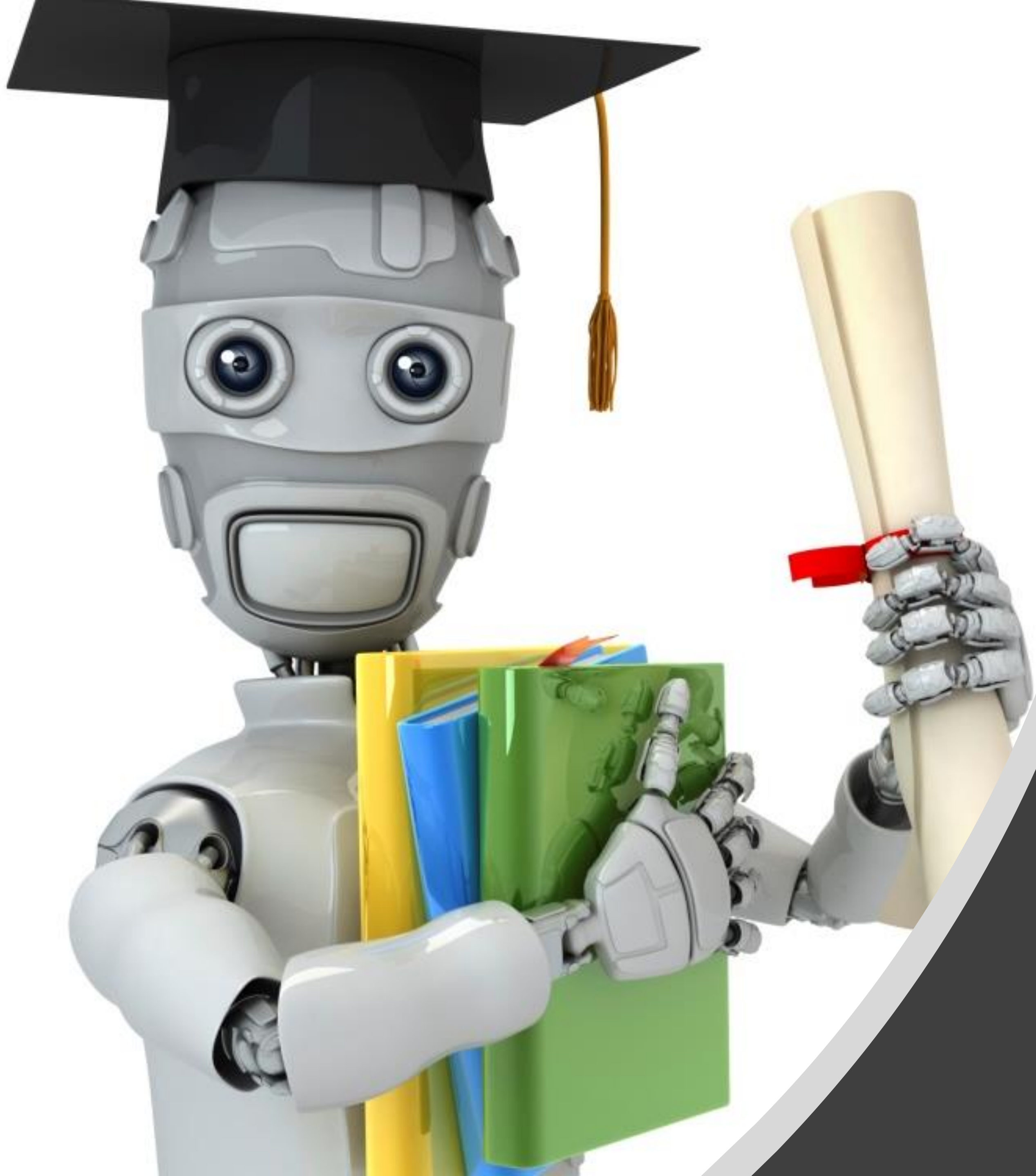
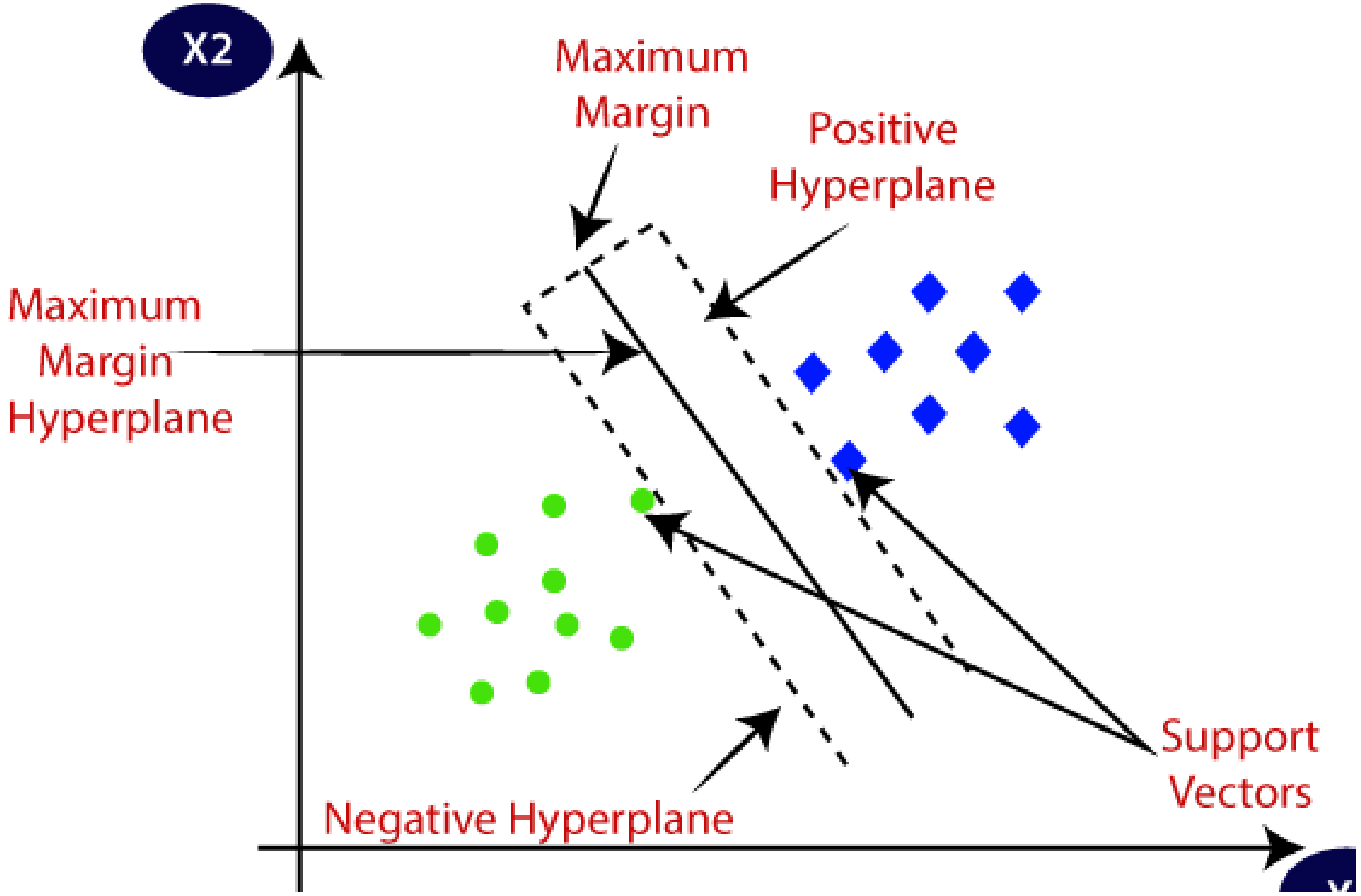


Machine Learning

Prepared By:

Dr. Sara Sweidan

• . Consider the following diagram in which there are two different categories that are classified using a decision boundary or hyperplane.



Support Vector Machine

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.
- However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create ***the best line*** or ***decision boundary*** that can segregate n-dimensional **space into classes** so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a ***hyperplane***.
- SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as **Support Vector Machine**.

The mathematical function used for the transformation is known as the ***kernel function***. Following are the popular functions.

- Linear
- Polynomial
- Radial basis function (RBF)
- Sigmoid

A linear kernel function is recommended when linear separation of the data is straightforward. In other cases, one of the other functions should be used. You will need to experiment with the different functions to obtain the best model in each case, as they each use different algorithms and parameters

Explain Hyperplanes and Support Vectors.

