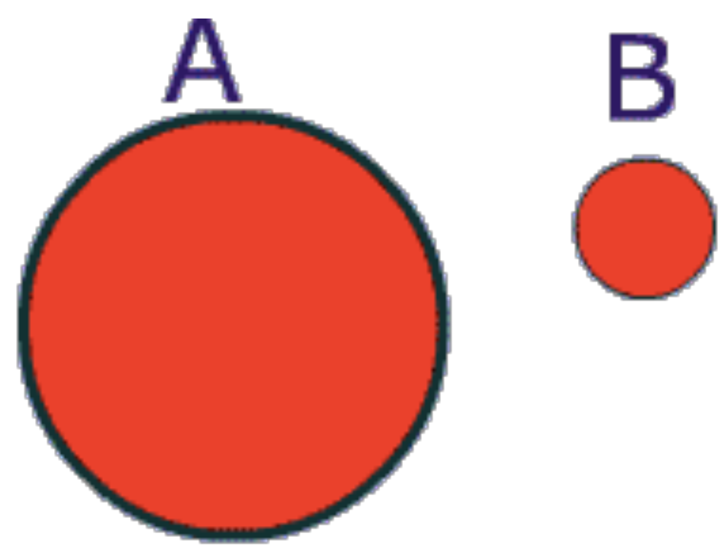
SVM Math

What is Magnitude ?

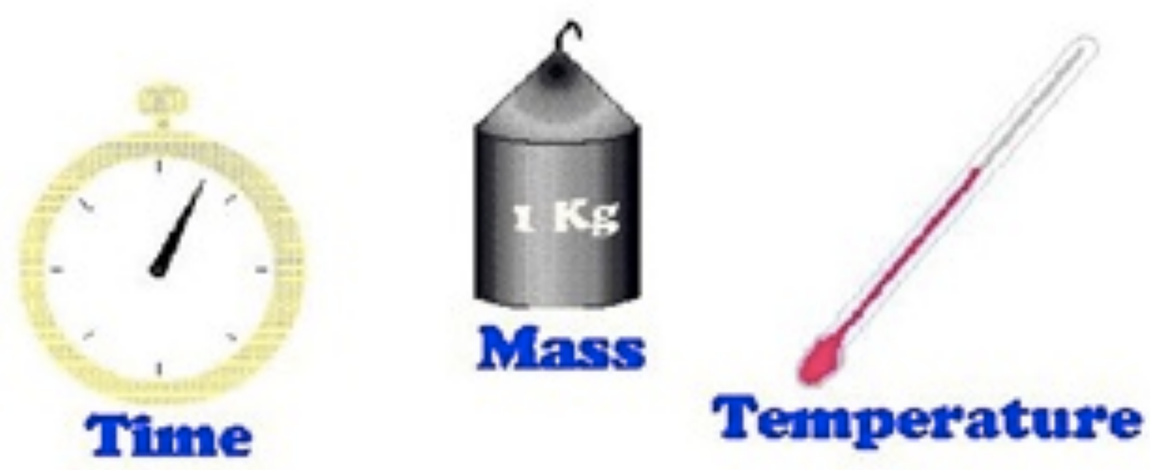


magnitude of a vector is shown by two vertical bars on either side of the vector

$|a|$

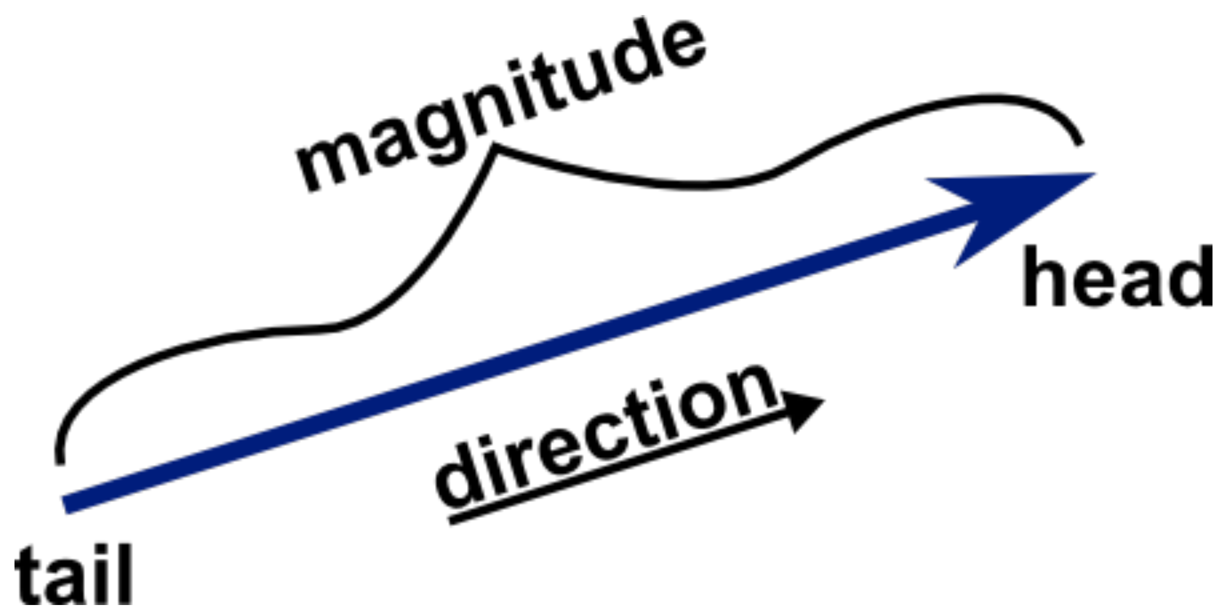
In mathematics, magnitude is the size of a mathematical object, a property by which the object can be compared as larger or smaller than other objects.

What is a scalar?

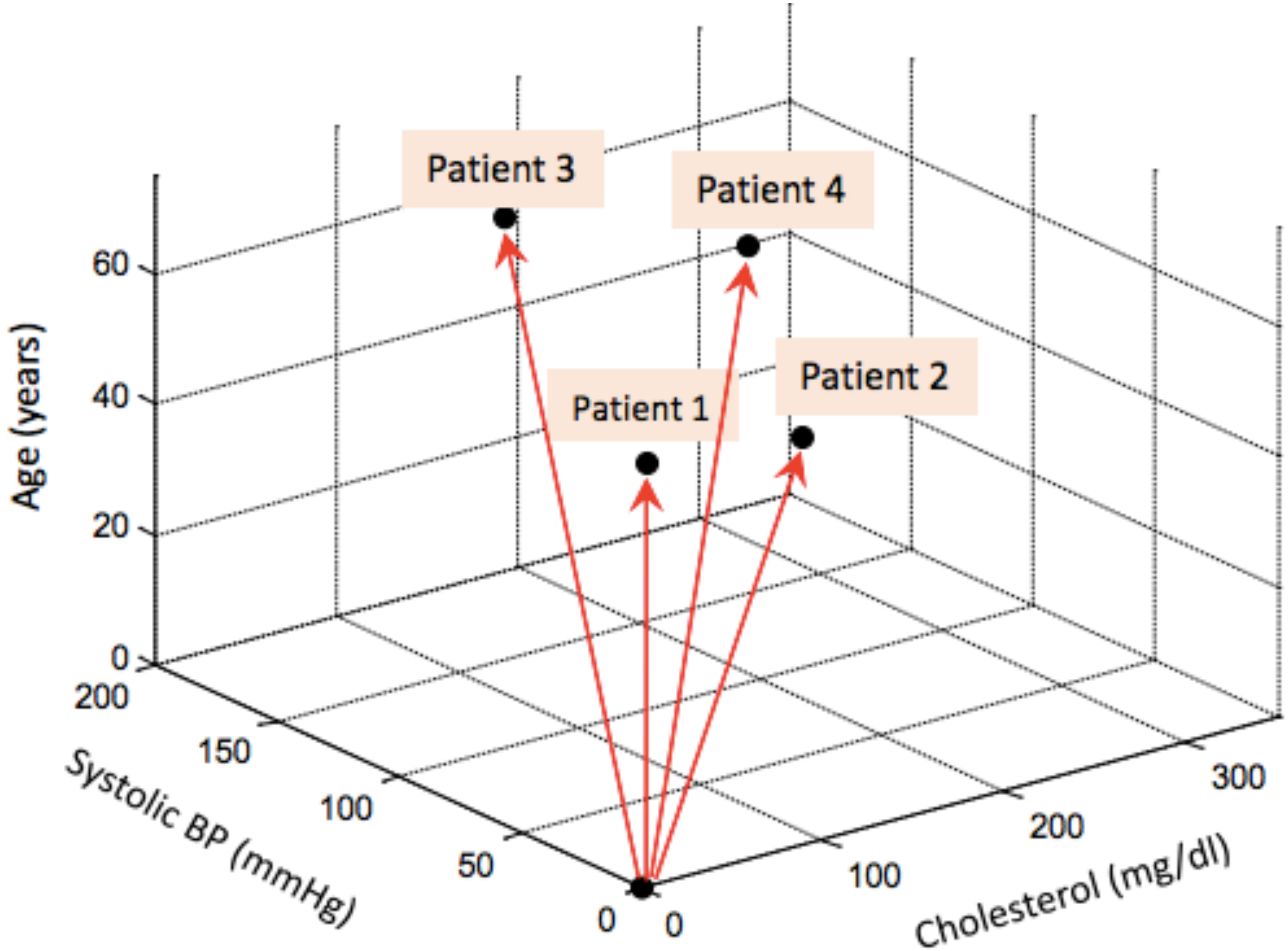


A scalar is a quantity that is fully described by a magnitude only. It is described by just a single number. Some examples of scalar quantities include speed, volume, mass, temperature, power, energy, and time.

What is a vector ?



Vectors are used to represent quantities that have both a magnitude and a direction. Good examples of quantities that can be represented by vectors are force and velocity.



Patient id	Cholesterol (mg/dl)	Systolic BP (mmHg)	Age (years)	Tail of the vector	Arrow-head of the vector
1	150	110	35	(0,0,0)	(150, 110, 35)
2	250	120	30	(0,0,0)	(250, 120, 30)
3	140	160	65	(0,0,0)	(140, 160, 65)
4	300	180	45	(0,0,0)	(300, 180, 45)

