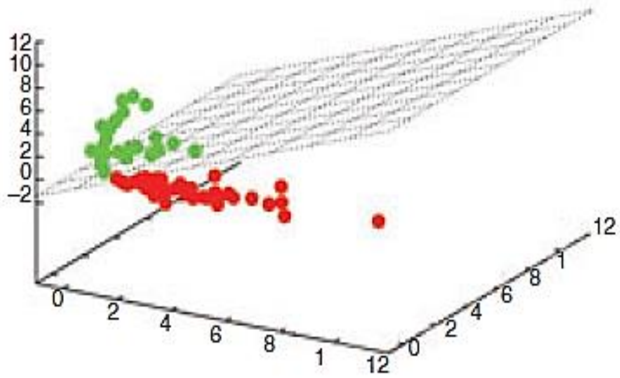


How it works?

1D

3D





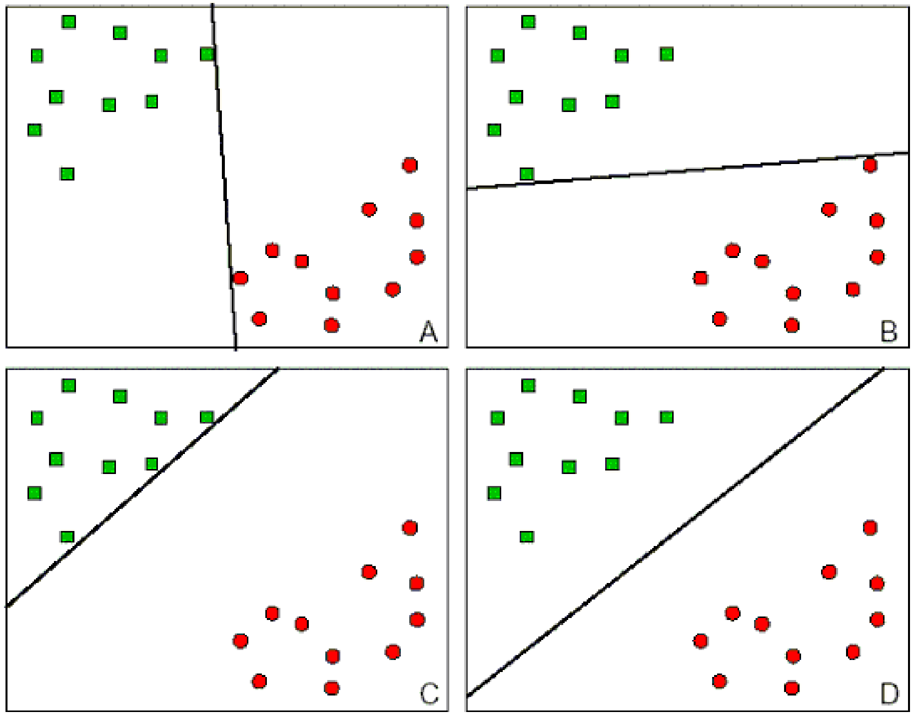
Dot

Hyperplane

How it works?

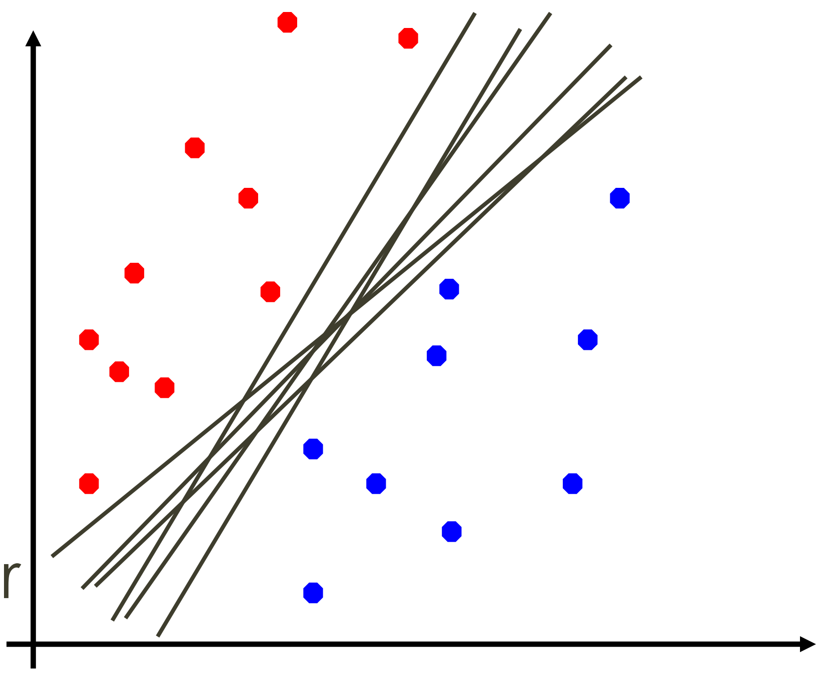
- For 2D as shown we can separate two classes by a single line like wise for 1D we can use a single dot and for 3D we have to use a "hyperplane" to separate two different classes.
- Now let's look at some problems regarding this technique.