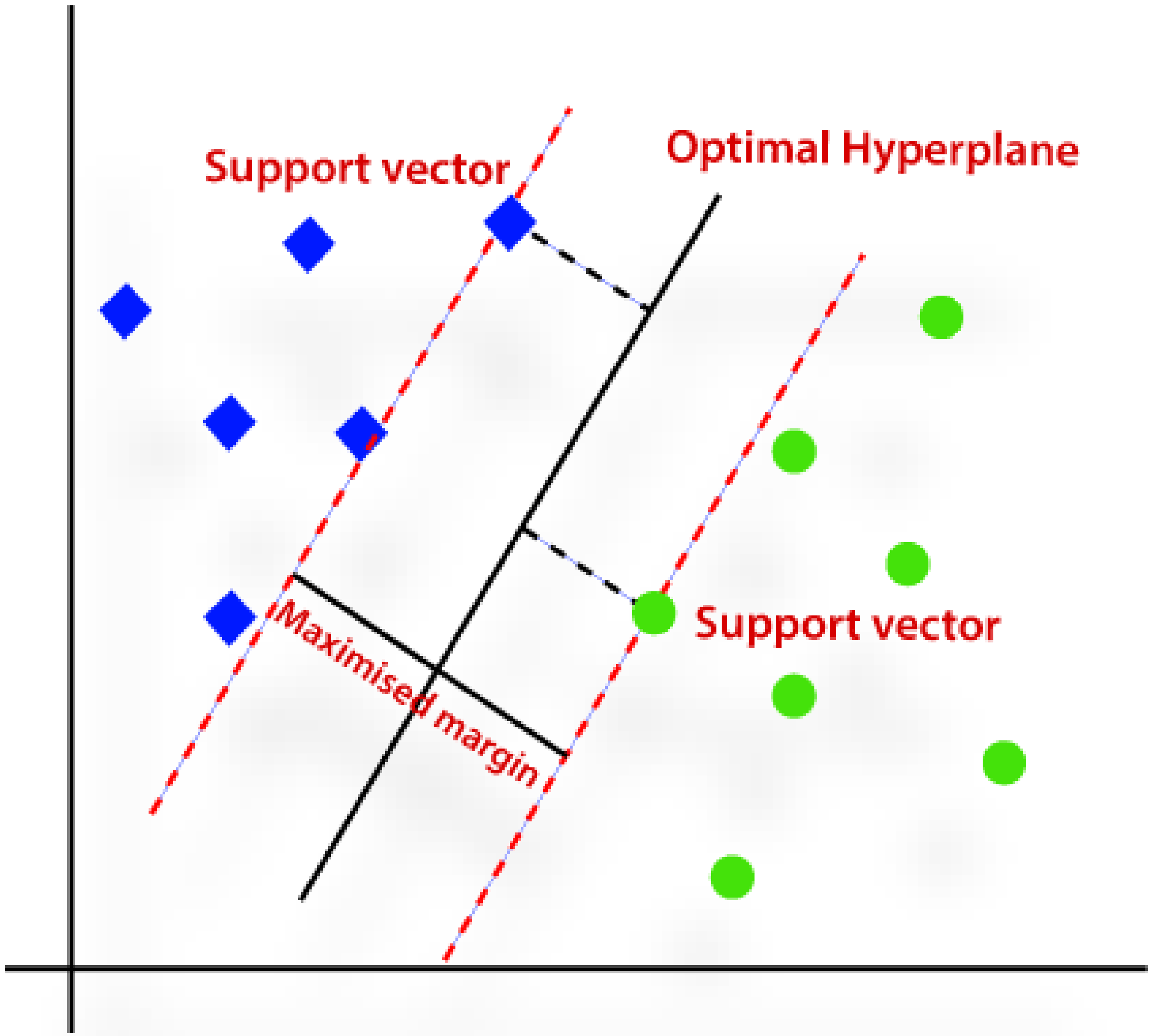
Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the **hyperplane of SVM**.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Explain Hyperplanes and Support Vectors.

- **Support Vectors:** The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.



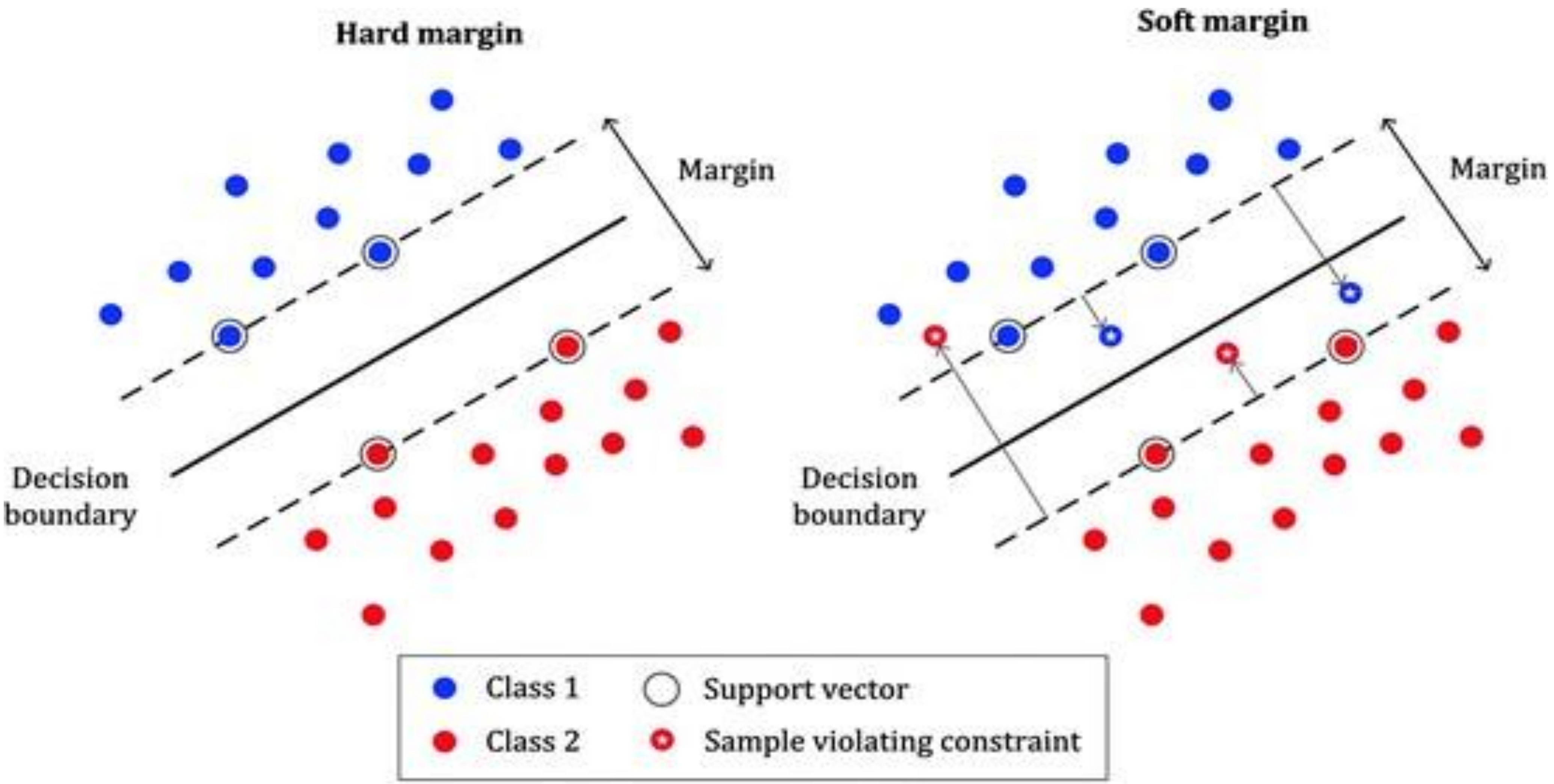
Hard and soft margins with the help of sketch

The distance of the vectors from the hyperplane is called the **margin** which is a separation of a line to the closest class points. We would like to choose a hyperplane that maximizes the margin between classes.

1- Soft Margin – As most of the real-world data are not fully linearly separable, we will allow some margin violation to occur which is called soft margin classification. It is better to have a large margin, even though some constraints are violated. Margin violation means choosing a hyperplane, which can allow some data points to stay on either the incorrect side of the hyperplane and between the margin and correct side of the hyperplane.

2- Hard Margin – If the training data is linearly separable, we can select two parallel hyperplanes that separate the two classes of data, so that the distance between them is as large as possible.

Hard and soft margins with the help of sketch



Support Vector Machine terminology

Support Vector Machines are part of the supervised learning model with an associated learning algorithm. It is the most powerful and flexible algorithm used for classification, regression, and detection of outliers. It is used in case of high dimension spaces; where each data item is plotted as a point in n-dimension space such that each feature value corresponds to the value of specific coordinate.

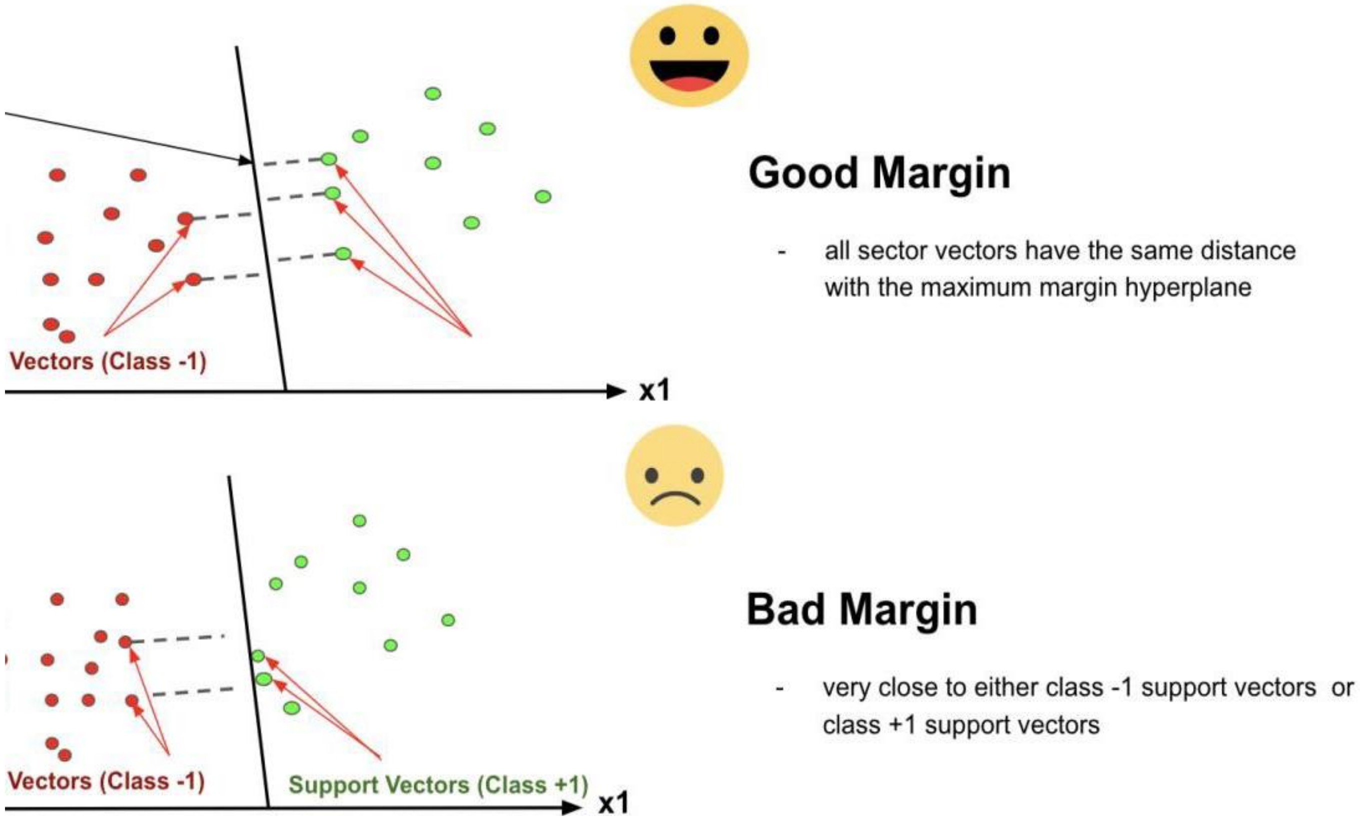
The classification is made on the basis of a **hyperplane/line** as wide as possible, which distinguishes between two categories more clearly. Basically, support vectors are the observational points of each individual, whereas the support vector machine is the boundary that differentiates one class from another class.