Magnitude of a Vector

The magnitude, or length, of the vector is given by,

$$\textbf{Magnitude} = \sqrt{x^2 + y^2} \quad \text{(for 2D vectors)}$$

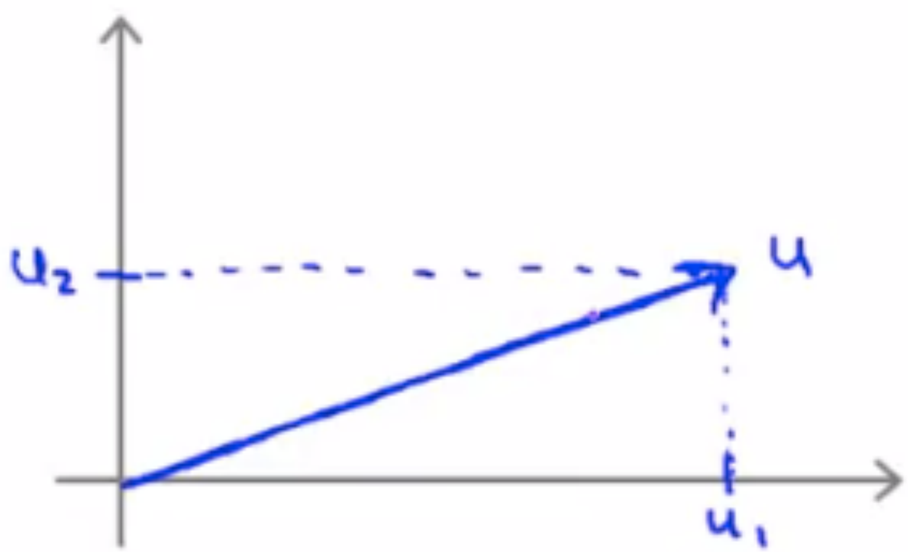
Euclidian length

$$\textbf{Magnitude} = \sqrt{x^2 + y^2 + z^2} \quad \text{(for 3D vectors)}$$

So if you have a vector given by the coordinates (3, 4), its magnitude is 5

$$v = \sqrt{3^2 + 4^2} = 5$$

Vector Inner Product



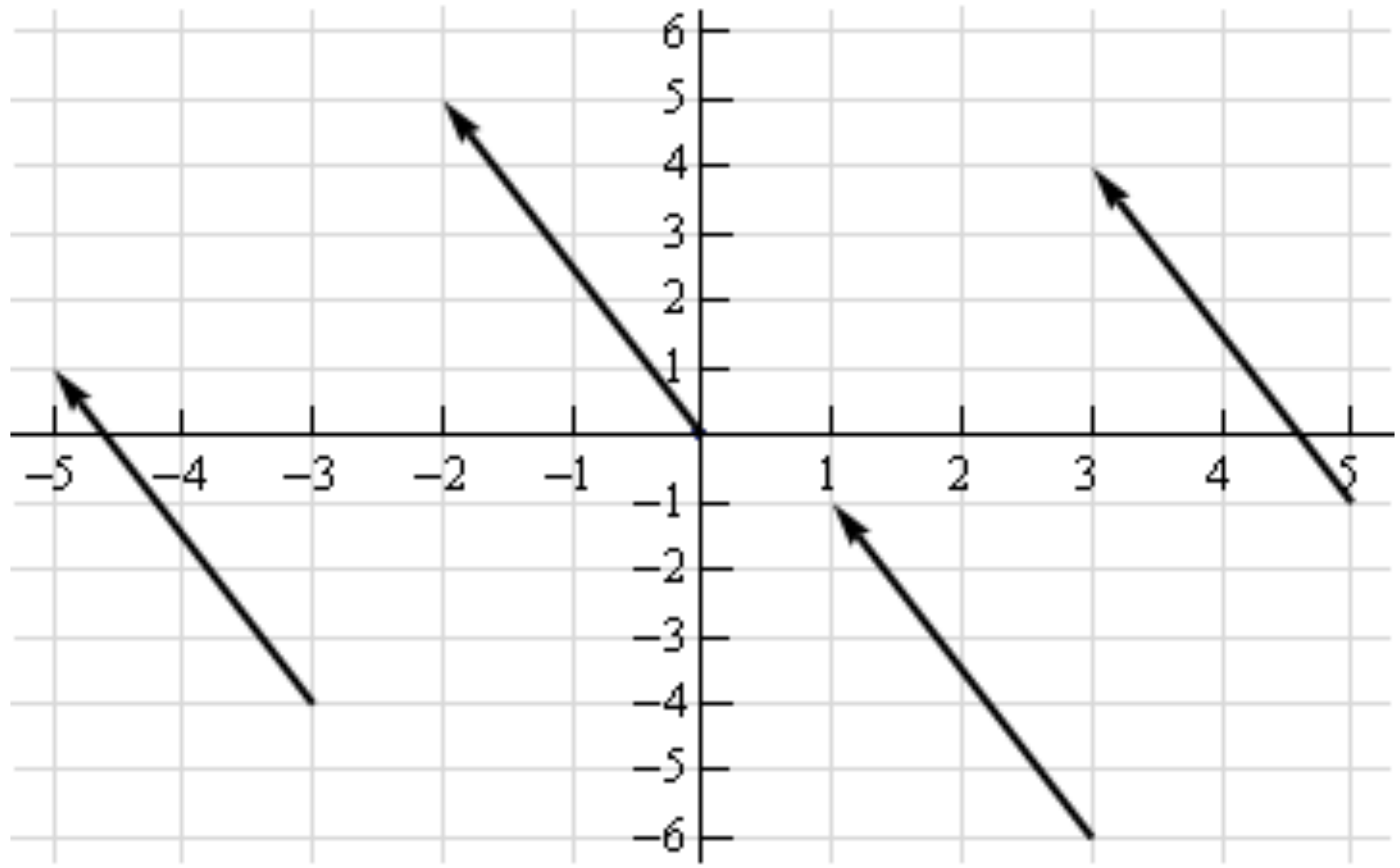
$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$u^T v = ?$

$\|u\|$ = length of vector u

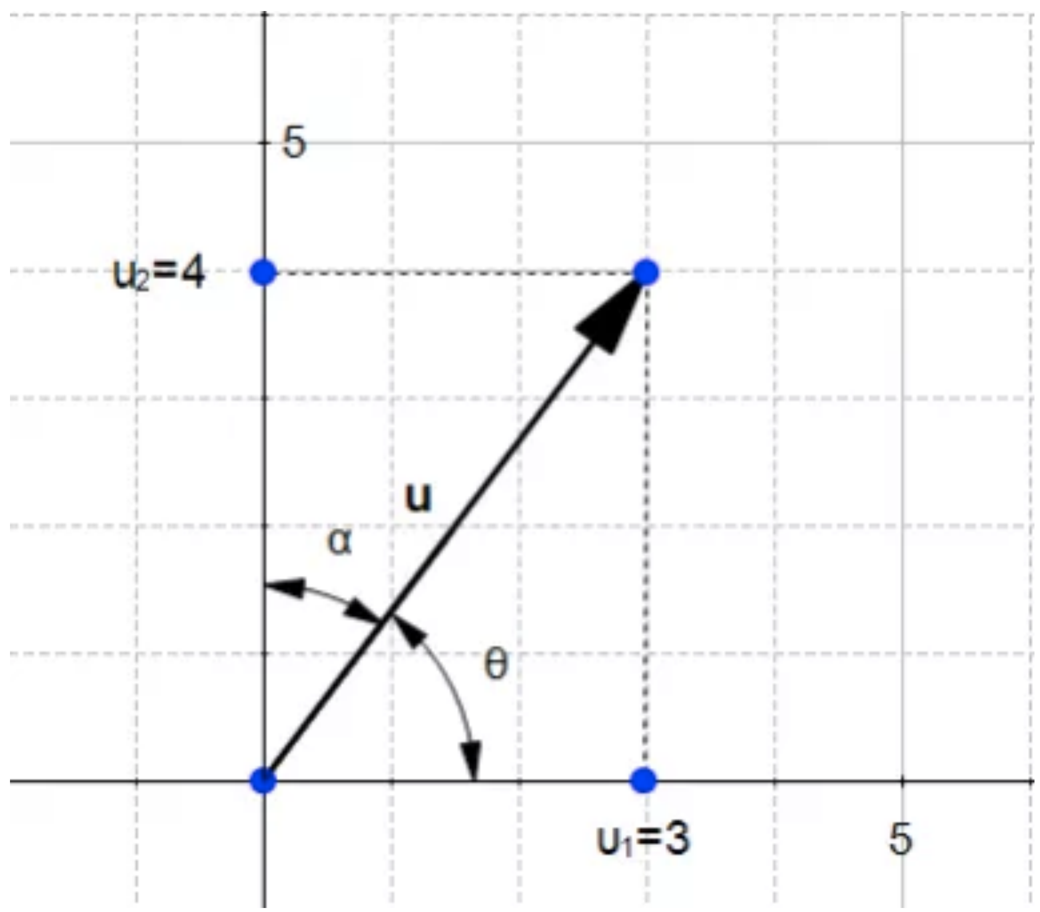
$= \sqrt{u_1^2 + u_2^2} \in \mathbb{R}$

all are same vector $\vec{v} = \langle -2, 5 \rangle$



Vectors only impart magnitude and direction. They don't impart any information about where the quantity is applied. Be careful to distinguish vector notation, from the notation we use to represent coordinates of points, . The vector denotes a magnitude and a direction of a quantity while the point denotes a location in space.

Find the direction



vector u(u1,u2) with u1=3 and u2 =4

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 can use the cosine of the angles.

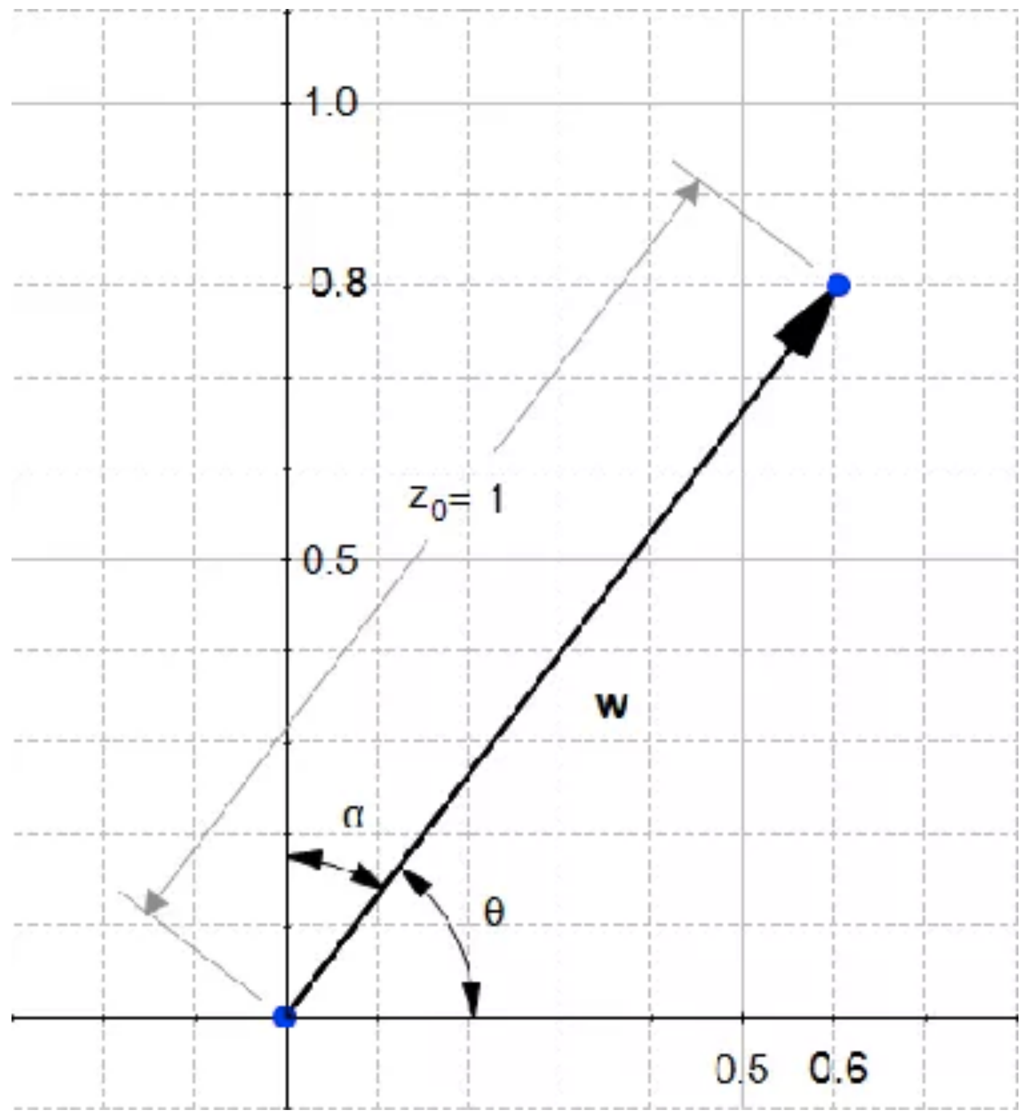
In a right triangle, the cosine of an angle β is defined by

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

we can form two right triangles, and in both case the adjacent side will be on one of the axis. Which means that the definition of the cosine implicitly contains the axis related to an angle. We can rephrase our naïve definition to :

$$cos(\theta) = \frac{u_1}{||u||}$$

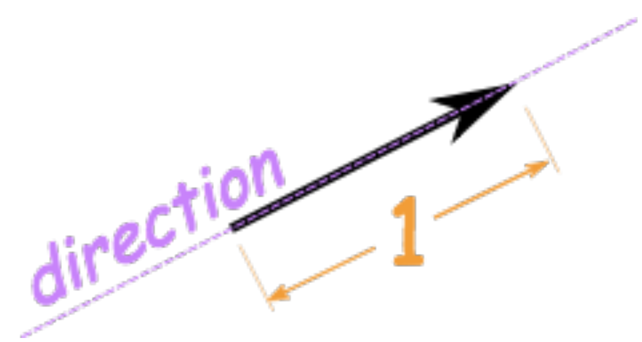
$$cos(\alpha) = \frac{u_2}{||u||}$$



$$cos(\theta) = \frac{u_1}{\|u\|} = \frac{3}{5} = 0.6 \qquad cos(\alpha) = \frac{u_2}{\|u\|} = \frac{4}{5} = 0.8$$

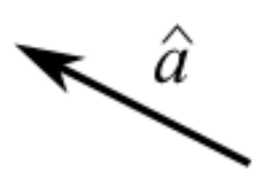
We can see that w as indeed the same look as u except it is smaller. Something interesting about direction vectors like **w** is that their norm is equal to 1. That's why we often call them **unit vectors**.

What is a Unit Vector ?



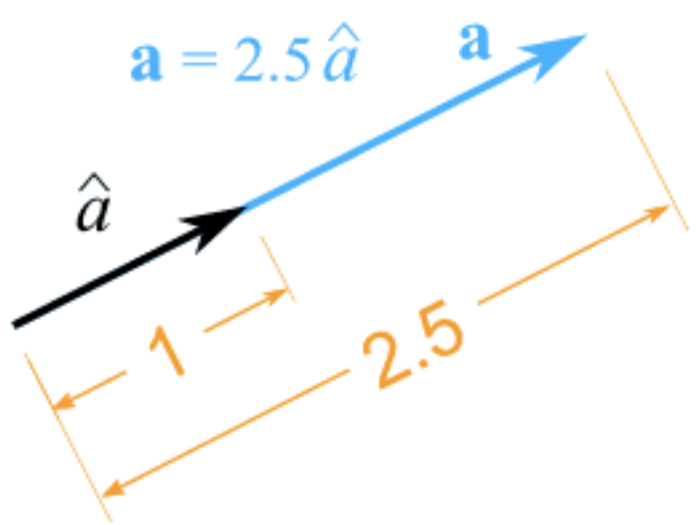
A **Unit Vector** has a magnitude of **1**:. Its only purpose is to describe a direction in space.

The symbol is usually a lowercase letter with a "hat"

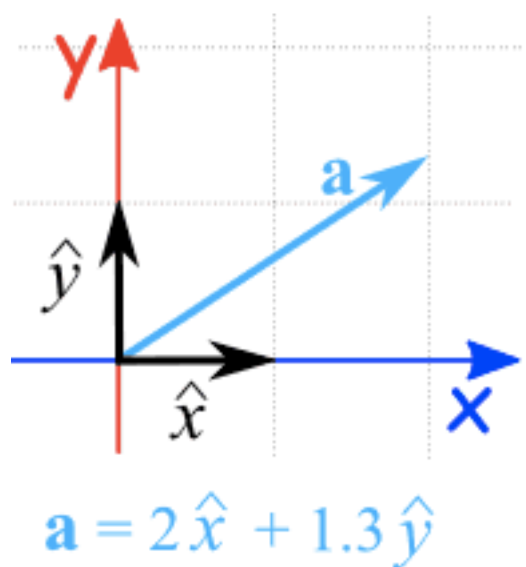


Scaling In 1 Dimensions

A vector can be "scaled" off the unit vector. Here vector **a** is shown to be 2.5 times a unit vector. Notice they still point in the same direction:



Scaling In 2 Dimensions



Here we show that the vector **a** is made up of 2 "x" unit vectors and 1.3 "y" unit vectors.