Problems



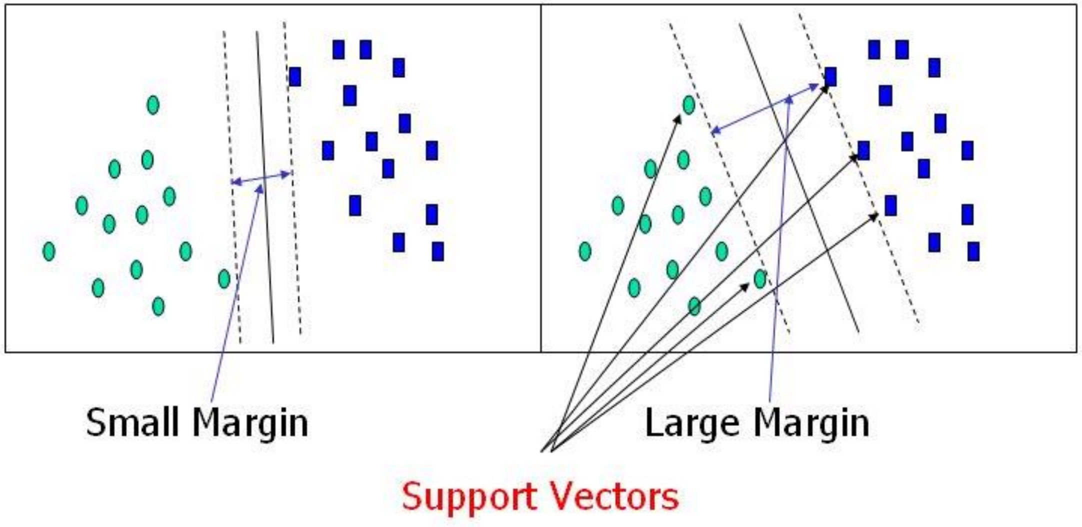
Problems

- As you can see here many possible lines can be drawn between this two classes.
- The problem is which one to consider for portioning.



Problems

- The solution to this problem is given as choose the line that has maximum margin from both class points.

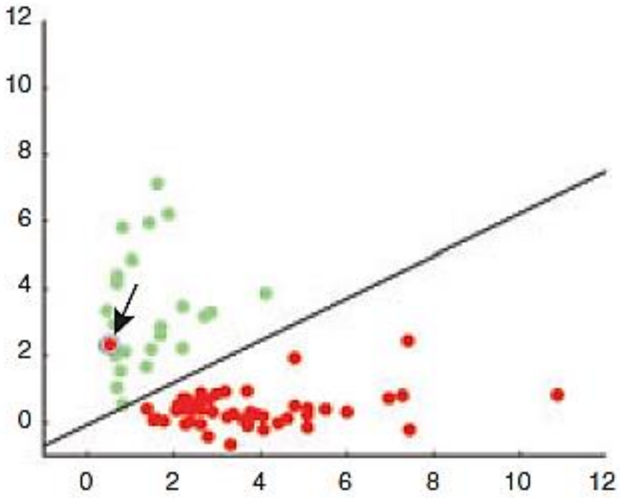


Problems

- The reason for choosing such line is obvious that it is more robust to outliners and do it has strong generalization ability.

Problems

- The second kind of problem is because of outliners. As many practical data cant be separated cleanly.
- To come up with this kind of problem we have a technique called "Soft Margin".