Support Vector Machine terminology

Some significant terminology of SVM is given below:

- **Support Vectors:** These are the data point or the feature vectors lying nearby to the hyperplane. These help in defining the separating line.
- **Hyperplane:** It is a subspace whose dimension is one less than that of a decision plane. It is used to separate different objects into their distinct categories. The best hyperplane is the one with the maximum separation distance between the two classes.
- **Margins:** It is defined as the distance (perpendicular) from the data point to the decision boundary. There are two types of margins: good margins and margins. **Good margins** are the one with huge margins and the **bad margins** in which the margin is minor.

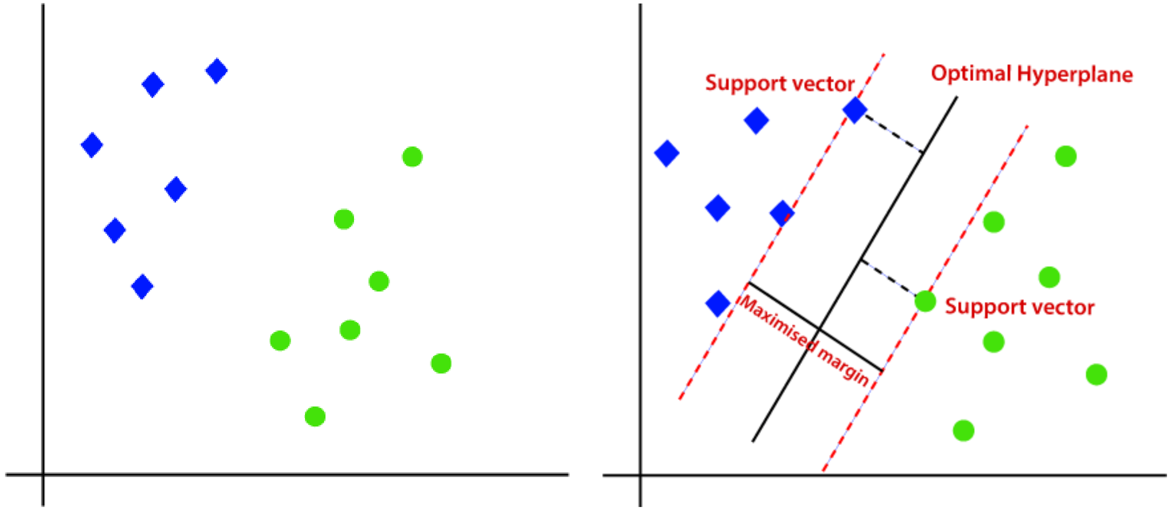
Support Vector Machine terminology

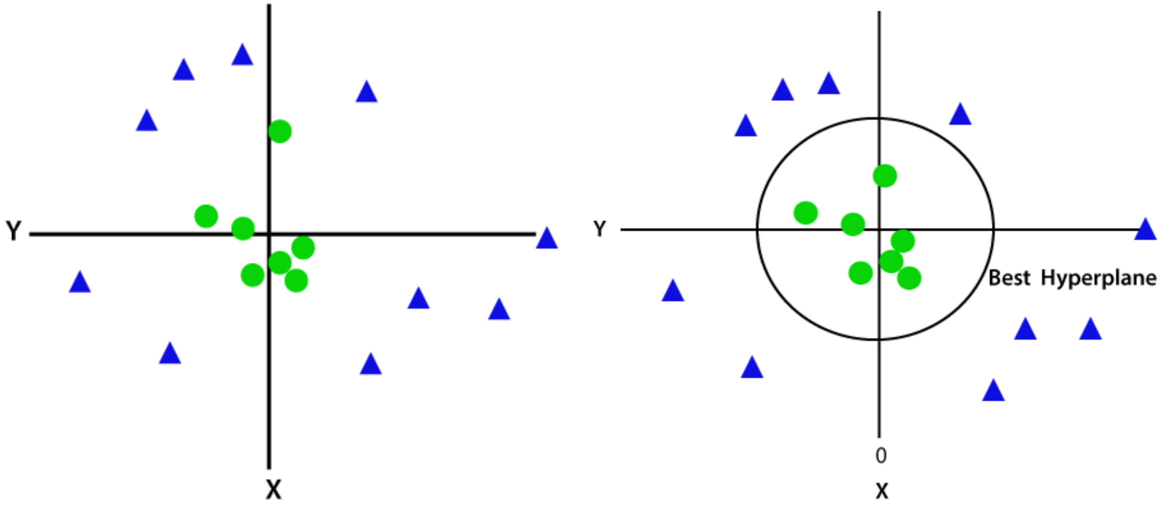


The main goal of SVM is to find the maximum marginal hyperplane, so as to segregate the dataset into distinct classes.

It undergoes the following steps:

- Firstly the SVM will produce the hyperplanes repeatedly, which will separate out the class in the best suitable way.
- Then we will look for the best option that will help in correct segregation.





Linear SVM

non-linear SVM

Advantages	limitations
-SVM‘s are very good when we have no assumptions on the data. -Works well with even unstructured and semi structured data like text, Images and trees. - SVM with an appropriate kernel function, can solve any complex problem. -It scales relatively well to high dimensional data. SVM is more effective in high dimensional spaces. -SVM models have generalization in practice; the risk of over-fitting is less in SVM. -SVM works relatively well when there is a clear margin of separation between classes.	-Choosing a good kernel function is not easy. -SVM algorithm is not suitable for large data sets. -Long training time for large datasets. -The SVM hyper parameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact. -SVM does not perform very well when the data set has more noise i.e. target classes are overlapping. -As the support vector classifier works by putting data points, above and below the classifying hyperplane there is no probabilistic explanation for the classification.

calculating hyperplane and margin

Hyper parameters of SVM are considered as Kernel, Regularization, Gamma and Margin.

Kernel: The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays role.

For **linear kernel** the equation for prediction for a new input using the dot product between the input (x) and each support vector (x_i) is calculated as follows:

f(x) = B(0) + sum(a_i * (x, x_i))

This is an equation that involves calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and a_i (for each input) must be estimated from the training data by the learning algorithm.

Kernel functions: linear, polynomial, rbf, sigmoid

Kernel function $k(x_n, x_i)$, which transform the original data space with higher dimension, this process includes the transformation function with dot product $\emptyset(x)$.

$$k(x_n, x_i) = \emptyset(x_n)\ \emptyset(x_i)$$

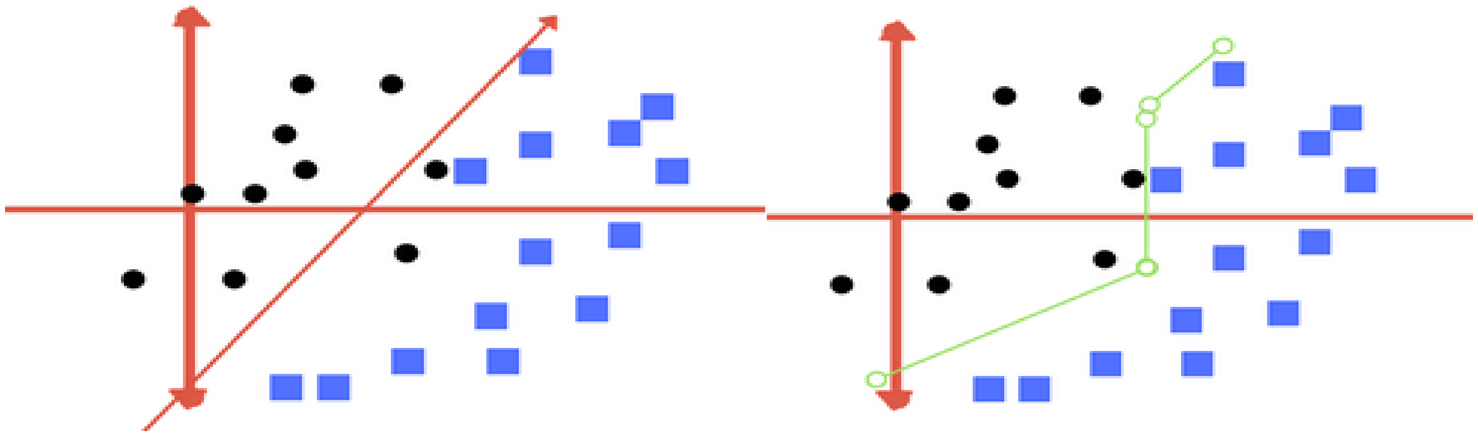
The aim is the data which already transformed into higher dimensions can be separated easily. Thus the hyperplane function will be:

$$f(x_i) = \sum_{n=1}^{N} \alpha_n y_n k(x_n, x_i) + b$$

Where, x_n : support vector data, y_n : label of membership class [-1,+1] with $n = 1,2,3,\ldots N$,
α_n: multiplier

calculating hyperplane and margin

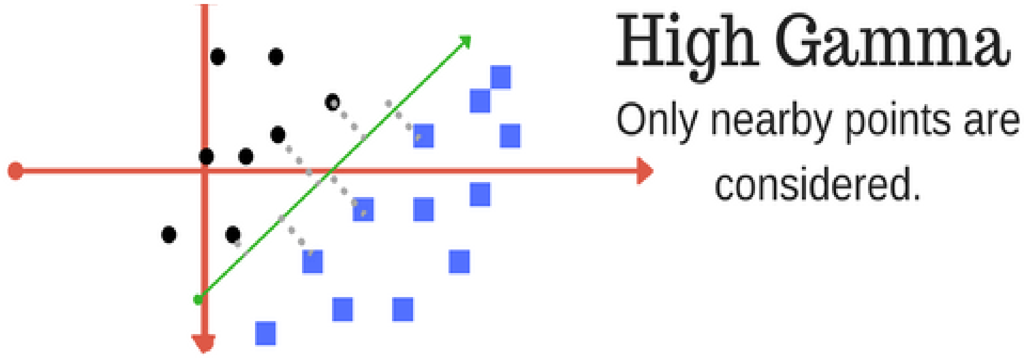
Regularization: The Regularization parameter (often termed as C parameter in python's sklearn library) tells the SVM optimization how much you want to avoid misclassifying each training example. For large values of C, the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly. Conversely, a very small value of C will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points.