Let's say we have a vector v notated as v = <a, b, c>.

$$\hat{u} =< \frac{a}{|\vec{v}|}, \frac{b}{|\vec{v}|}, \frac{c}{|\vec{v}|} >$$

Given: Vector v = <-1, 3, -4>. Determine its unit vector.

$$|\vec{v}| = \sqrt{a^2 + b^2 + c^2}$$

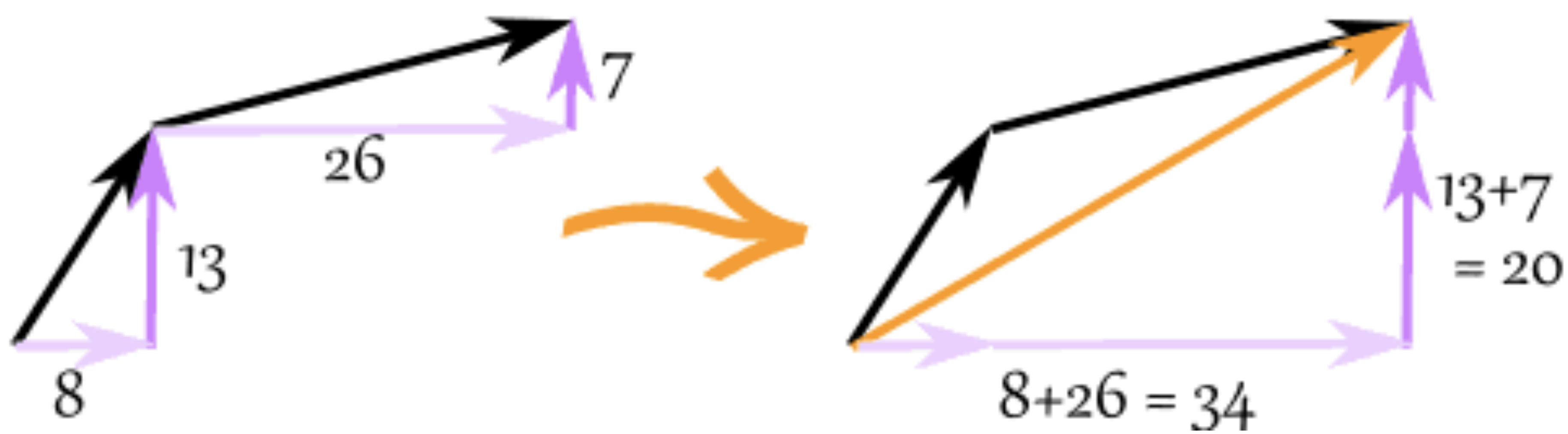
$$|\vec{v}| = \sqrt{(-1)^2 + 3^2 + (-4)^2}$$

$$|\vec{v}| = \sqrt{1 + 9 + 16}$$

$$|\vec{v}| = \sqrt{26}$$

$$\hat{u} =< -\frac{1}{\sqrt{26}}, \frac{3}{\sqrt{26}}, -\frac{4}{\sqrt{26}} >$$

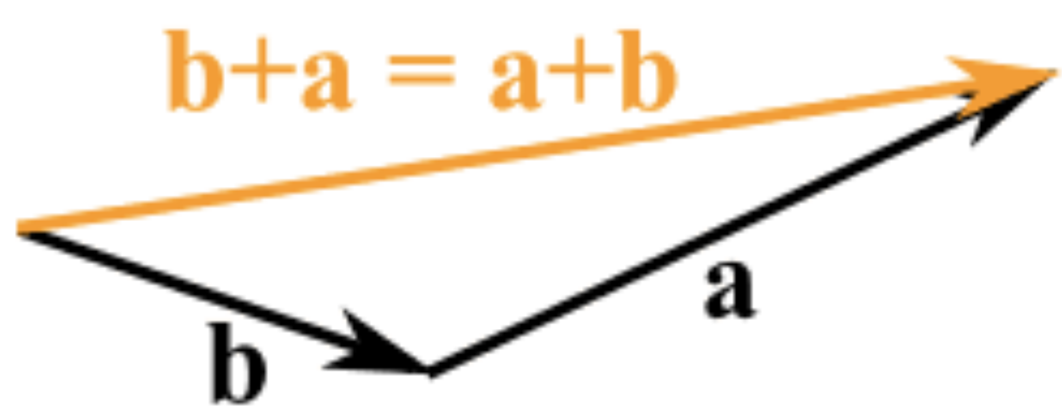
Adding Vectors



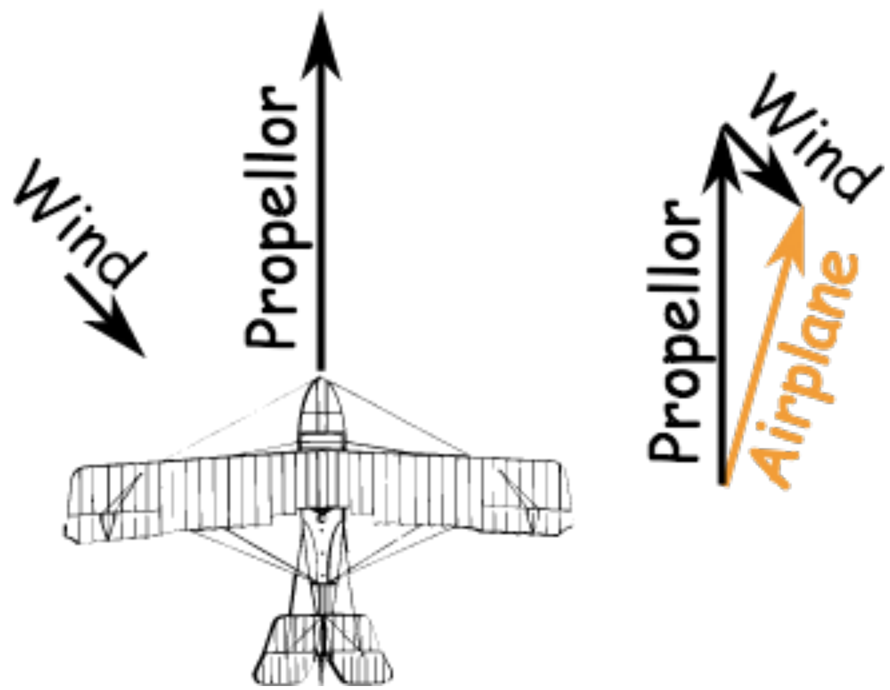
The vector (8,13) and the vector (26,7) add up to the vector (34,20)

add two vectors

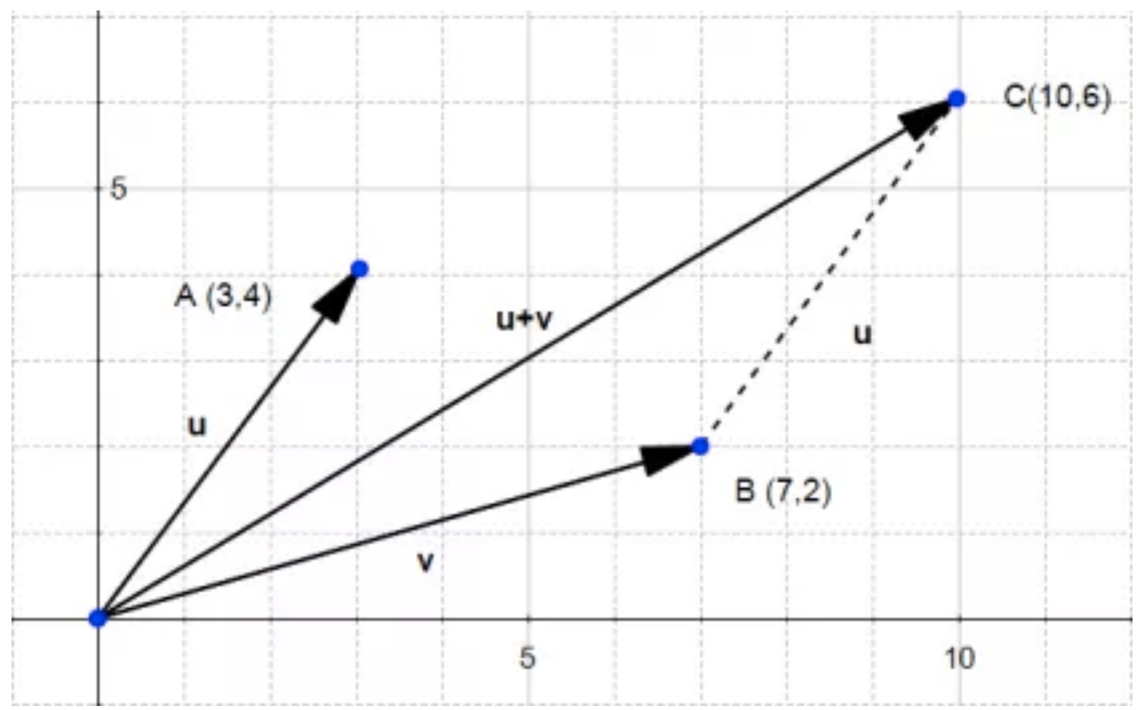
And it doesn't matter which order we add them, we get the same result:



A plane is flying along, pointing North, but there is a wind coming from the North-West.

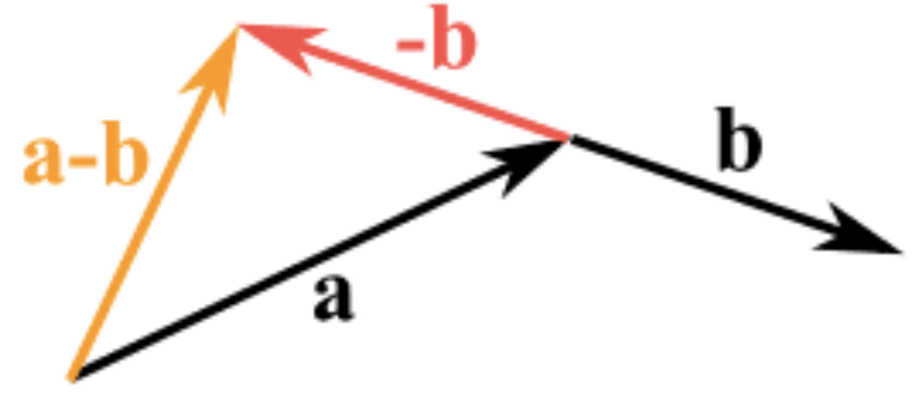


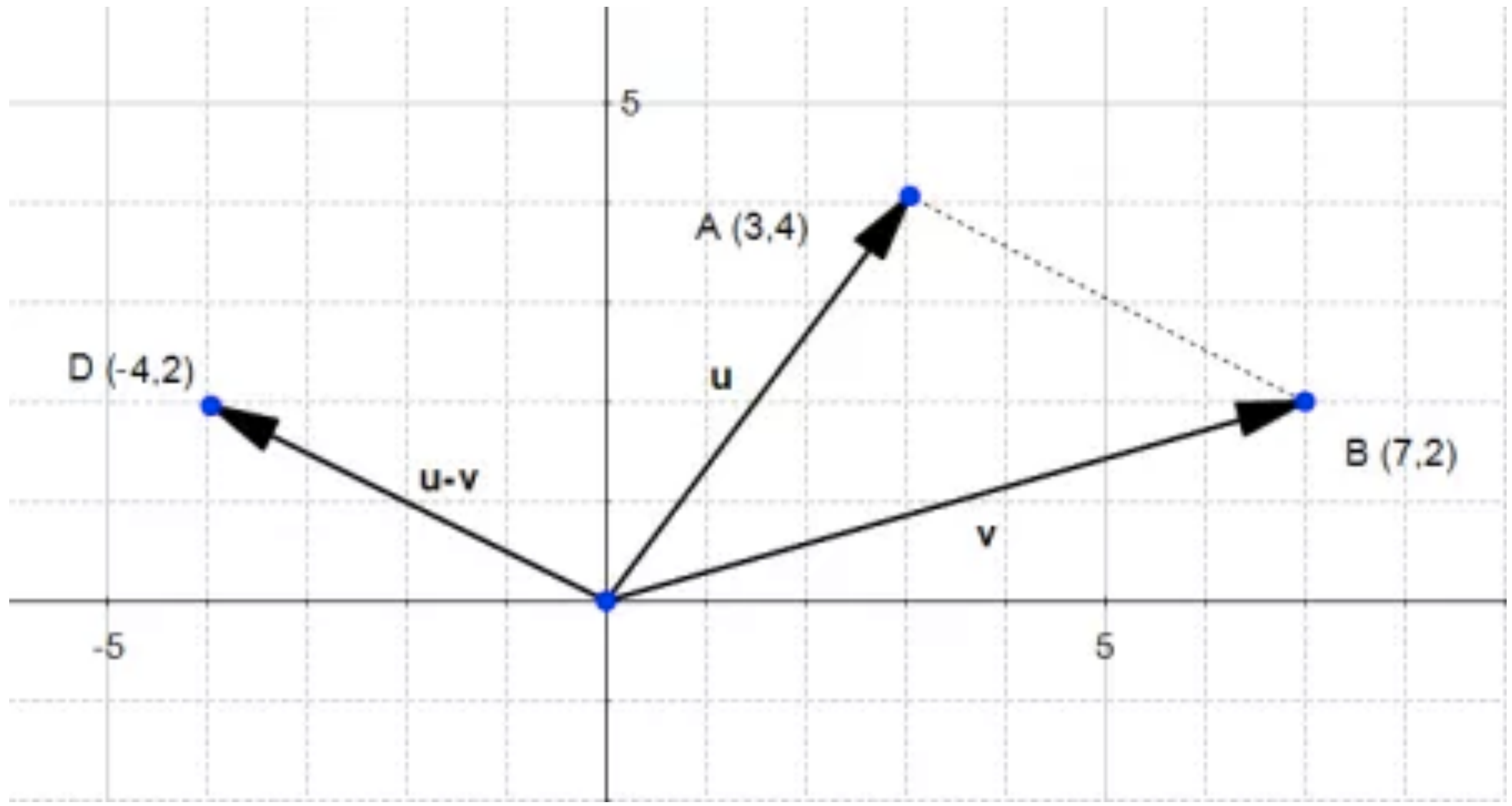
The two vectors (the velocity caused by the propeller, and the velocity of the wind) result in a slightly slower ground speed heading a little East of North.