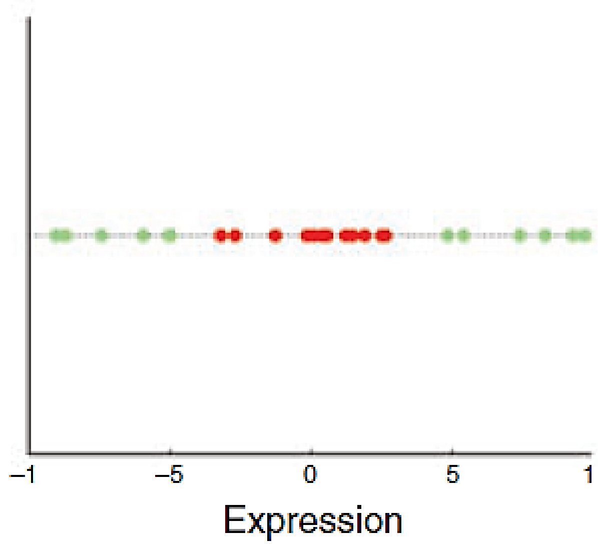


Problems

- This allows some data points to push their way through the margin of the separating hyperplane without affecting the final result.
- Of course, we don't want the SVM to allow for too many misclassifications.
- This value is generally user dependent which defines how many observations are allowed to violate the separating hyperplane.

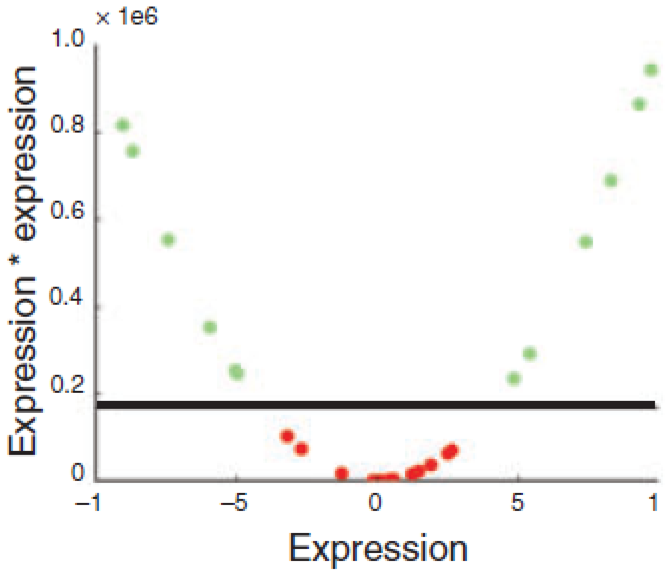
Kernel function

- To understand this let's first consider an example of classification based on only one value(1D).
- Here as we can see there is no way to place a single dot such that it separates two different classes.



Kernel function

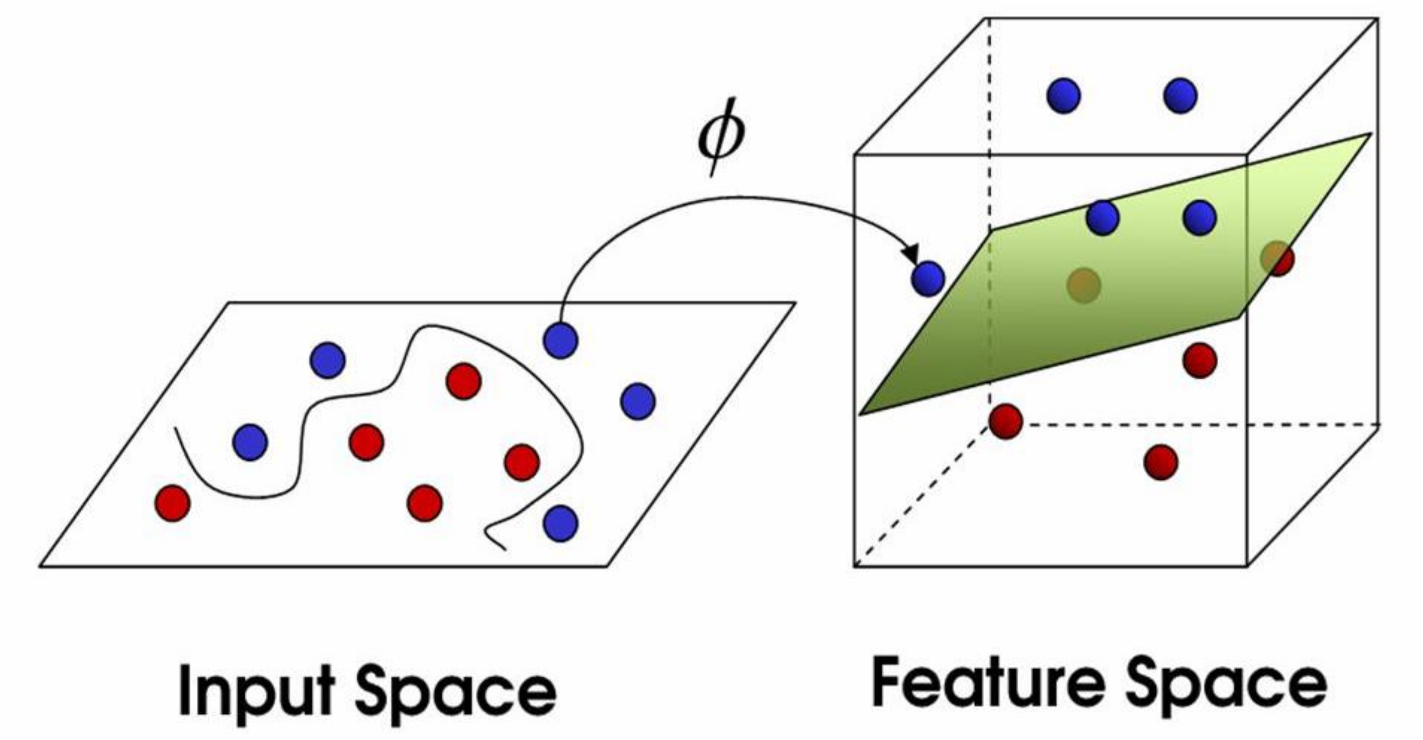
- To solve this kind of problem we have to use kernel function like here if we use square than we will get some diagram like:



Kernel function

- Now as we see this function simply transforms 1D data to 2D. And we can then classify data easily.
- In general, a kernel function projects data from a low-dimensional space to a space of higher dimension.
- If proper function is chosen than data will be separable in higher dimension space.

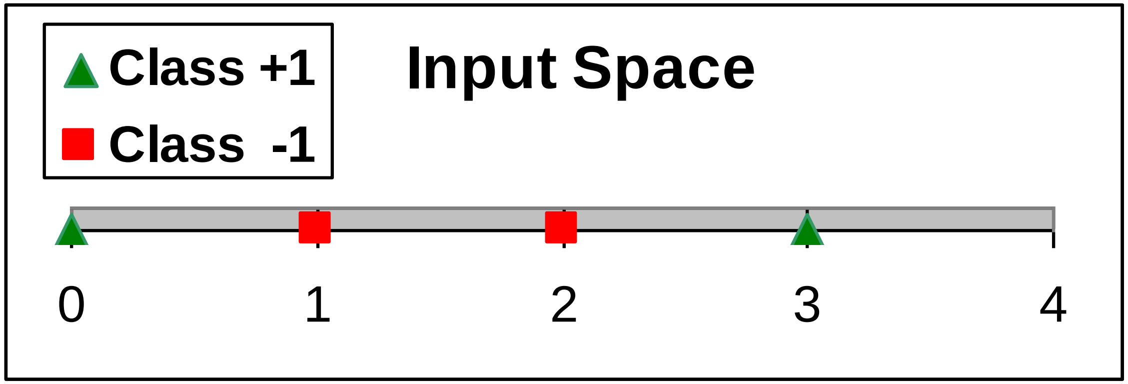
Kernel function



Simple SVM Example

Class	X_1
+1	0
-1	1
-1	2
+1	3

- How would SVM separates these points?
  - use the kernel trick
    - $\Phi(X_1) = (X_1, X_1^2)$
    - It becomes 2D.