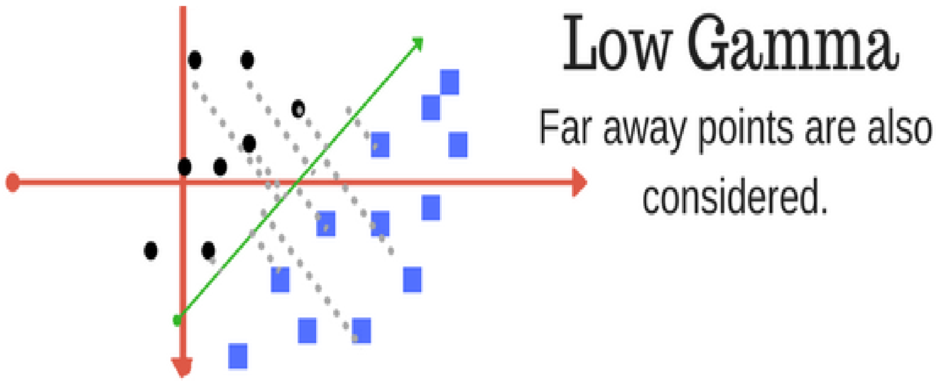


low regularization value high regularization value

calculating hyperplane and margin

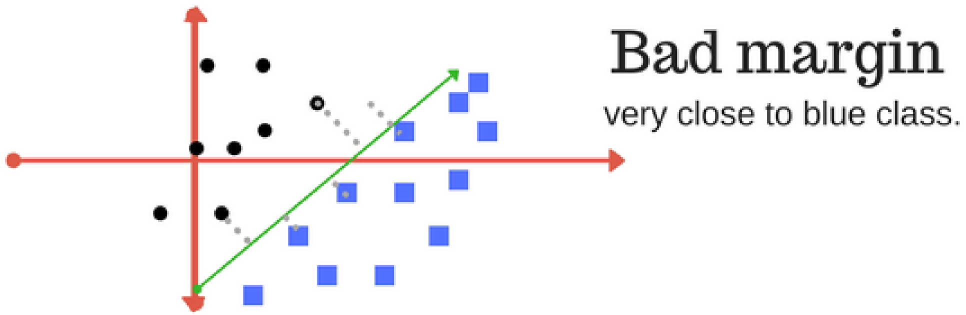
Gamma: The gamma parameter defines how far the influence of a single training example reaches. low gamma, points far away from plausible separation line are considered in calculation for the separation line. Whereas high gamma means the points close to plausible line are considered in calculation.

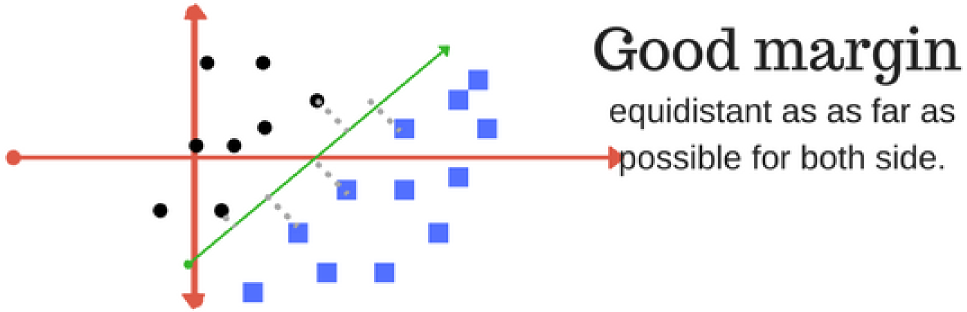




calculating hyperplane and margin

A margin is a separation of line to the closest class points.
A **good margin** is one where this separation is larger for both the classes. Images below gives to visual example of good and bad margin. A good margin allows the points to be in their respective classes without crossing to other class.





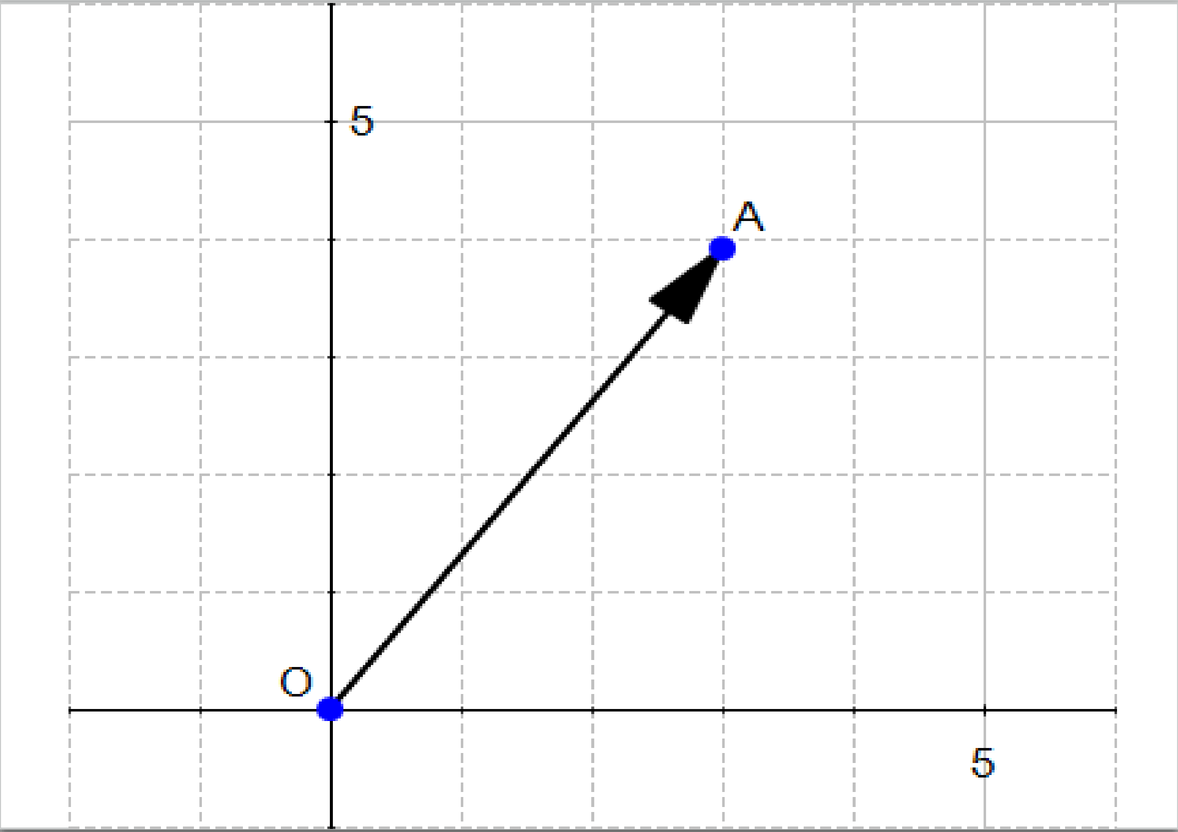
calculating hyperplane and margin

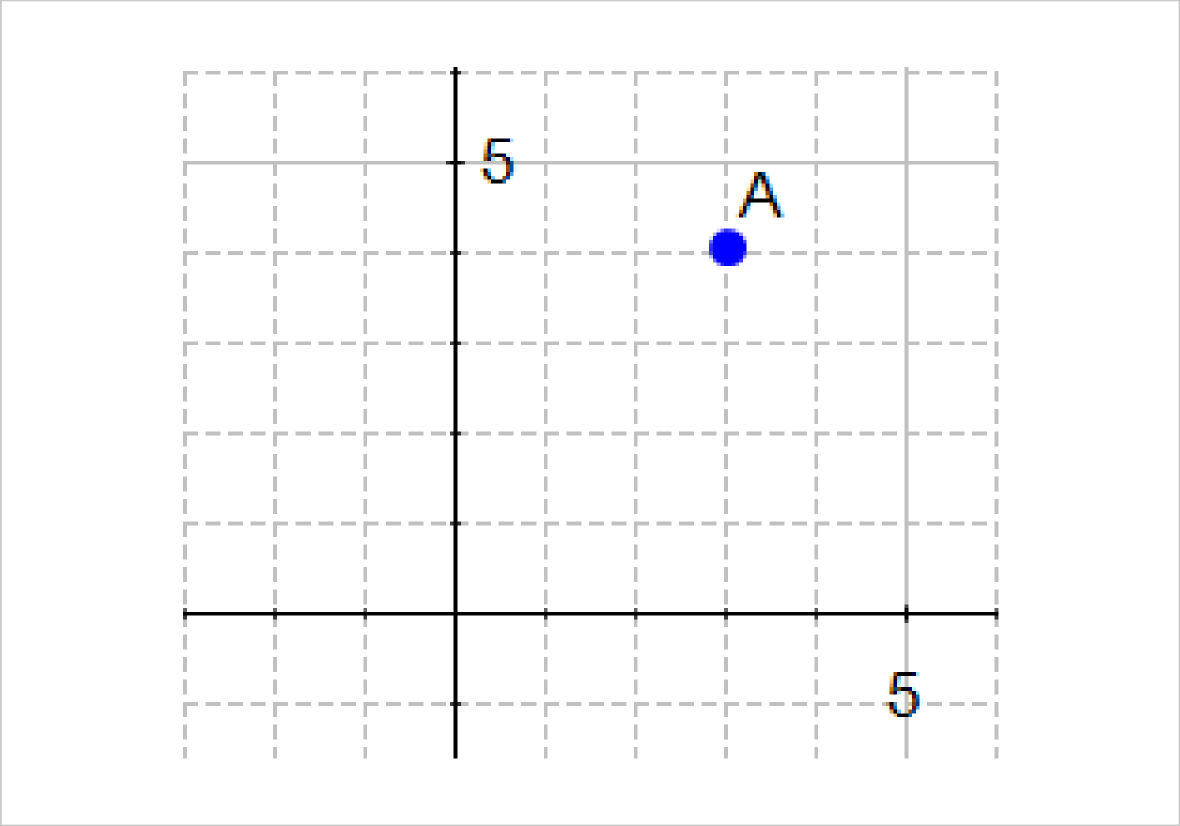
In Support Vector Machine, there is the word **vector.** That means it is important to understand vector well and how to use them.