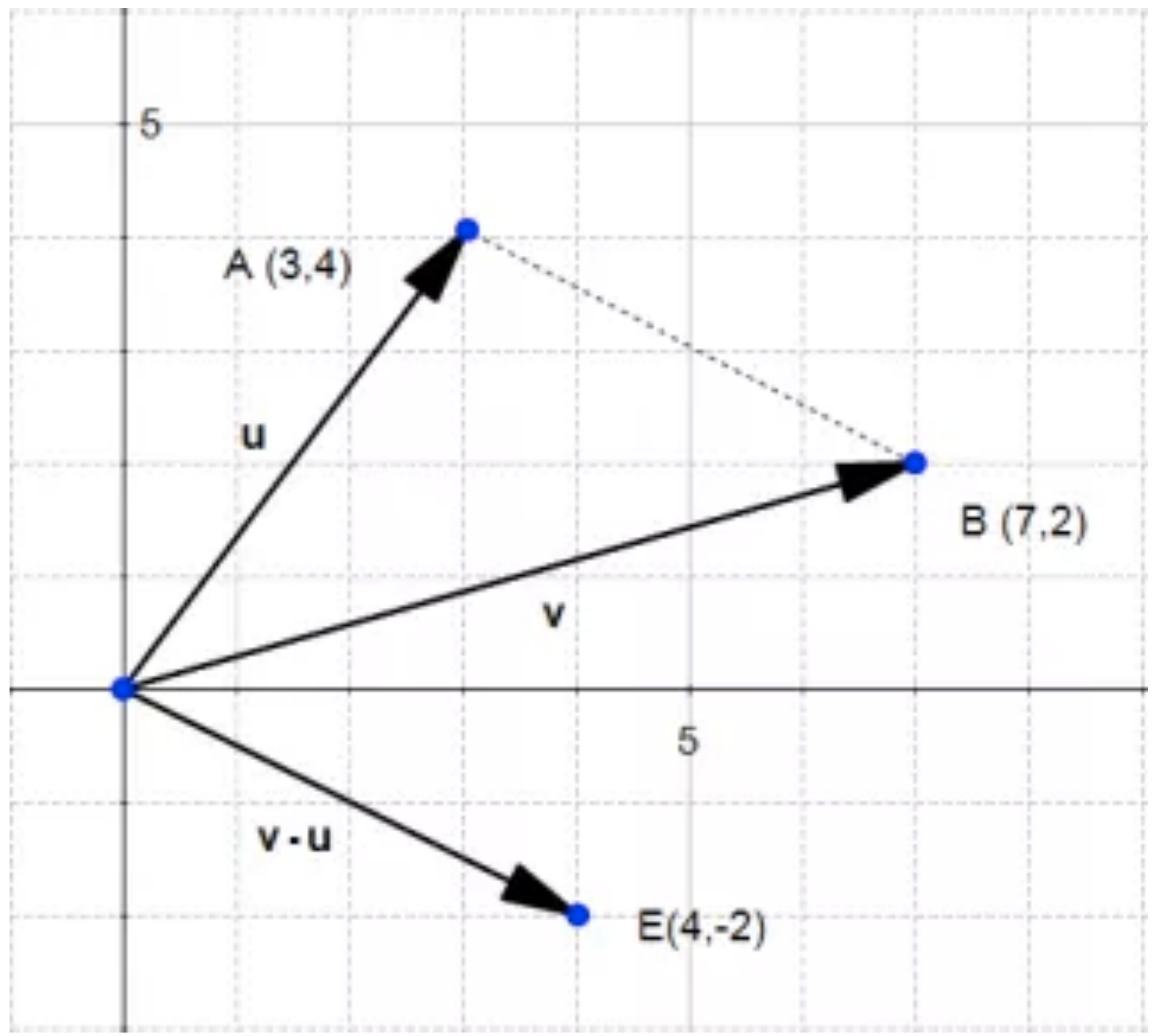
u + v = (u1 + v1, u2 + v2)

Subtracting

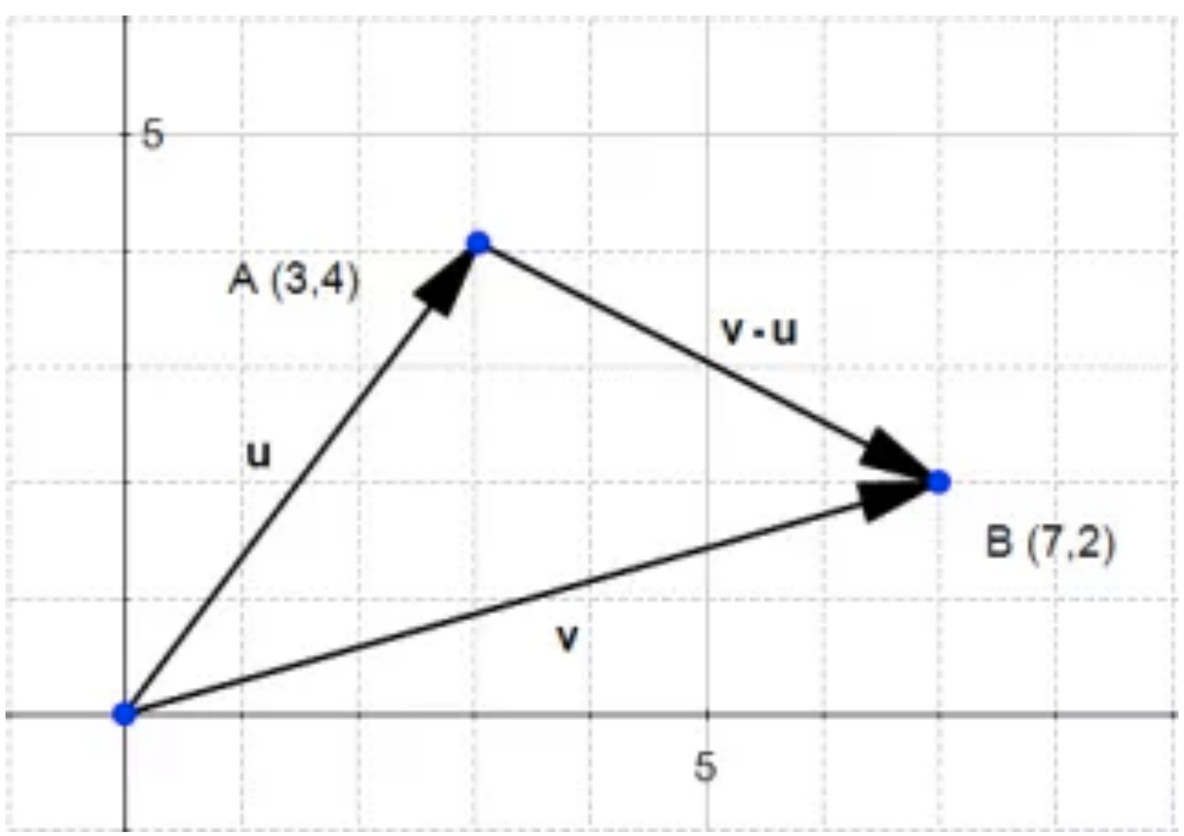


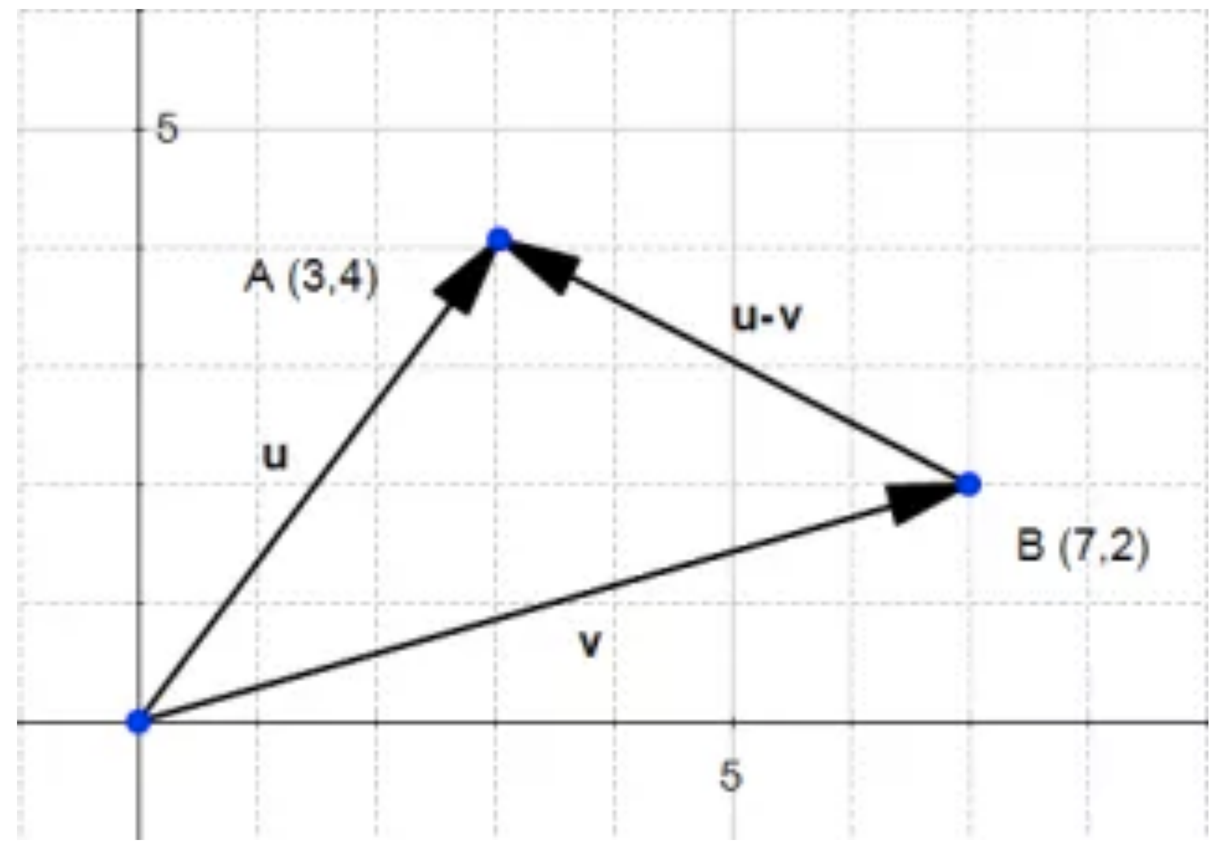


u - v = (u1 - v1, u2 - v2)



v - u = (v1 - u1, v2 - u2)





However, since a vector has a magnitude and a direction, we often consider that parallel translate of a given vector (vectors with the same magnitude and direction but with a different origin) are the same vector, just drawn in a different place in space.

Vector Multiplication

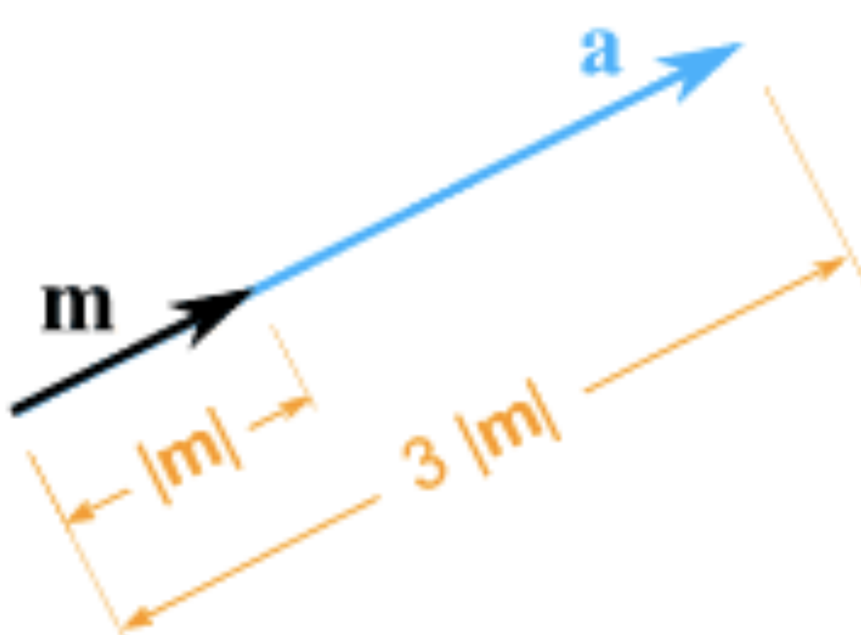
There are three different types of multiplication: dot product, cross product, and multiplication of vector by a scalar

The dot product of two vectors u and v is given as u . v = uv cos θ where θ is the angle between the vectors u and v

The cross product of two vectors u and v is given as u × v = uv sin θ where θ is the angle between the vectors u and v

When a vector is multiplied by a scalar, only the magnitude of the vector is changed, but the direction remains the same