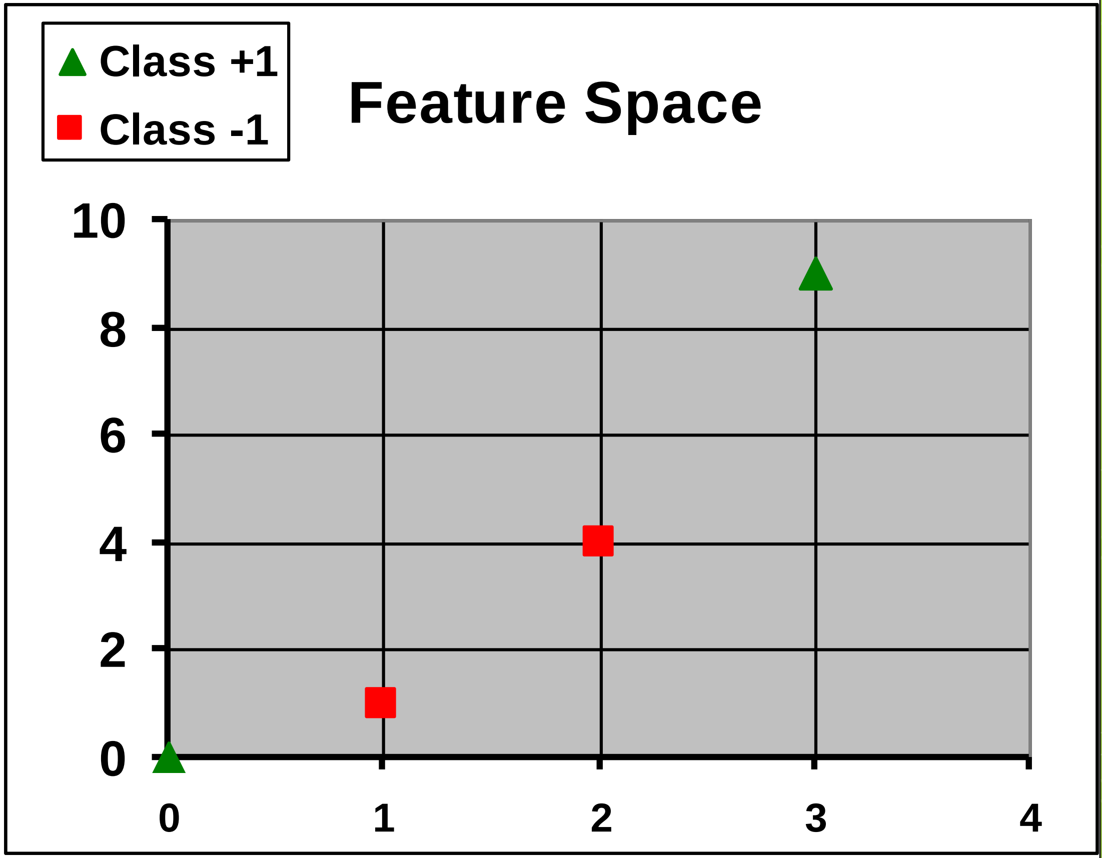


Source: Author's Research

Simple Points in Feature Space

Class	X_1	X_1^2
+1	0	0
-1	1	1
-1	2	4
+1	3	9

All points here are support vectors.



SVM Calculation

- Positive : $\langle \mathbf{w} \bullet \mathbf{x} \rangle + b = +1$
- Negative : $\langle \mathbf{w} \bullet \mathbf{x} \rangle + b = -1$
- Hyperplane : $\langle \mathbf{w} \bullet \mathbf{x} \rangle + b = 0$
- find the unknowns, **w** and **b**
- Expending the equations:
  - $w_1x_1 + w_2x_2 + b = +1$
  - $w_1x_1 + w_2x_2 + b = -1$
  - $w_1x_1 + w_2x_2 + b = 0$

Use Linear Algebra to Solve **w** and b

- $w_1x_1 + w_2x_2 + b = +1$

  $w_1 0 + w_2 0 + b = +1$

  $w_1 3 + w_2 9 + b = +1$

- $w_1x_1 + w_2x_2 + b = -1$

  $w_1 1 + w_2 1 + b = -1$

  $w_1 2 + w_2 4 + b = -1$

- Solution is $w_1 = -3$, $w_2 = 1$, $b = 1$
- SVM algorithm can find the solution that returns a Hyperplane with the largest margin

Use Solutions to Draw the Planes

Positive Plane:	Negative Plane:	Hyperplane:
$\langle \mathbf{w} \bullet \mathbf{x} \rangle + b = +1$	$\langle \mathbf{w} \bullet \mathbf{x} \rangle + b = -1$	$\langle \mathbf{w} \bullet \mathbf{x} \rangle + b = 0$
$w_1x_1 + w_2x_2 + b = +1$	$w_1x_1 + w_2x_2 + b = -1$	$w_1x_1 + w_2x_2 + b = 0$
➔ $-3x_1 + 1x_2 + 1 = +1$	➔ $-3x_1 + 1x_2 + 1 = -1$	➔ $-3x_1 + 1x_2 + 1 = 0$
➔ $x_2 = 3x_1$	➔ $x_2 = -2 + 3x_1$	➔ $x_2 = -1 + 3x_1$

Positive Plane:

X_1	X_2
0	0
1	3
2	6
3	9

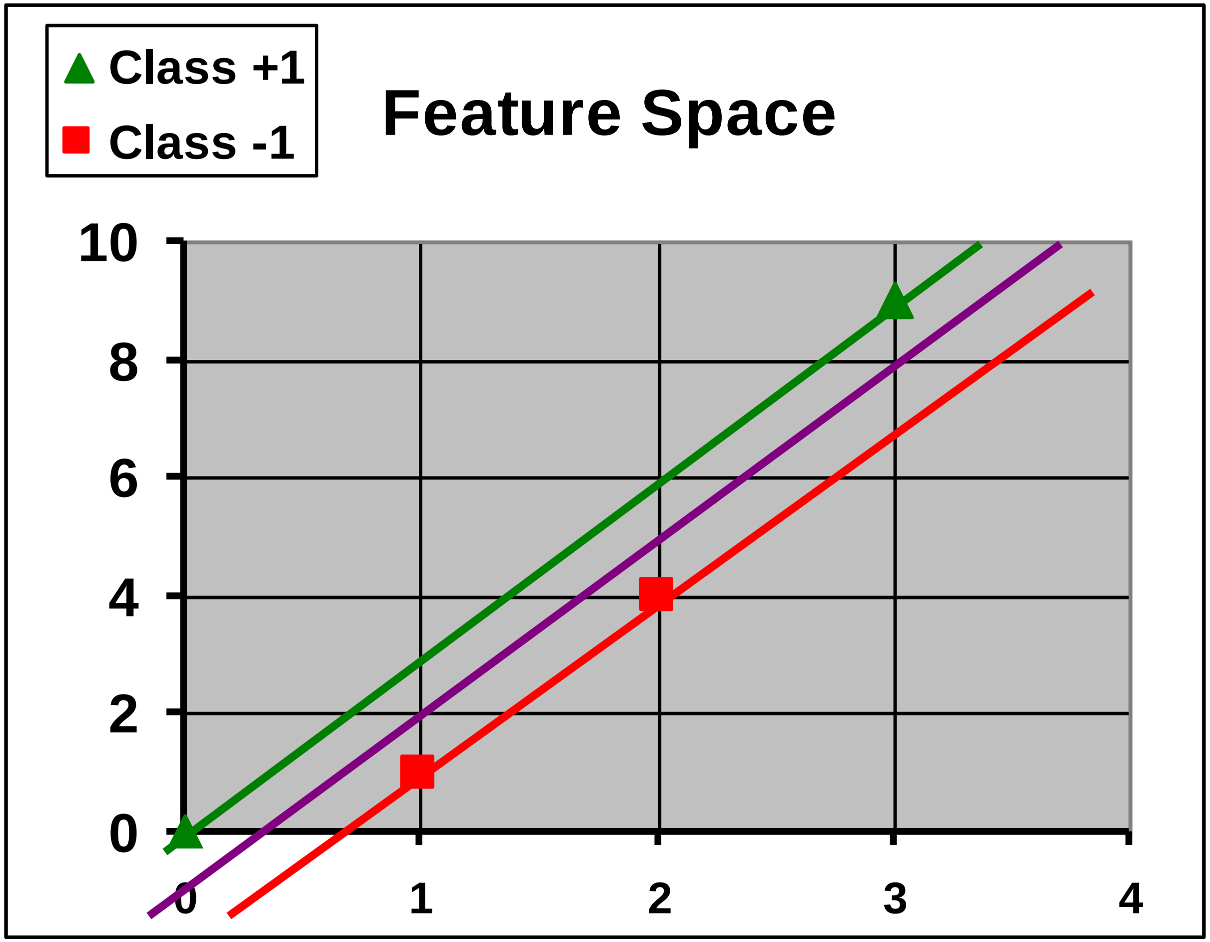
Negative Plane:

X_1	X_2
0	-2
1	1
2	4
3	7

Hyperplane:

X_1	X_2
0	-1
1	2
2	5
3	8

Simple Data Separated by a Line



SVM Applications

- SVM has been used successfully in many real-world problems

  - text (and hypertext) categorization
  - image classification
  - bioinformatics (Protein classification, Cancer classification)
  - hand-written character recognition
  - Gene Expression Data Classification
  - Face Detection and Face Recognition

Advantages

- The absence of local minima.
- Is more accurate when applied to real life problems.
- SVMs deliver a unique solution. This is an advantage compared to Neural Networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples.

Disadvantages

- Choice of kernel function and parameter is important. Otherwise it will lead to state where no classification is possible.
- A second limitation is speed and size, both in training and testing sets.
- It is sensitive to noise
- A relatively small number of mislabeled examples can dramatically decrease the performance.

References

- http://www.svms.org/
- http://www.wikipedia.org/wiki/Support_vector_machine
- http://www.naturebiotechnology

Thank you

- Any questions??