- What is a vector?
  - its norm
  - its direction
- How to add and subtract vectors?
- What is the dot product?
- How to project a vector onto another?
- What is the equation of the hyperplane?
- How to compute the margin?

What is a vector?

- Definition: Any point $x=(x1, x2), x \neq 0$, specifies a vector in the plane, namely the vector starting at the origin and ending at x.
- If we define a point $A(3,4)$
- This definition means that there exists a vector between the origin and A.





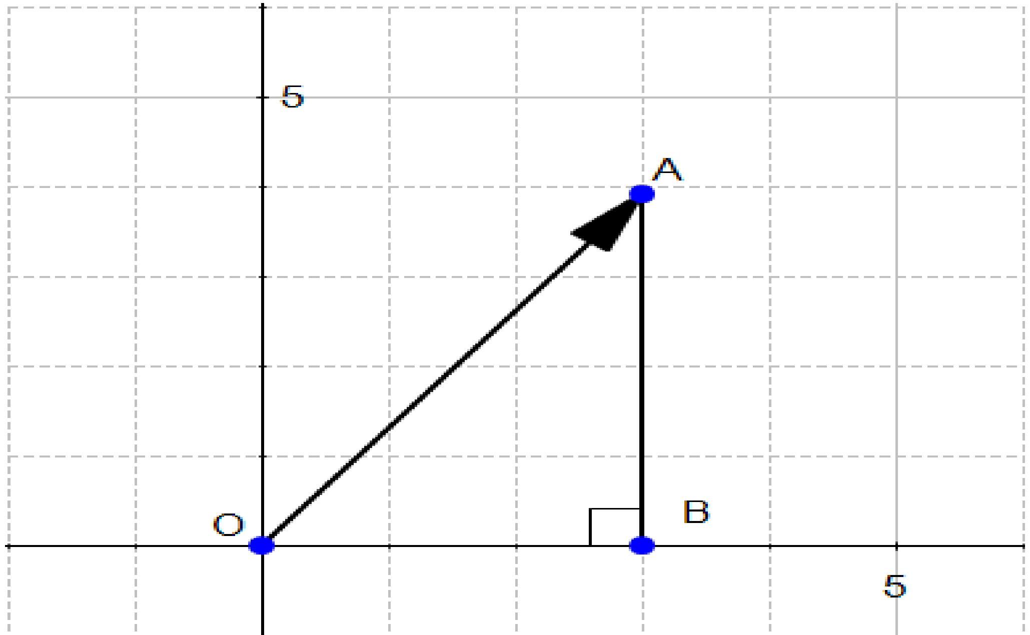
What is a vector?

Definition: A vector is an object that has both a magnitude and a direction.

We will now look at these two concepts.

1) The magnitude

The magnitude or length of a vector *x* is written ‖*x*‖ and is called its norm.



$$OA^2 = OB^2 + AB^2$$

$$OA^2 = 3^2 + 4^2$$

$$OA^2 = 25$$

$$OA = 5$$

$$\|OA\| = 5$$

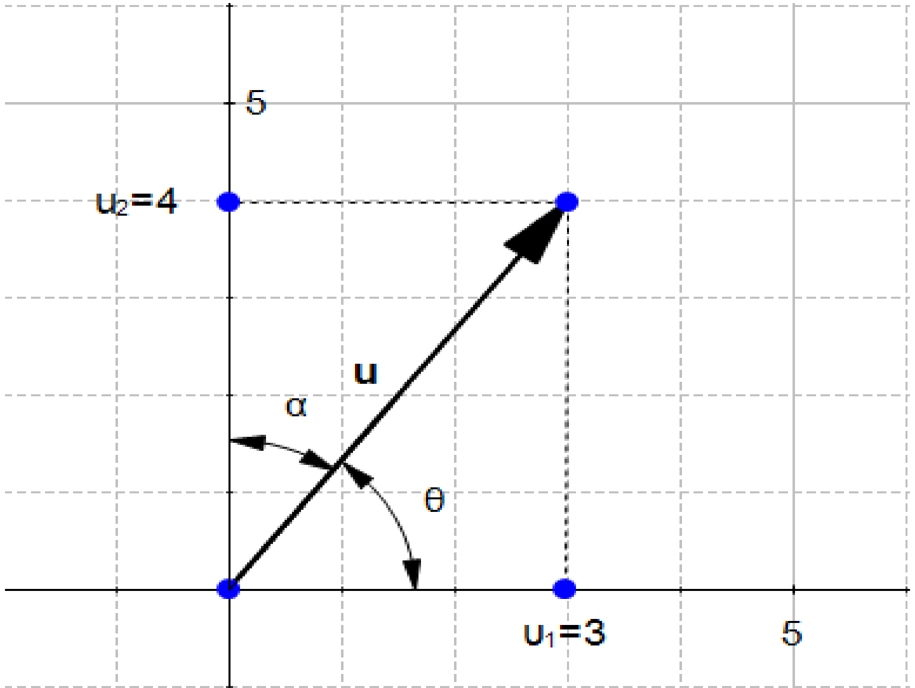
What is a vector?

Definition: A vector is an object that has both a magnitude and a direction.

We will now look at these two concepts.

1) The direction

Definition: The direction of a vector u(u_1,u_2) is the vector $\frac{\boldsymbol{u_1}}{\|\boldsymbol{u}\|}$, $\frac{\boldsymbol{u_2}}{\|\boldsymbol{u}\|}$



The direction of the vector **u** is defined by the angle θ with respect to the horizontal axis, and with the angle α with respect to the vertical axis. This is tedious. Instead of that we will use the cosine of the angles.

$\cos(\theta) = \frac{\boldsymbol{u_1}}{\|\boldsymbol{u}\|}$, $\cos(\alpha) = \frac{\boldsymbol{u_2}}{\|\boldsymbol{u}\|}$

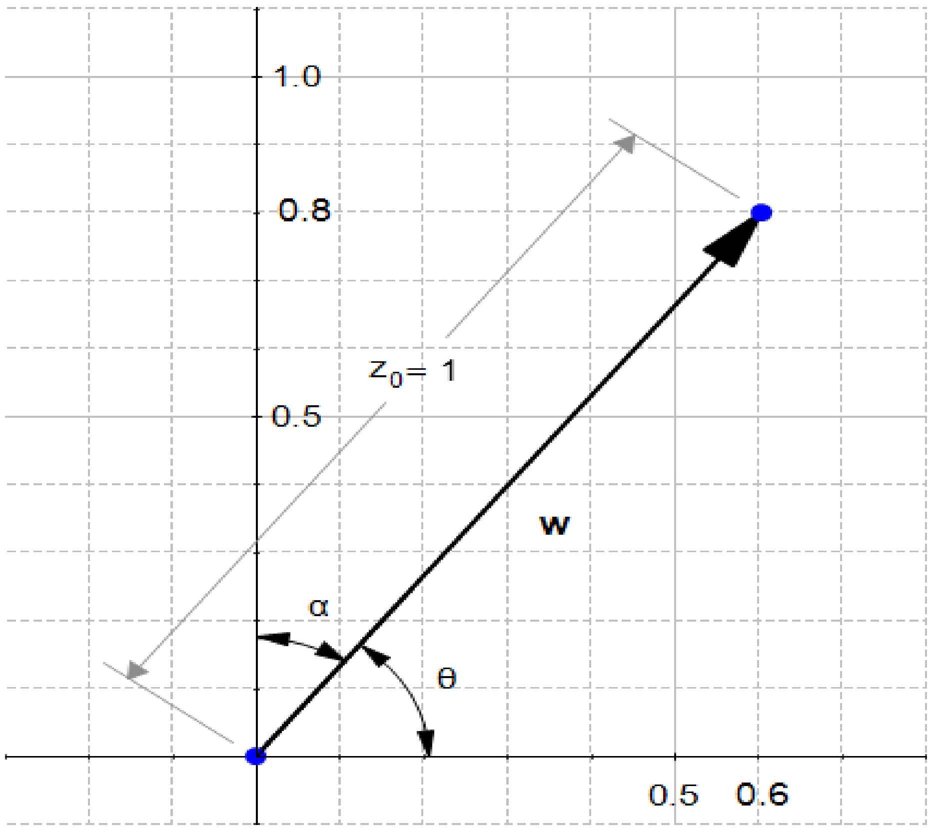
We will now compute the direction **w** of the vector **u**

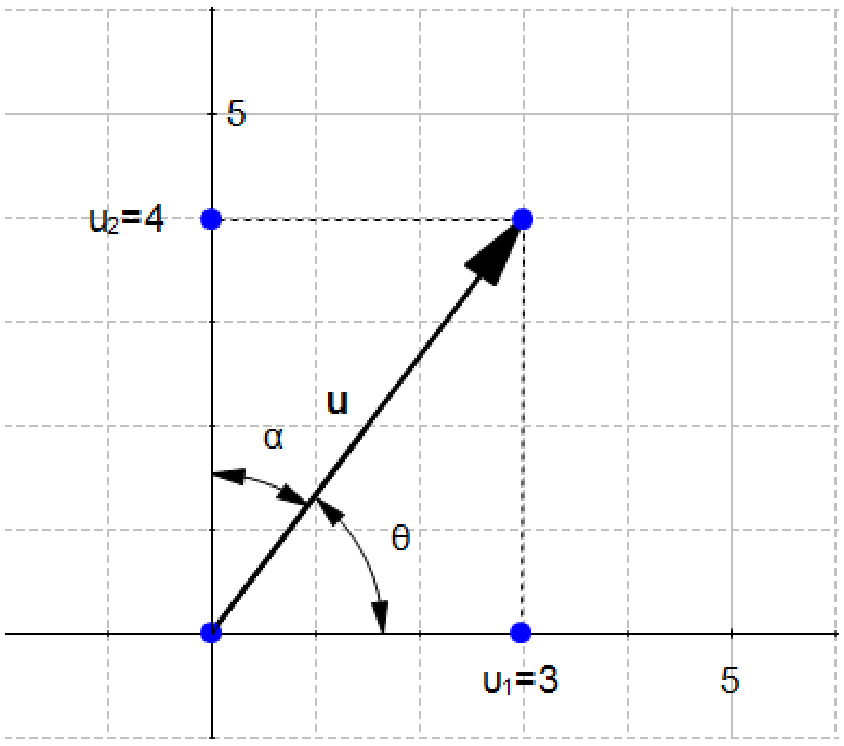
$\cos(\theta) = \frac{\mathbf{3}}{\mathbf{5}} = 0.6$, $\cos(\alpha) = \frac{\mathbf{4}}{\mathbf{5}} = 0.8$

So the direction of u(3,4) is **w**(0.6,0.8)

What is a vector?