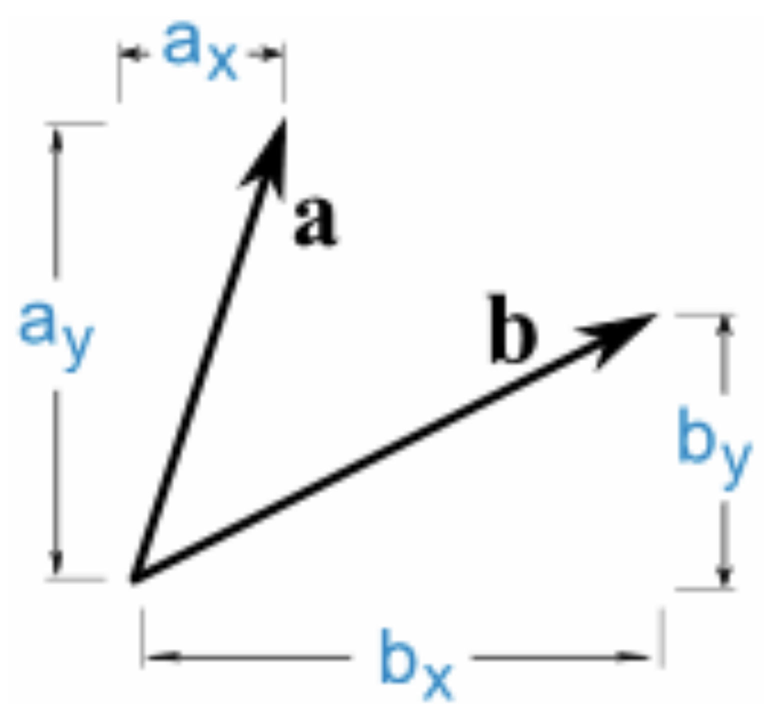
Multiplying a Vector by a Scalar

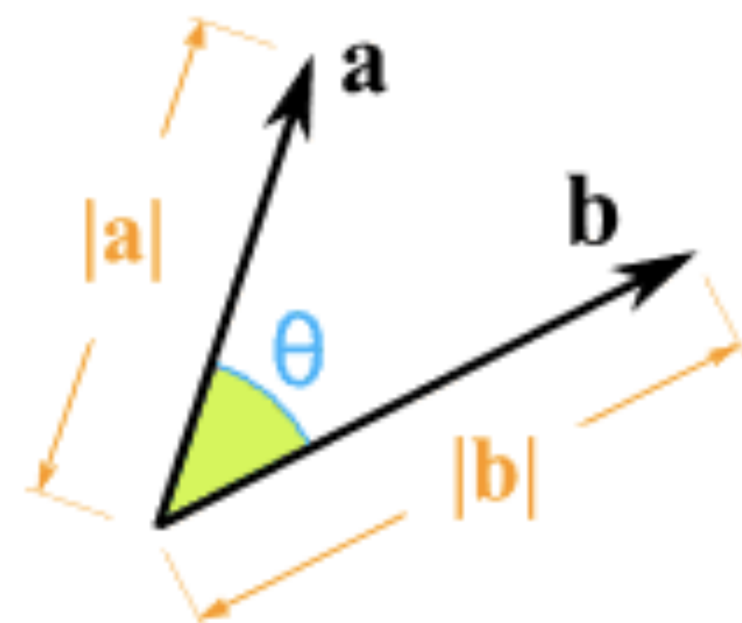


$a = 3m = (3\times7,3\times3) = (21,9)$

multiply the vector m = (7,3) by the scalar 3.
It still points in the same direction, but is 3 times longer

Dot Product

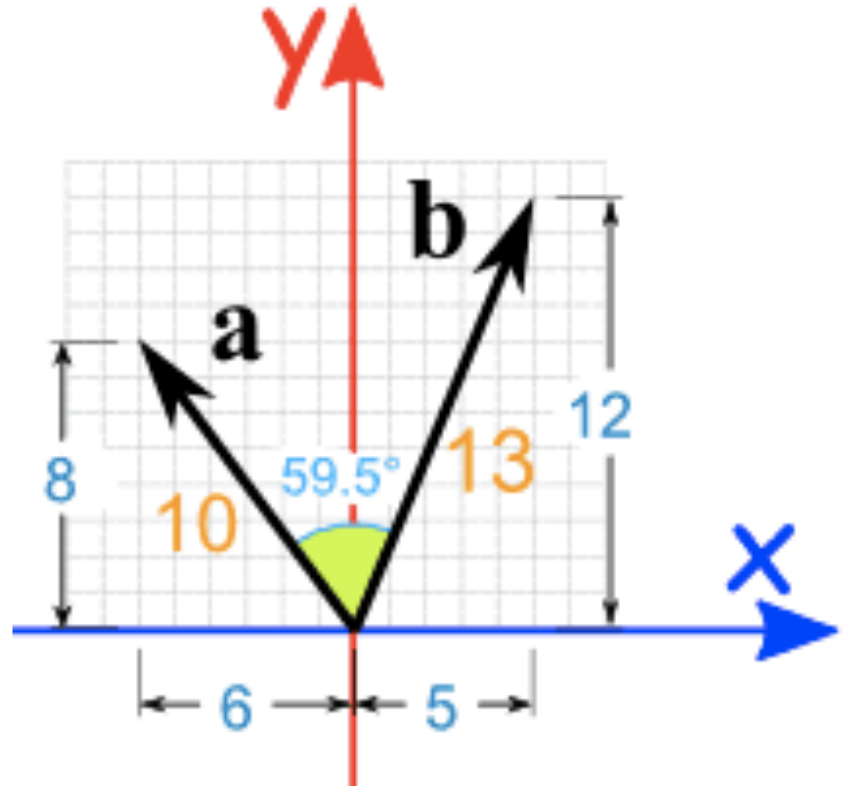




$\mathbf{a} \cdot \mathbf{b} = a_x \times b_x + a_y \times b_y$

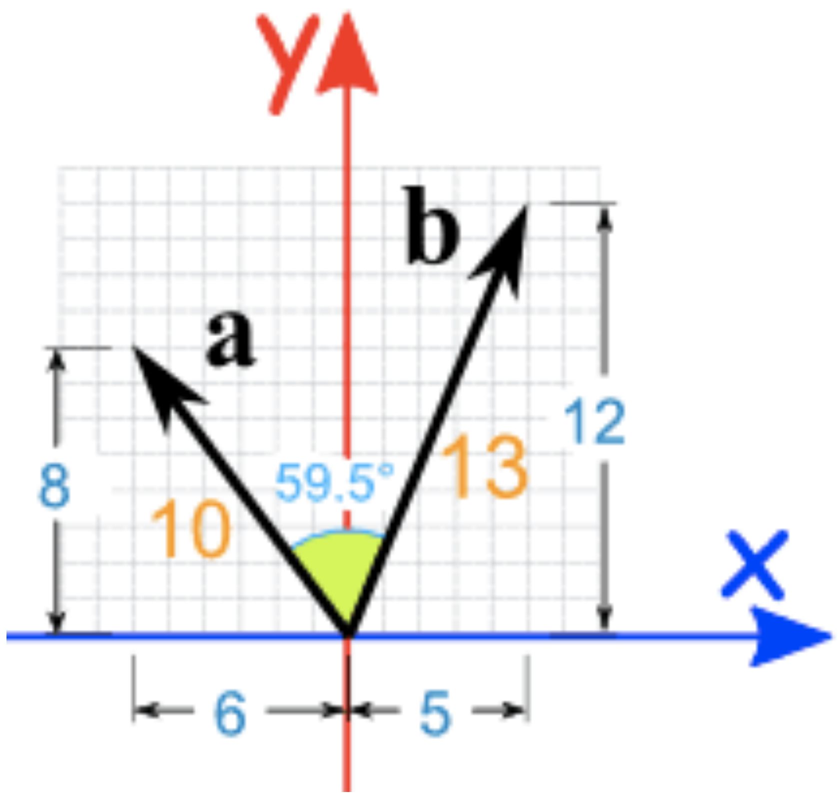
$a \cdot b = |a| \times |b| \times \cos(\theta)$

2 ways to calculate dot product, Both methods work!



$$a \cdot b = |a| \times |b| \times \cos(\theta)$$

$$a \cdot b = 10 \times 13 \times \cos(59.5°)$$
$$a \cdot b = 10 \times 13 \times 0.5075...$$
$$a \cdot b = 65.98... = 66 \text{ (rounded)}$$



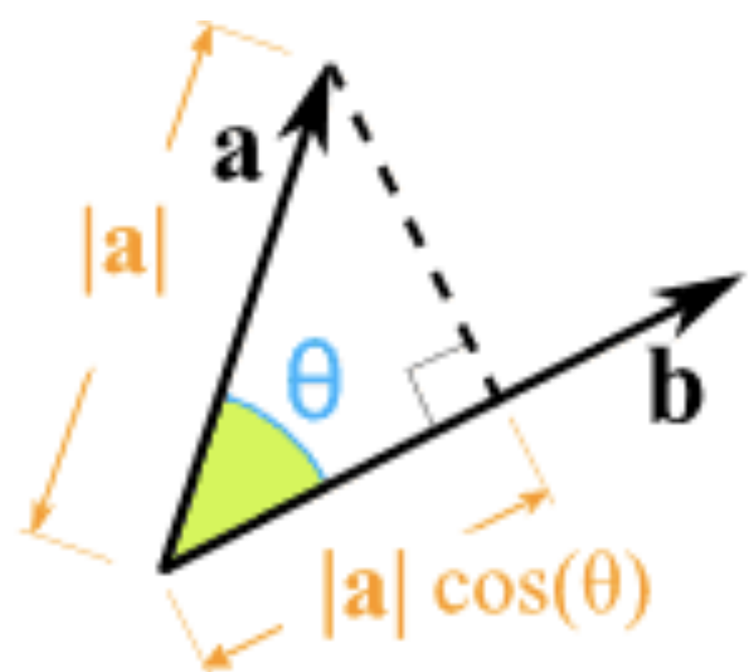
$$\mathbf{a} \cdot \mathbf{b} = a_x \times b_x + a_y \times b_y$$

$$\mathbf{a} \cdot \mathbf{b} = -6 \times 5 + 8 \times 12$$
$$\mathbf{a} \cdot \mathbf{b} = -30 + 96$$
$$\mathbf{a} \cdot \mathbf{b} = 66$$

Both methods came up with the same result (after rounding)

Also note that we used **minus 6** for a_x (it is heading in the negative x-direction)

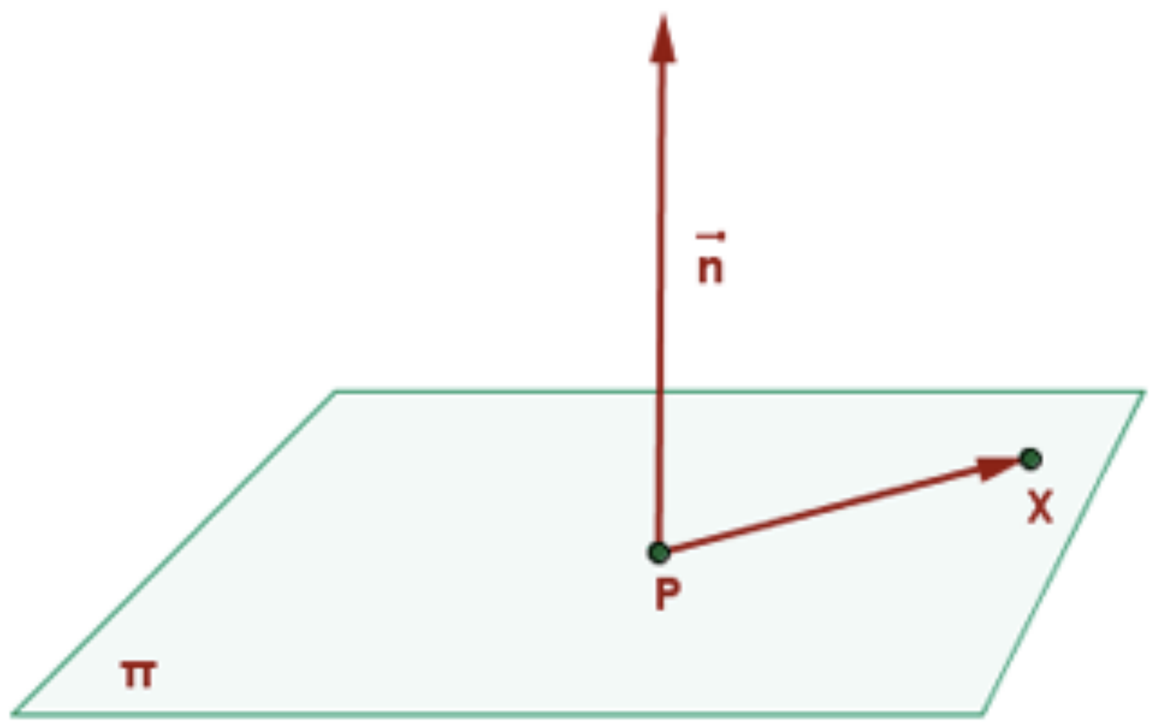
Why cos(θ) ?



OK, to multiply two vectors it makes sense to multiply their lengths together but only when they point in the same direction. we make one "point in the same direction" as the other by multiplying by cos(θ):

What is Normal Vector ?