- If we draw this vector we get



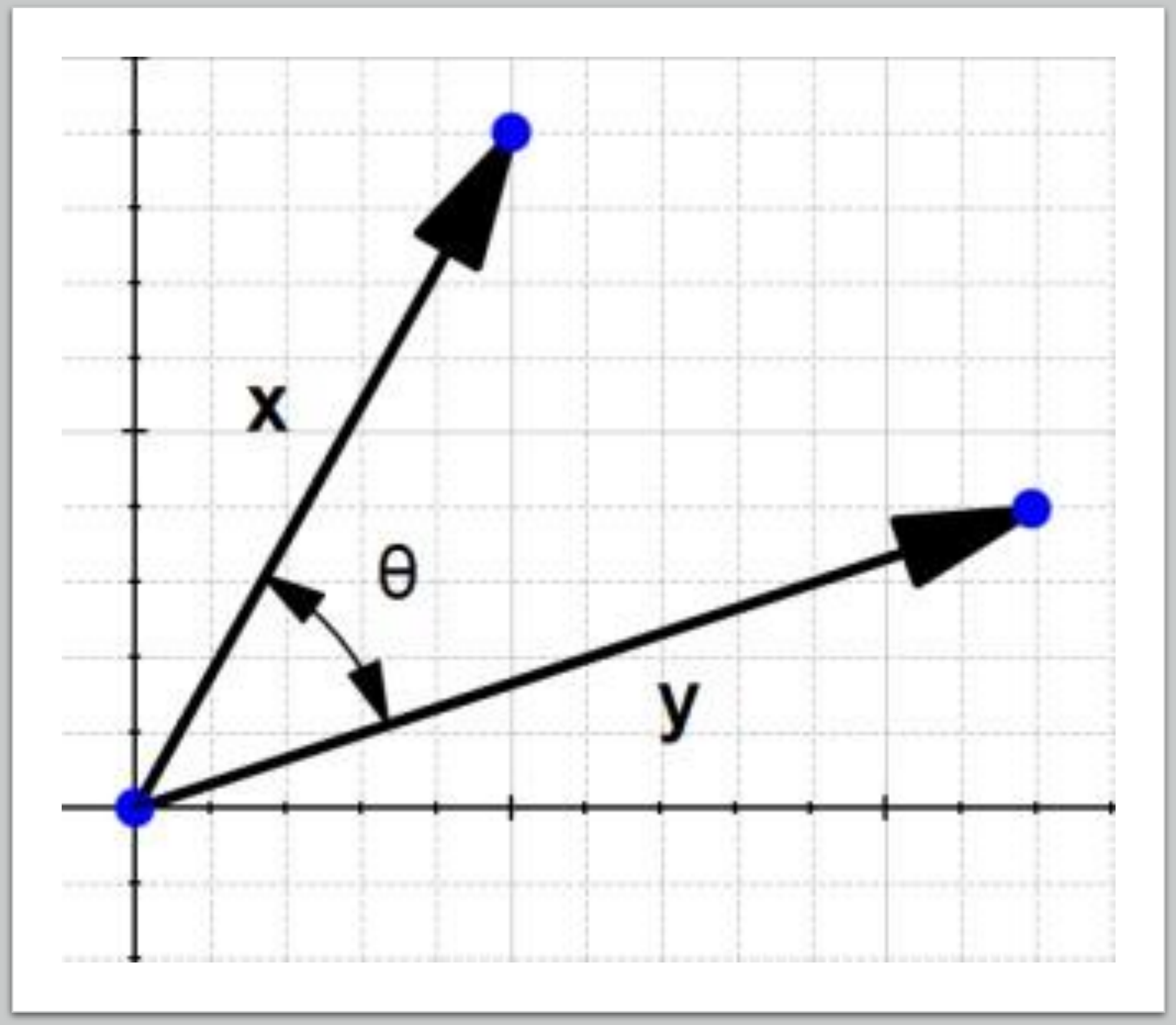


The dot product

- One **very** important notion to understand SVM is [the dot product.]
- Definition: Geometrically, it is the product of the Euclidian magnitudes of the two vectors and the cosine of the angle between them.
- This means if we have two vectors **x** and **y** and there is an angle θ (theta) between them, their dot product is:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \| x \| \| y \| \cos(\theta)$$

dot product = inner product (x,y)

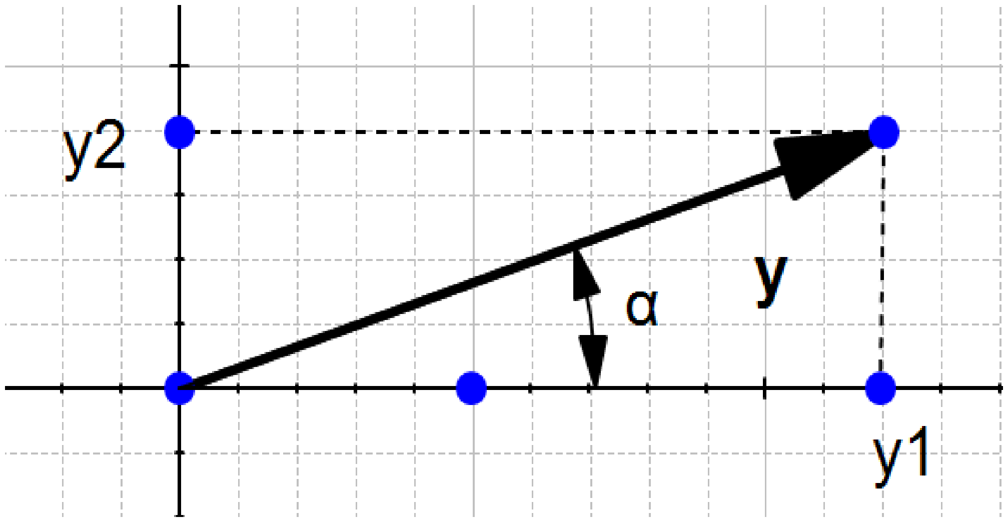


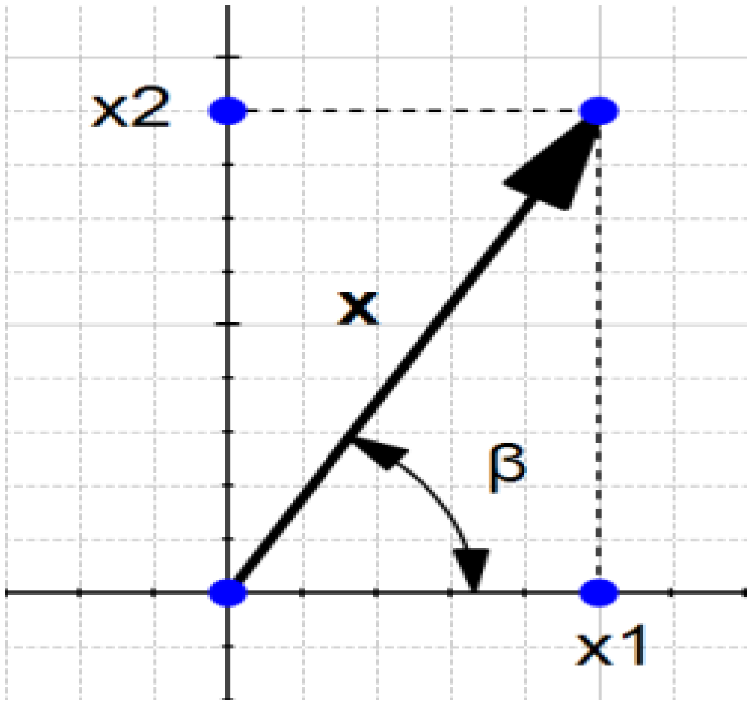
The dot product

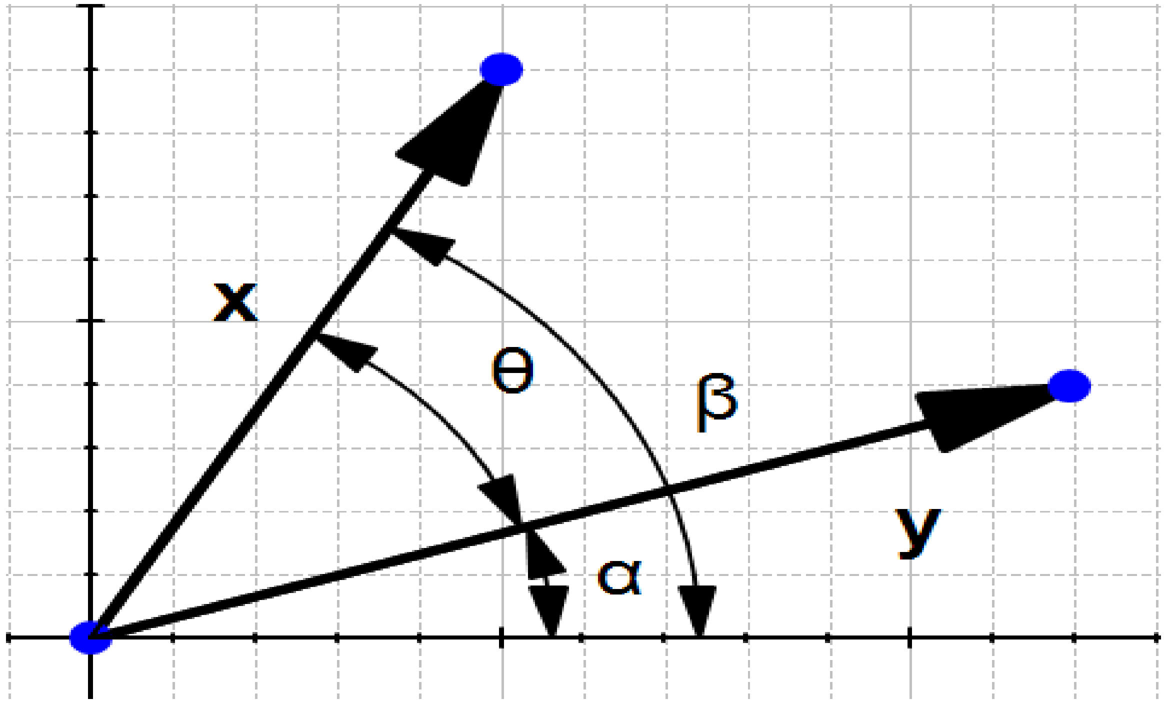
By definition:

cos(θ)=adjacent/hypotenuse

So, $\theta=\beta-\alpha$, $\cos(\theta) = \cos(\beta-\alpha)$







$cos(\theta) = cos(\beta - \alpha)$

$cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha)$

So, $cos(\theta) = \frac{\boldsymbol{x_1}}{\|\boldsymbol{x}\|}\frac{\boldsymbol{y_1}}{\|\boldsymbol{y}\|} + \frac{\boldsymbol{x_2}}{\|\boldsymbol{x}\|}\frac{\boldsymbol{y_2}}{\|\boldsymbol{y}\|}$

$$cos(\theta) = \frac{\boldsymbol{x_1 y_1} + \boldsymbol{x_2 y_2}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$$

$$\|\boldsymbol{x}\|\|\boldsymbol{y}\|.\ cos(\theta) = \boldsymbol{x_1 y_1} + \boldsymbol{x_2 y_2}$$

$$\boldsymbol{x_1 y_1} + \boldsymbol{x_2 y_2} = \sum_{i=1}^{2}(\boldsymbol{x_i y_i})$$

Which equal:

$$\|\boldsymbol{x}\|\|\boldsymbol{y}\|.\ cos(\theta) = \boldsymbol{x} . \boldsymbol{y}$$

$$cos(\beta) = \frac{adjacent}{hypotenuse} = \frac{x_1}{||x||}$$

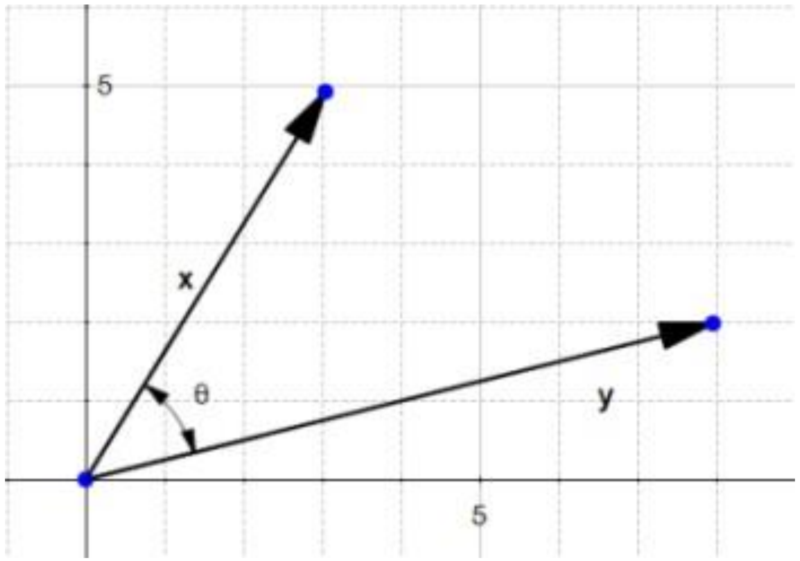
$$sin(\beta) = \frac{opposite}{hypotenuse} = \frac{x_2}{||x||}$$

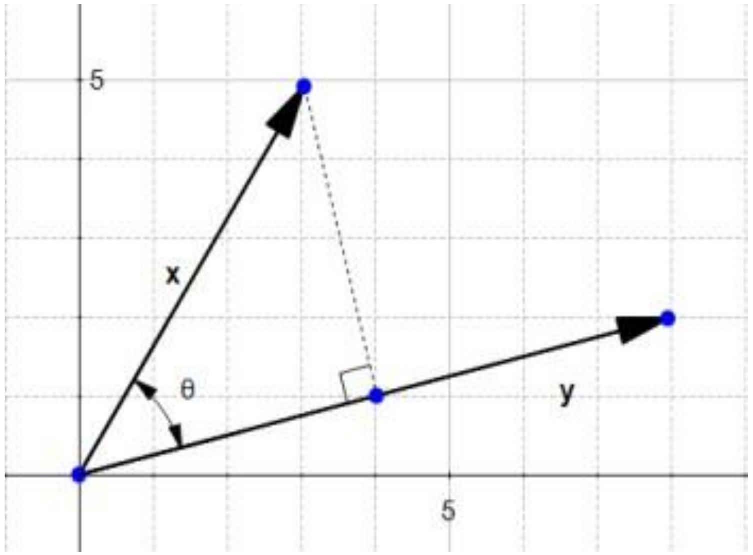
$$cos(\alpha) = \frac{adjacent}{hypotenuse} = \frac{y_1}{||y||}$$

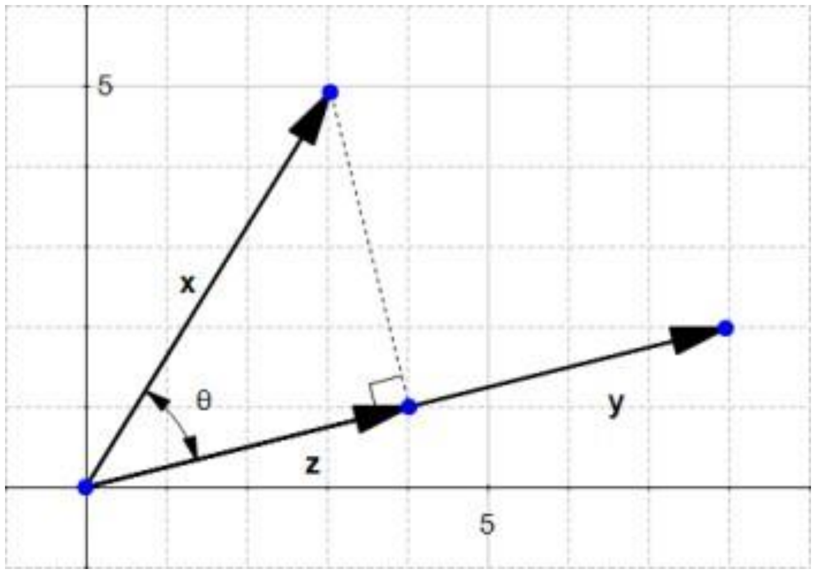
$$sin(\alpha) = \frac{opposite}{hypotenuse} = \frac{y_2}{||y||}$$

The orthogonal projection of a vector

Given two vectors **x** and **y**, we would like to find the orthogonal projection of **x** onto **y**.







To do this we project the vector **x** onto **y**

This gives us the vector **z**

By definition:

$\cos(\theta) = \frac{\|\mathbf{z}\|}{\|\mathbf{x}\|} \Rightarrow \| \mathbf{z} \| = \cos(\theta) \| \mathbf{x} \|$

From the previous section the dot product

$\cos(\theta) = \frac{x.y}{\|\mathbf{x}\|\|\mathbf{y}\|}$

we replace $\cos(\theta) \Rightarrow \| \mathbf{z} \| = \frac{x.y}{\|\mathbf{x}\|\|\mathbf{y}\|} \| \mathbf{x} \| \quad \Rightarrow \| \mathbf{z} \| = \frac{x.y}{\|\mathbf{y}\|}$

If we define u the direction of y and z

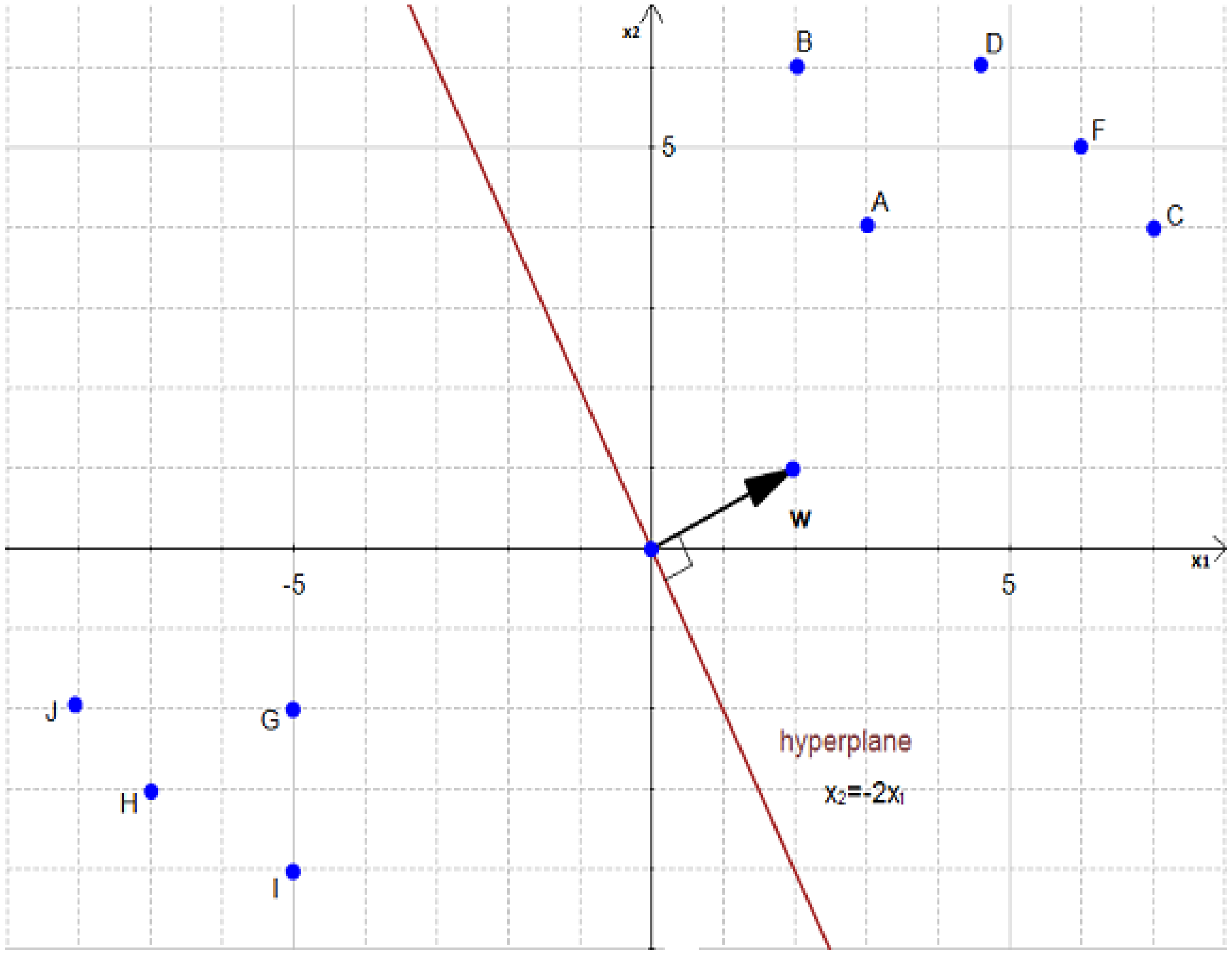
$$u = \frac{\boldsymbol{y}}{\|\mathbf{y}\|} \quad , u = \frac{\boldsymbol{z}}{\|\mathbf{z}\|}, \quad y = u.\ \|y\|,\ z = u.\ \|z\|, \|z\| = u.\ x$$