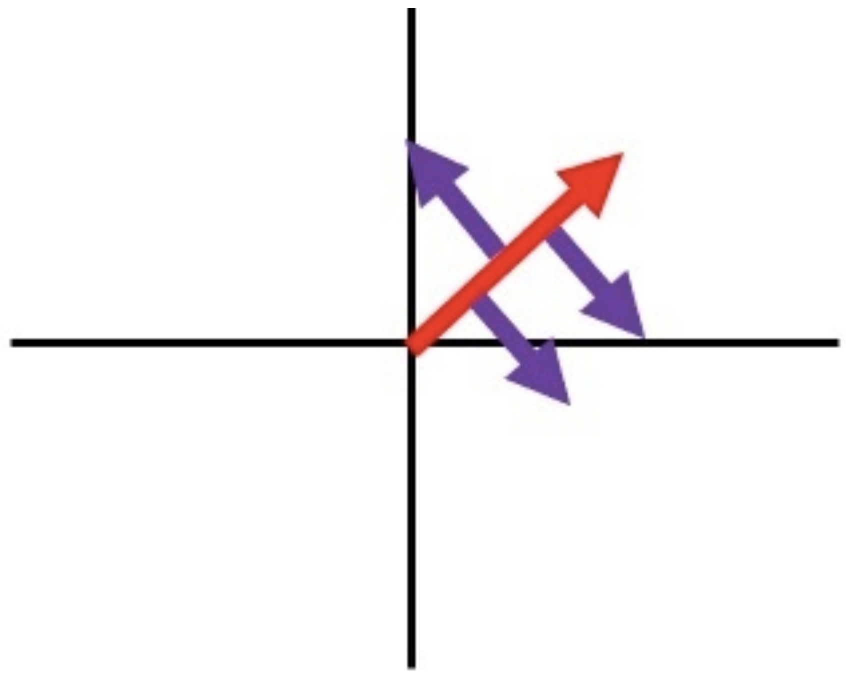
The vector is a normal vector to the plane if it is perpendicular to the plane



its dot product is zero.

$$\overrightarrow{PX} \cdot \vec{n} = 0$$

Technically, there is an infinite number of normal vectors to any vector because the only criteria for a normal vector is that is 90° to the original vector.

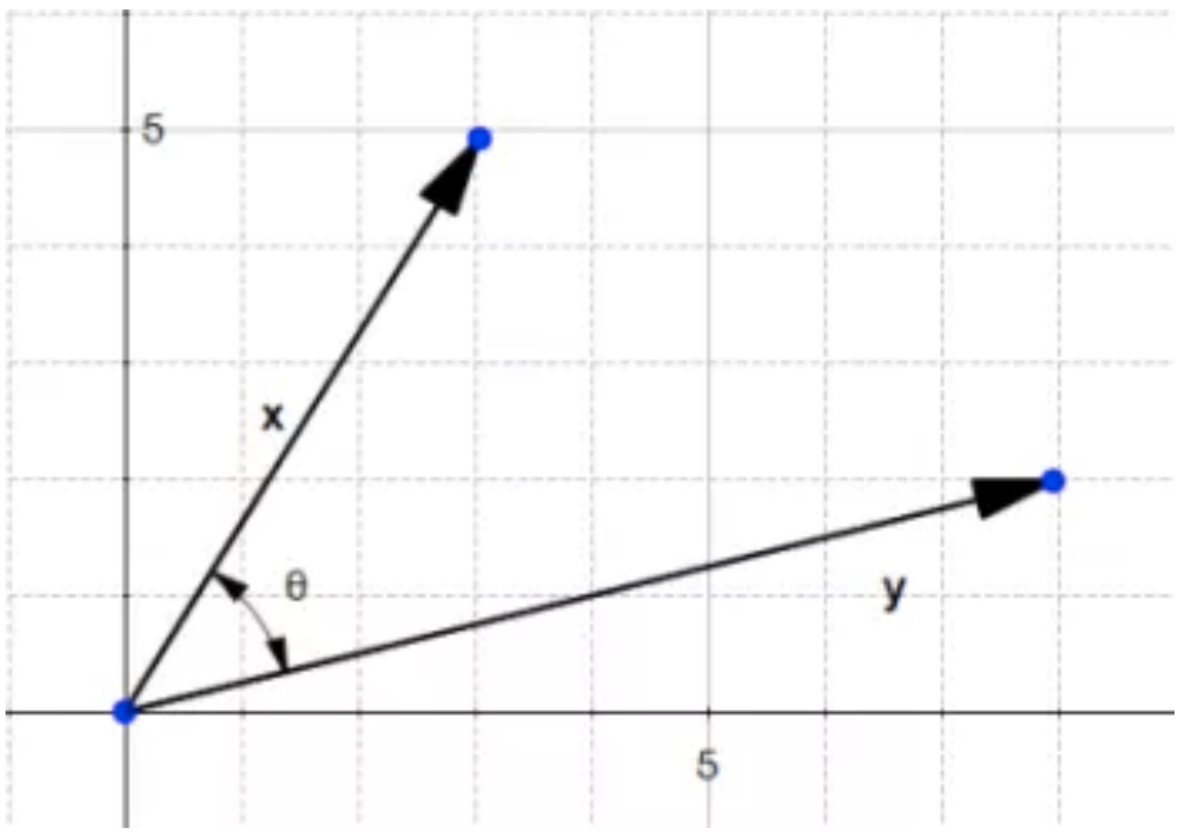


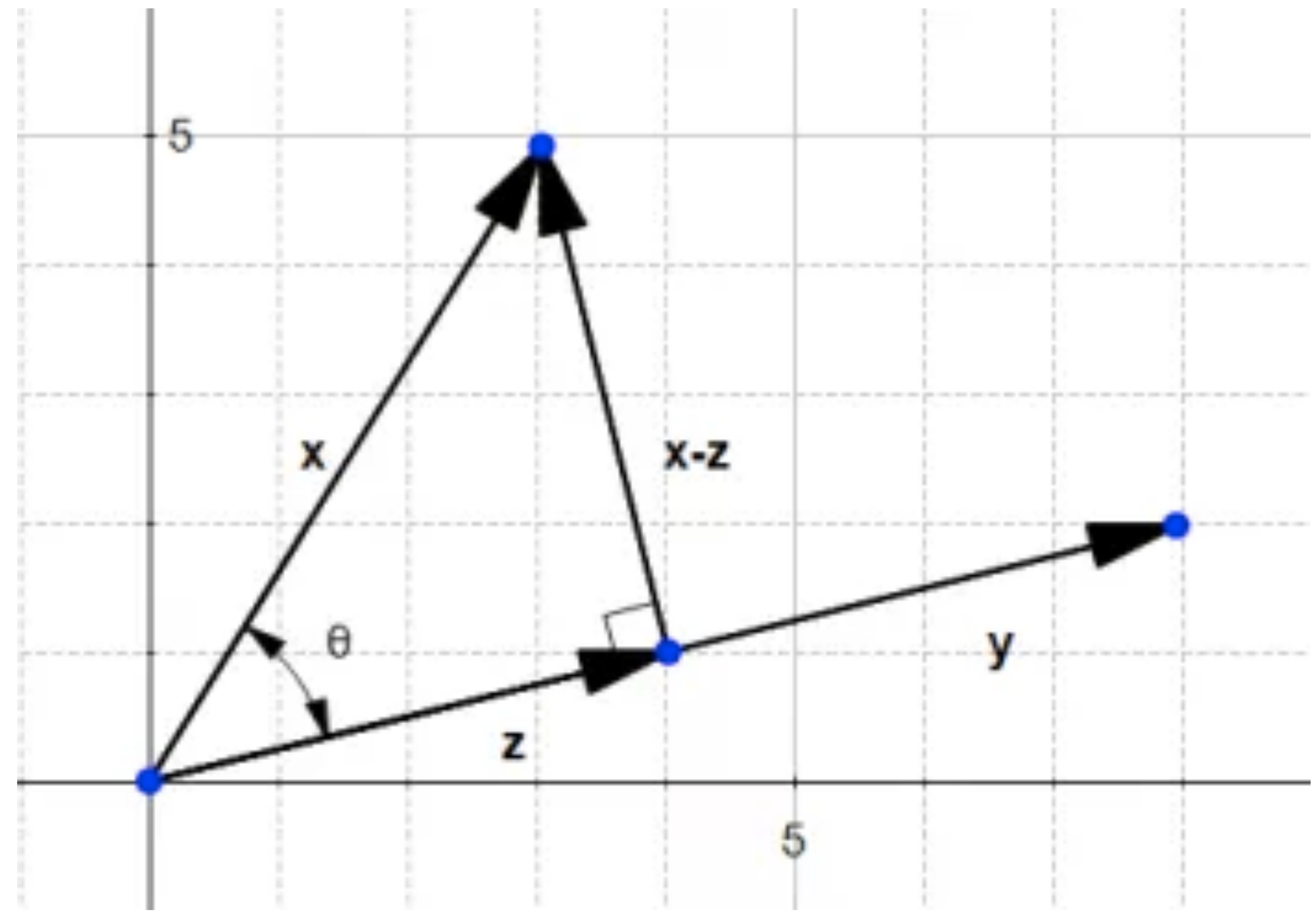
Since the angle between a given vector and any normal vector is 90°, the right side of this equation is 0 because the cosine of 90° is 0.

$$\vec{a} \bullet \vec{b} = 0$$

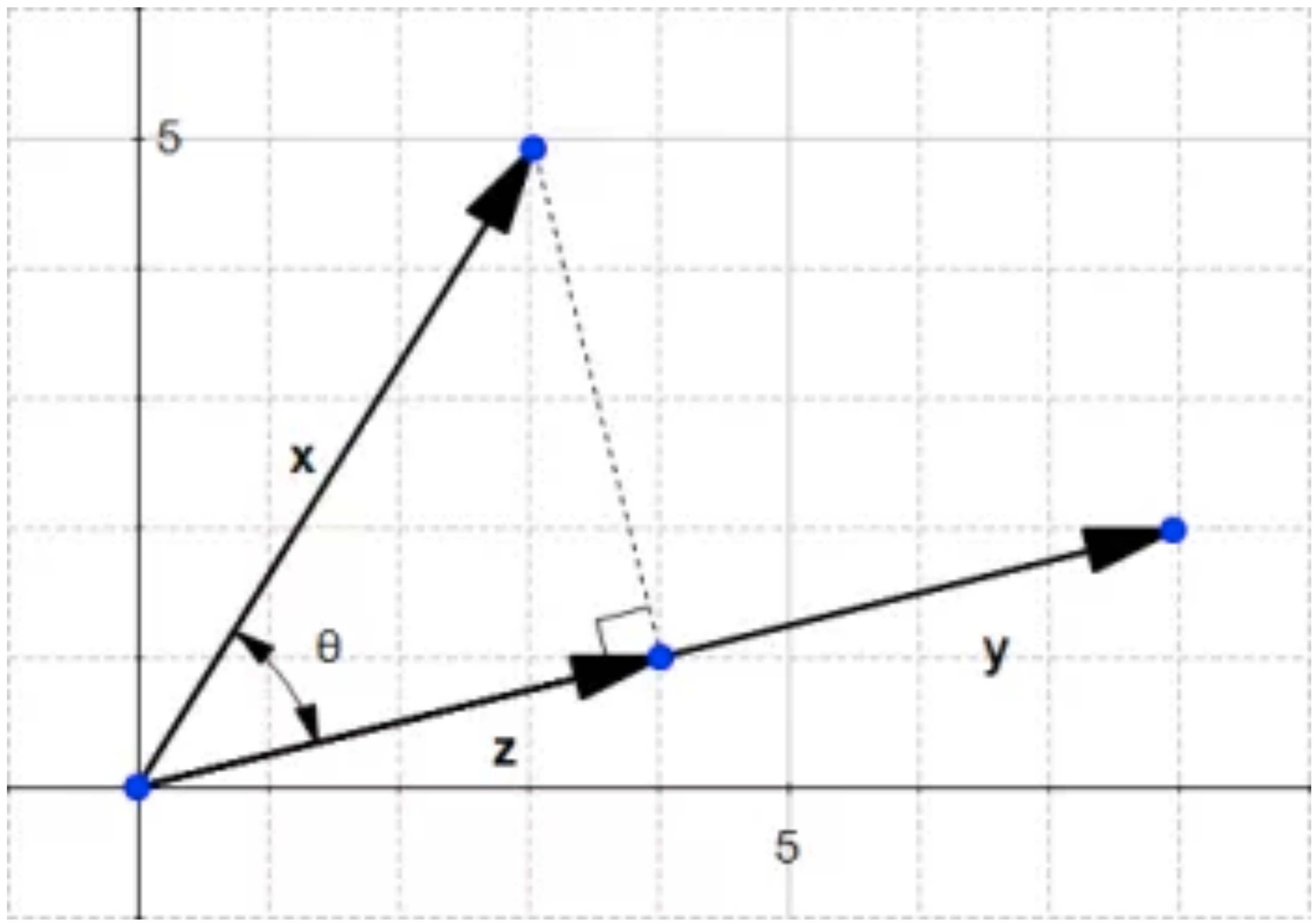
The dot product between a given vector and the normal vector equals zero

Orthogonal Projection of a Vector





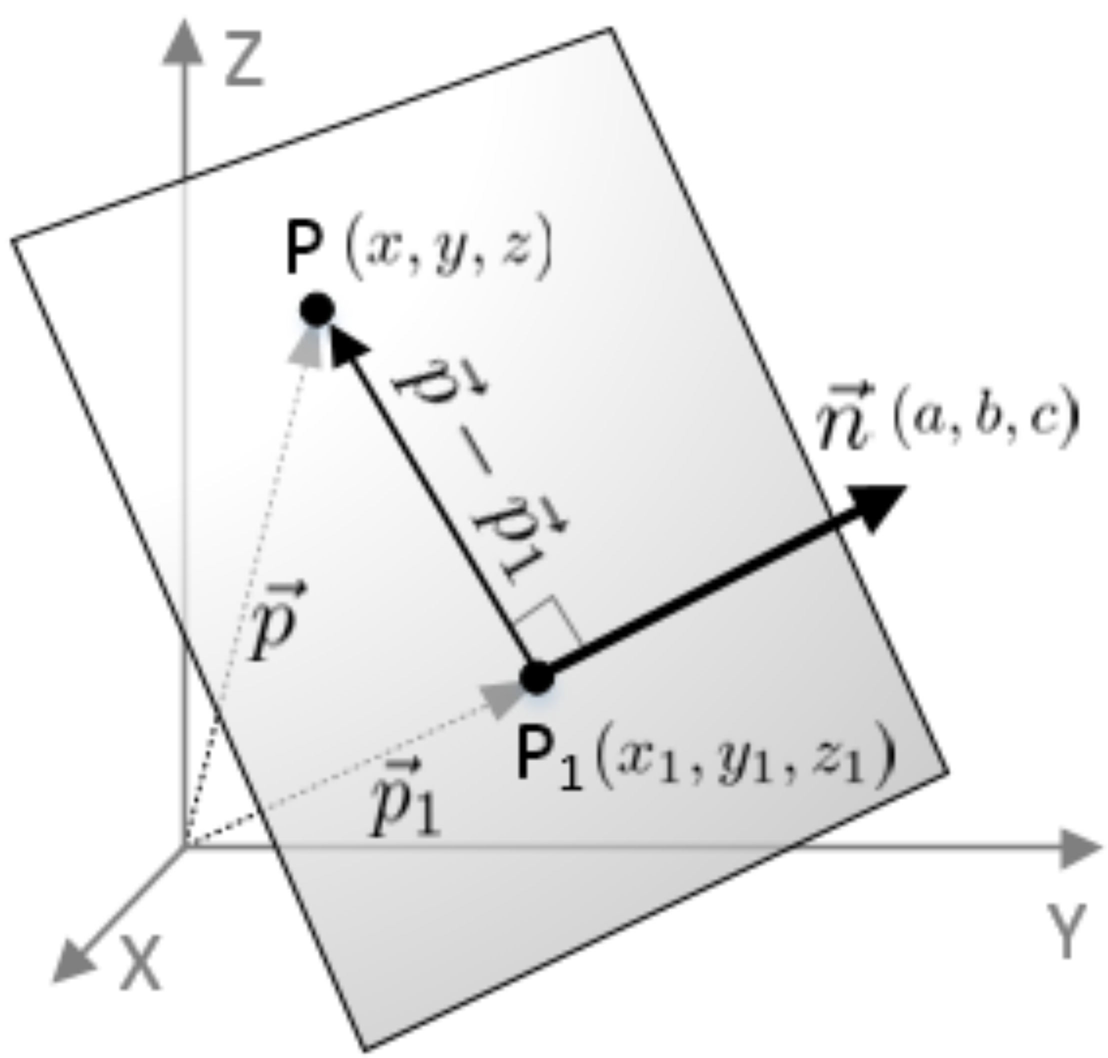
Given two vectors we would like to find the orthogonal projection of X on tp Y.



To do this we project the vector x on to y. This gives us the vector Z.

Equation of Hyperplane

The equation of a plane in 3D space is defined with normal vector (perpendicular to the plane) and a known point on the plane.



Let the normal vector of a plane n., and the known point on the plane, P_1. And, let any point on the plane as P.

We can define a vector connecting from P_1 to P, which is lying on the plane.

$$\vec{p} - \vec{p}_1 = (x - x_1,\ y - y_1,\ z - z_1)$$

Since the vector $\vec{p} - \vec{p}_1$ and the normal vector $\vec{n}$ are perpendicular each other, the dot product of two vector should be 0

$$\vec{n} \cdot (\vec{p} - \vec{p}_1) = 0 \qquad (\because \vec{n} \perp (\vec{p} - \vec{p}_1))$$

This dot product of the normal vector and a vector on the plane becomes the equation of the plane. By calculating the dot product, we get

$$(a, b, c) \cdot (x - x_1, y - y_1, z - z_1) = 0$$
$$a(x - x_1) + b(y - y_1) + c(z - z_1) = 0$$
$$ax + by + cz - (ax_1 + by_1 + cz_1) = 0$$

If we substitute the constant terms to , $d = -(ax_1 + by_1 + cz_1)$

then the plane equation becomes simpler;

$$ax + by + cz + d = 0$$

Distance of a Point from a Hyperplane

The shortest distance from an arbitrary point P_2 to a plane can be calculated by the dot product of two vectors

$(\vec{p}_2 - \vec{p}_1)$ and $\vec{n}$

, projecting the vector $(\vec{p}_2 - \vec{p}_1)$ to the normal vector $\vec{n}$ of the plane.

The distance D between a plane $ax + by + cz + d = 0$ and a point P_2 becomes

$$
\begin{aligned}
D &= |\vec{p}_2 - \vec{p}_1| \cos\theta \\
&= \frac{\vec{n} \cdot (\vec{p}_2 - \vec{p}_1)}{|\vec{n}|} \\
&= \frac{\vec{n} \cdot (\vec{p}_2 - \vec{p}_1)}{\sqrt{a^2 + b^2 + c^2}}
\end{aligned}
$$

$$\because \vec{n} \cdot (\vec{p}_2 - \vec{p}_1) = |\vec{n}||\vec{p}_2 - \vec{p}_1| \cos\theta$$

$$
\begin{aligned}
\vec{n} \cdot (\vec{p}_2 - \vec{p}_1) &= (a, b, c) \cdot (x_2 - x_1, y_2 - y_1, z_2 - z_1) \\
&= ax_2 + by_2 + cz_2 - (ax_1 + by_1 + cz_1)
\end{aligned}
$$

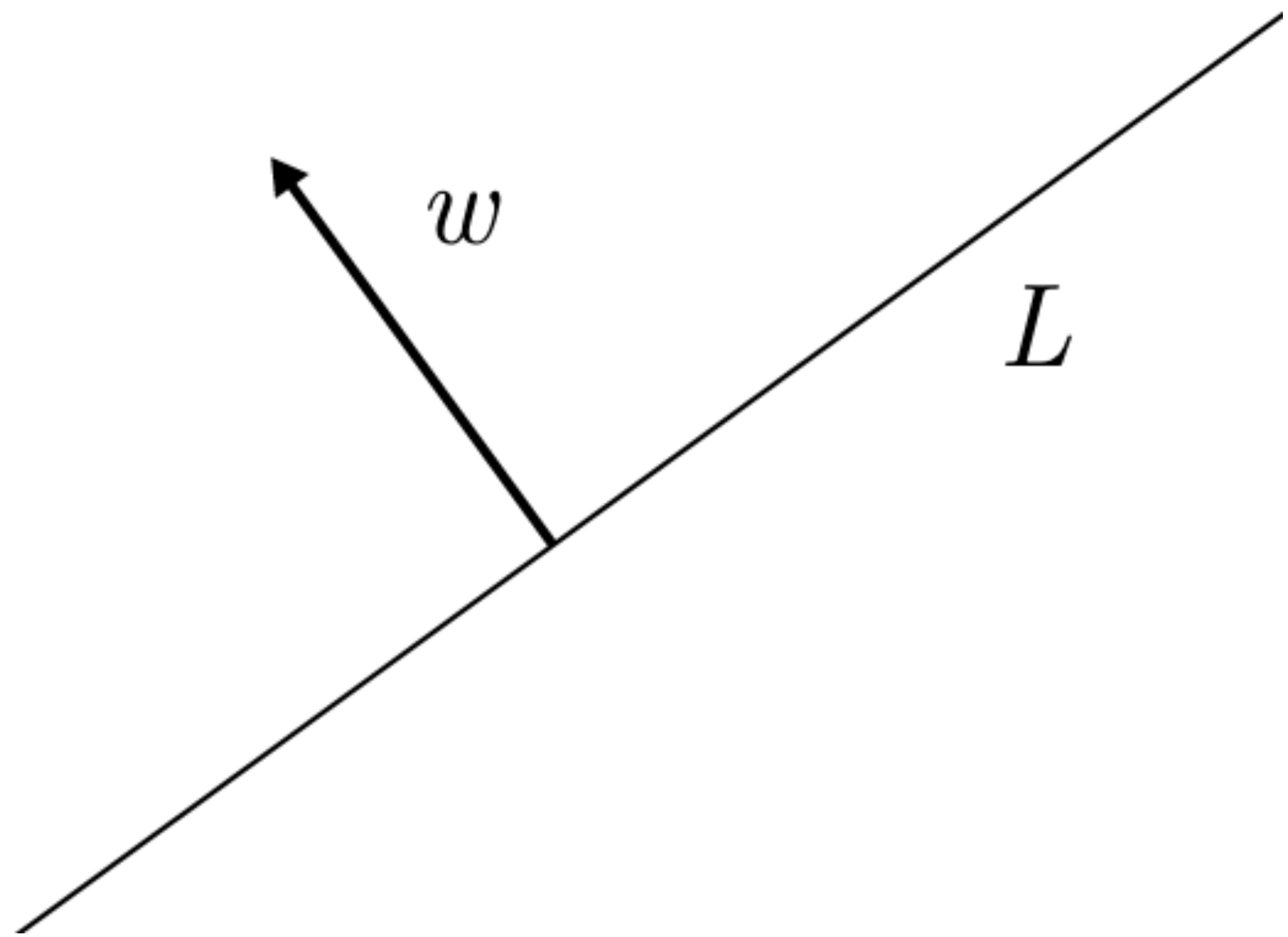
$$
D = \frac{ax_2 + by_2 + cz_2 - (ax_1 + by_1 + cz_1)}{\sqrt{a^2 + b^2 + c^2}}
$$

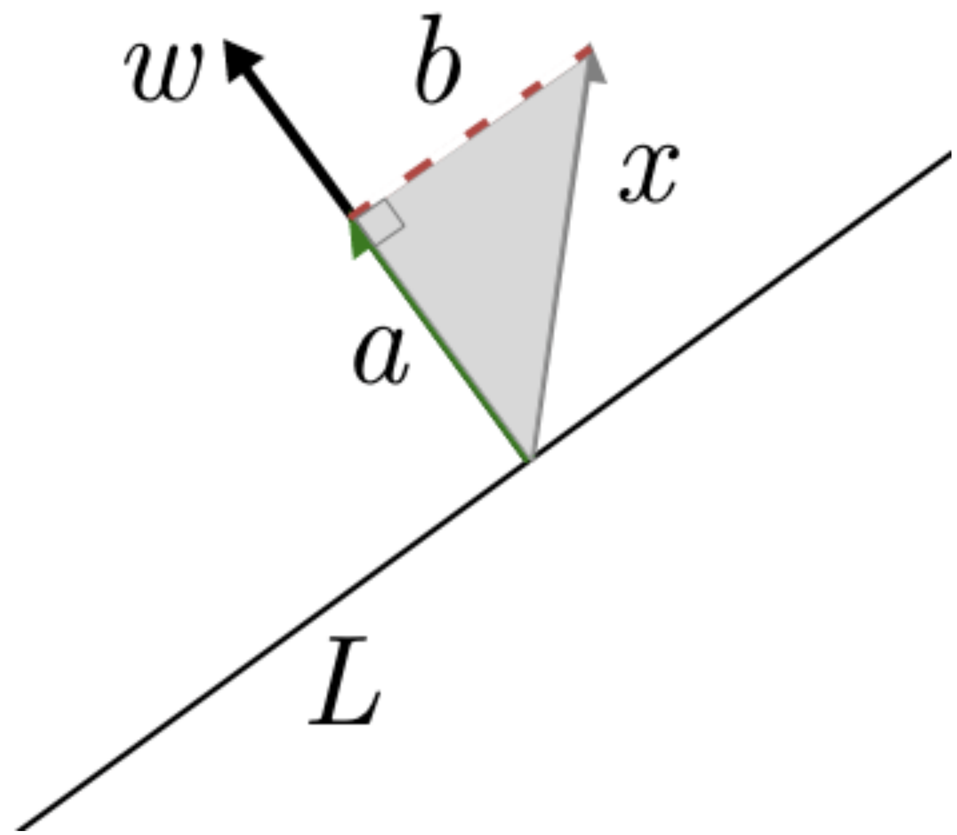
For example, the distance from a point (-1, -2, -3) to a plane x + 2y + 2z - 6 = 0 is;

$$
\begin{aligned}
D &= \frac{1x + 2y + 2z - 6}{\sqrt{1^2 + 2^2 + 2^2}}, \qquad P_2 = (-1, -2, -3) \\
&= \frac{1(-1) + 2(-2) + 2(-3) - 6}{\sqrt{1^2 + 2^2 + 2^2}} \\
&= \frac{-1 - 4 - 6 - 6}{3} = \frac{-17}{3}
\end{aligned}
$$

Notice this distance is signed; can be negative value. It is useful to determine the direction of the point. For example, if the distance is positive, the point is in the same side where the normal is pointing to. And, a negative distance means the point is in opposite side.

we have a line passing through the origin, with W being a unit vector perpendicular to L (“the normal” to the line).