So,

We can say vector z = (u.x) u is the orthogonal projection of x onto y

Compute the distance from a point to the hyperplane

- we have a hyperplane $x_2 = -2\ x_1$, which separates two groups of data

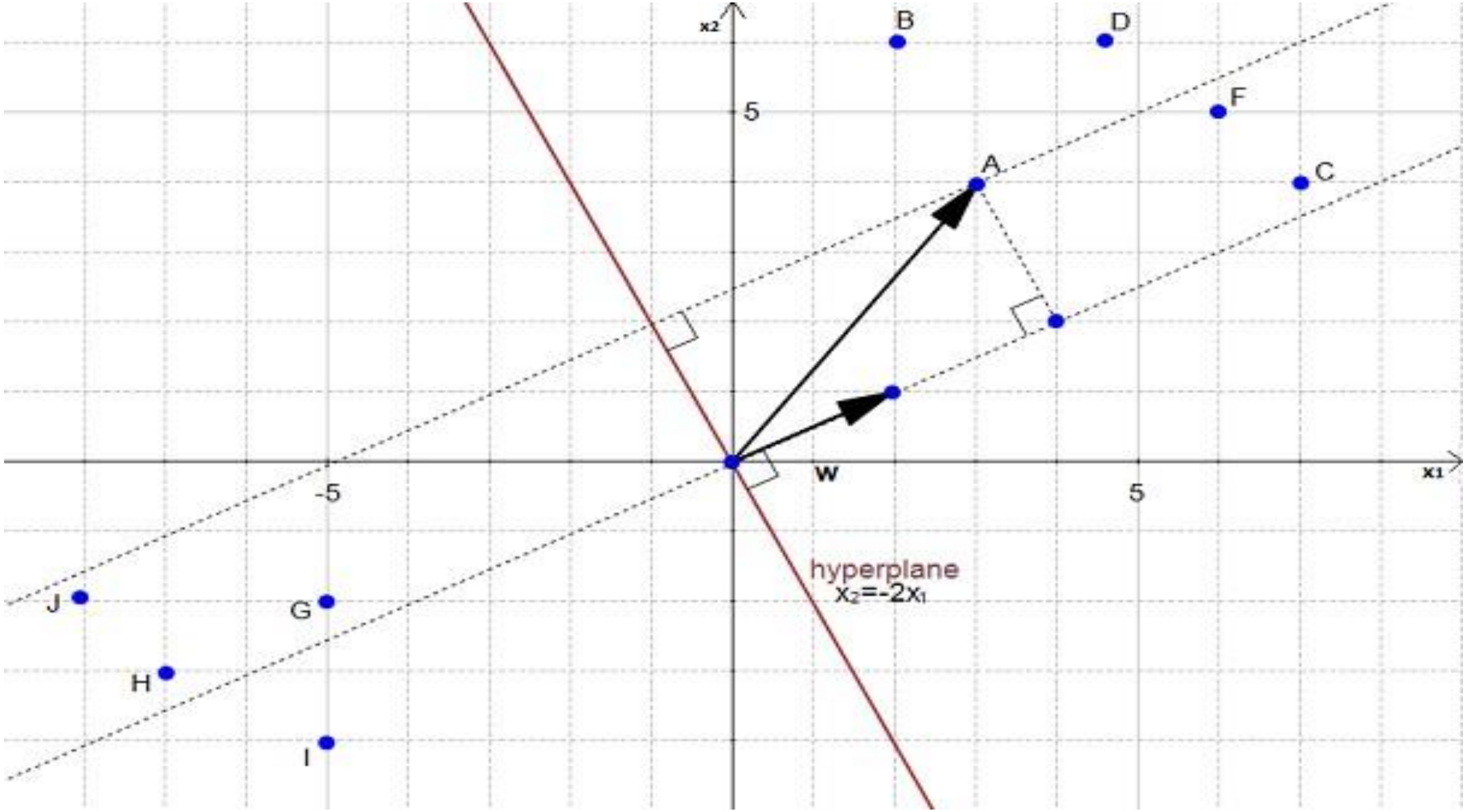


$x2 = -2x1$ which is equivalent to $\mathbf{w}^T\mathbf{x}=0$

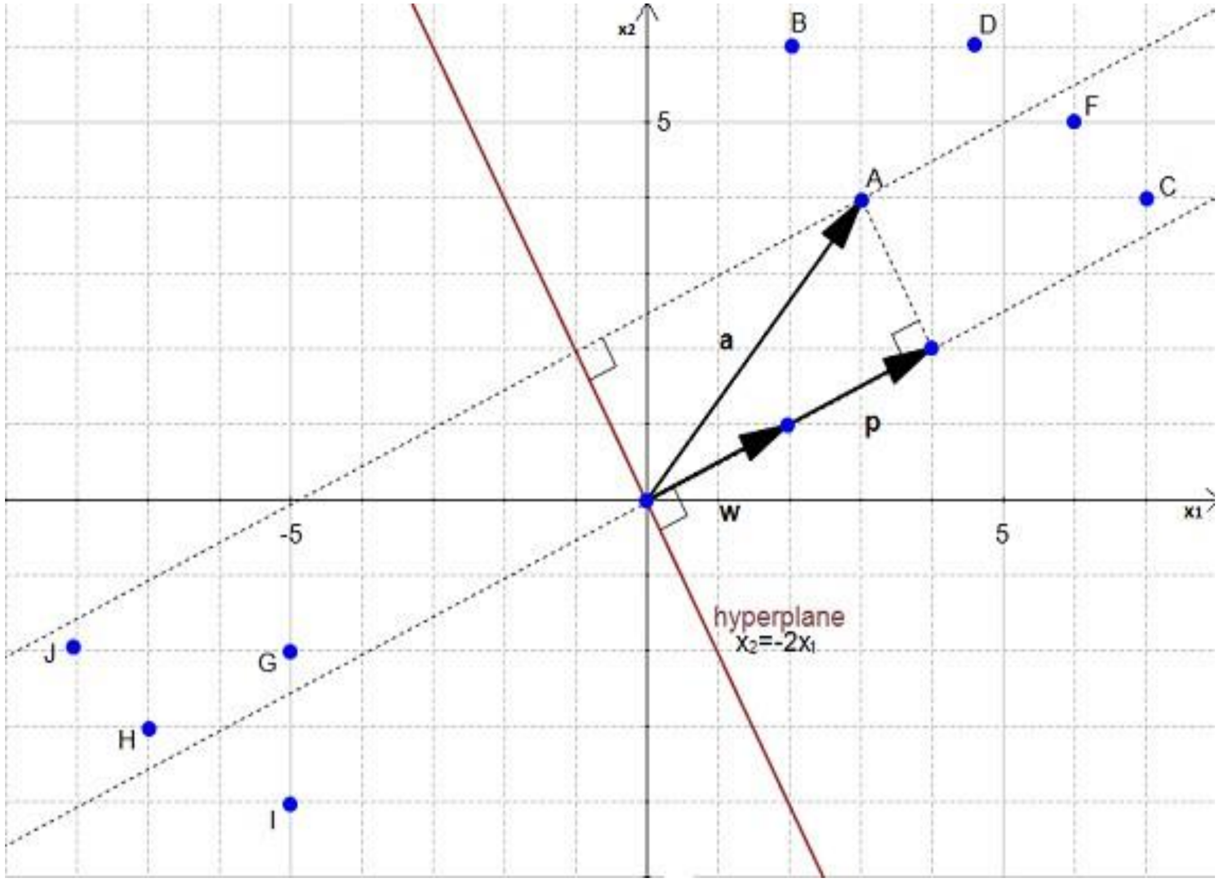
W $\begin{pmatrix}2\\1\end{pmatrix}$ x $\begin{pmatrix}x_1\\x_2\end{pmatrix}$

W (2,1) is a normal vector

We can view point A as a vector from origin to A. If we project it onto the normal vector **w**



We get the vector **p**



Our goal is to find the distance ‖p‖ between A(3,4) and hyperplane.
We start with w(2,1) which is normal to hyperplane, and a(3,4)

$\| w \| = \sqrt{2^2 + 1^2} = \sqrt{5}$
Let u the direction of w

$u = (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$
P is the orthogonal projection of a onto w so

P = $(u.\, a)\, u$
P= (3x $\frac{2}{\sqrt{5}}$ + 4x $\frac{1}{\sqrt{5}}$) u = $\frac{10}{\sqrt{5}}$. u
P= ($\frac{2}{\sqrt{5}}$ x $\frac{10}{\sqrt{5}}$, $\frac{1}{\sqrt{5}}$ x $\frac{10}{\sqrt{5}}$) = ($\frac{20}{5}$, $\frac{10}{5}$)
P= (4,2)
$\| p \| = \sqrt{4^2 + 2^2} = 2\sqrt{5}$
Margin = 2 $\| p \| = 4\sqrt{5}$


THANK YOU!