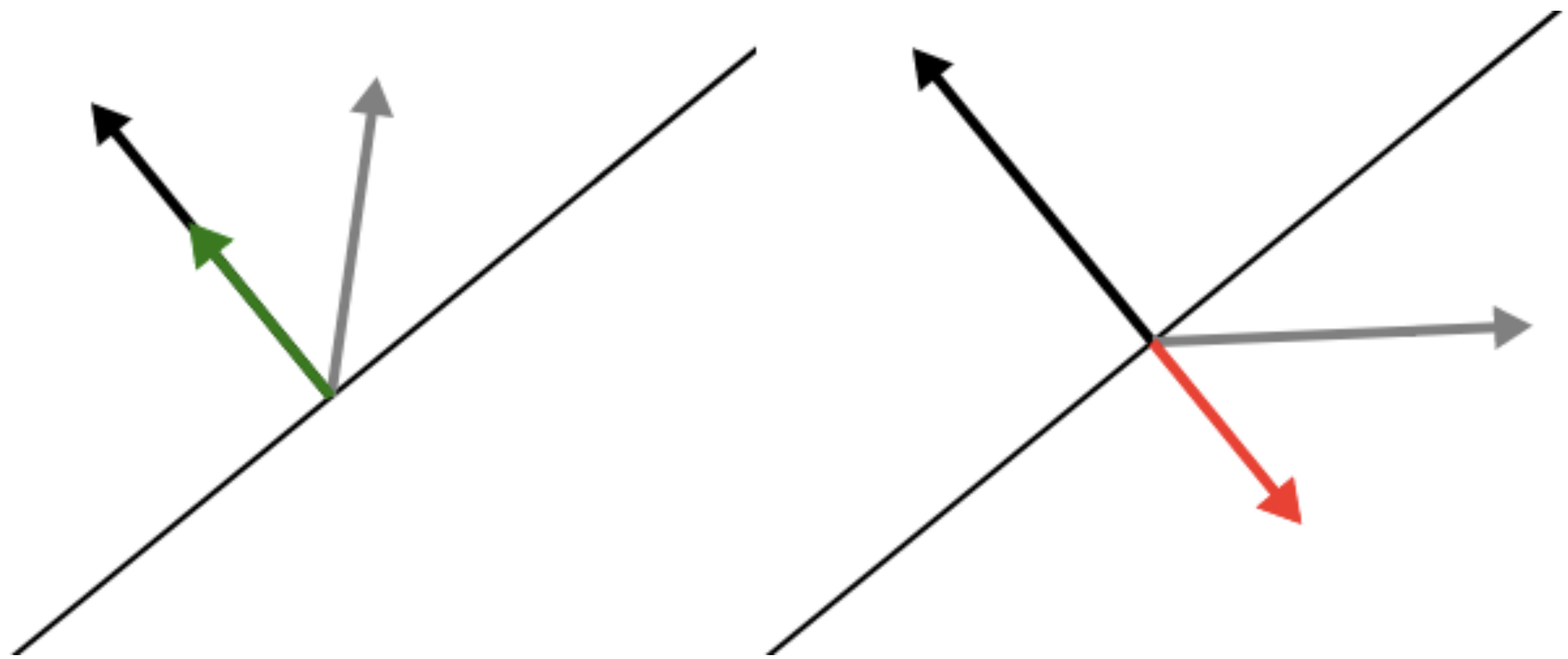




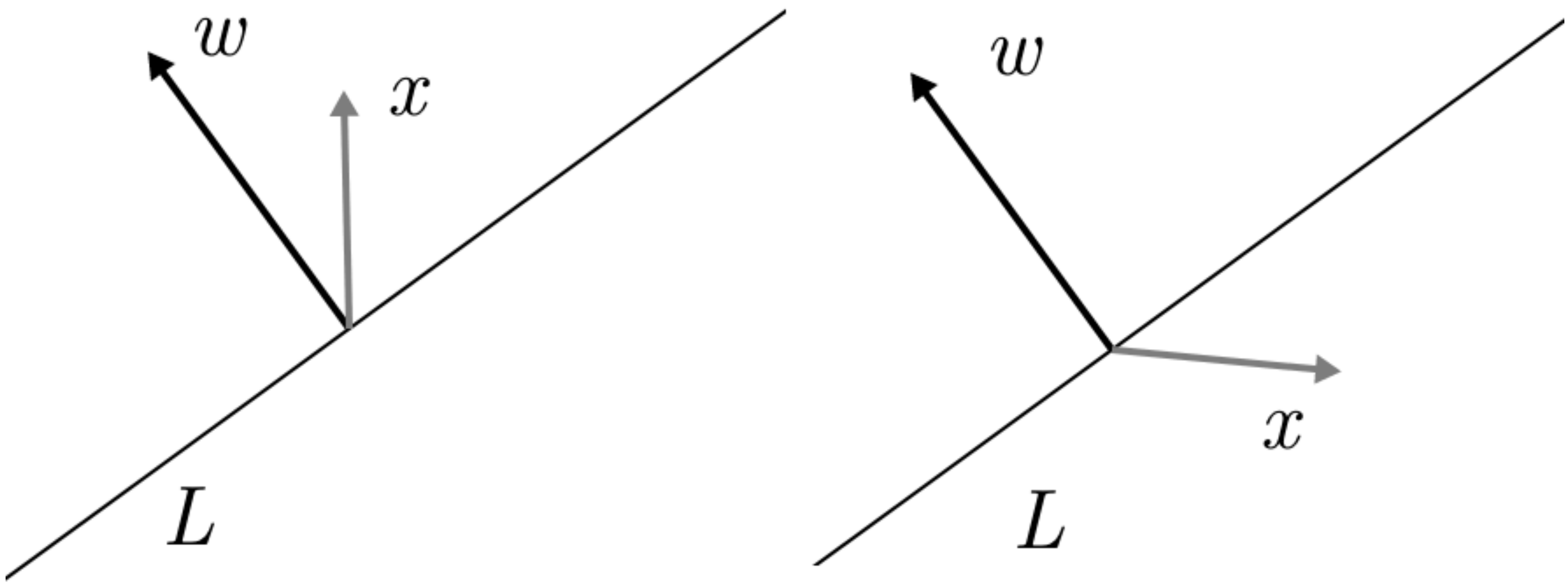
if W is a unit vector, then the length of a—that is, the length of the projection of X onto W is exactly the inner product of (X,W).

The projection of x onto w is a .
note : x = a + b.

if the angle between X and W is larger than 90 degrees, the projected vector will point in the opposite direction of W , so it’s really a “signed” length.



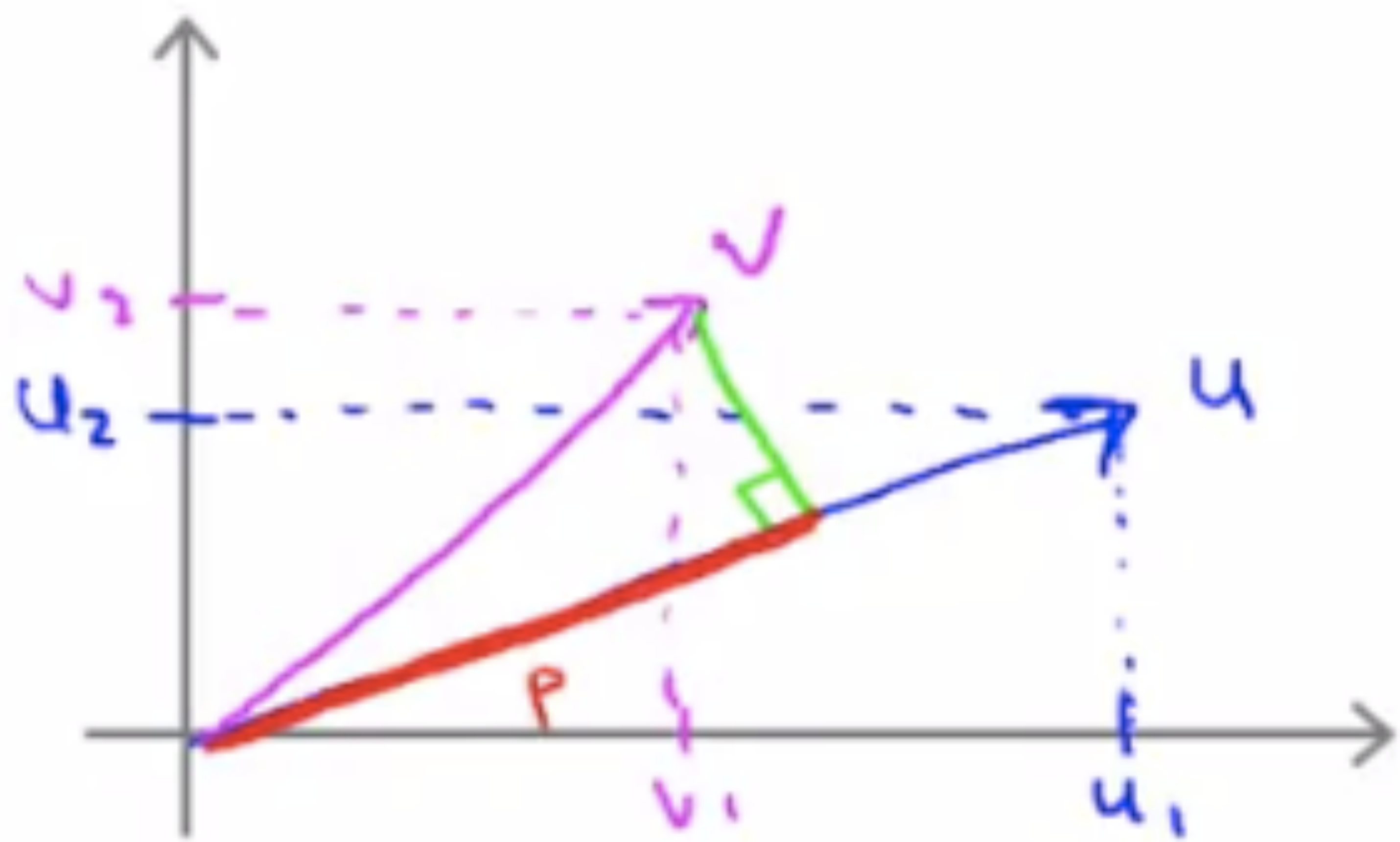
If you take any vector X, then the dot product (X,W) is positive if is on the same side of as , and negative otherwise. The dot product is zero if and only if is exactly on the line , including when is the zero vector.



Demo

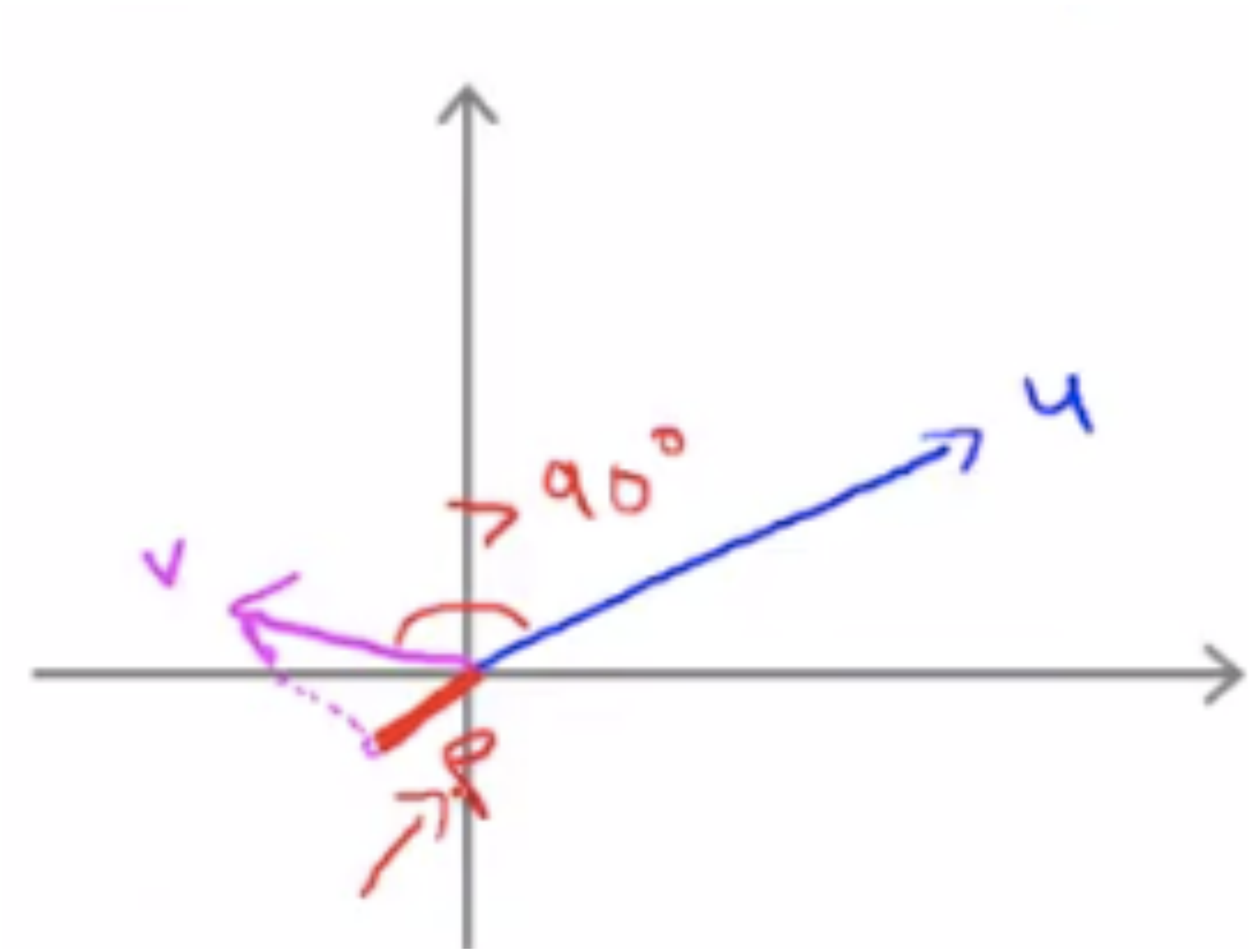
http://j2kun.github.io/decision-rule/index.html

compute the inner product between U and V. Here's how you can do it. Take the vector V and project it down onto the vector U. So I'm going to take an orthogonal projection or a 90 degree projection, and project it down onto U.



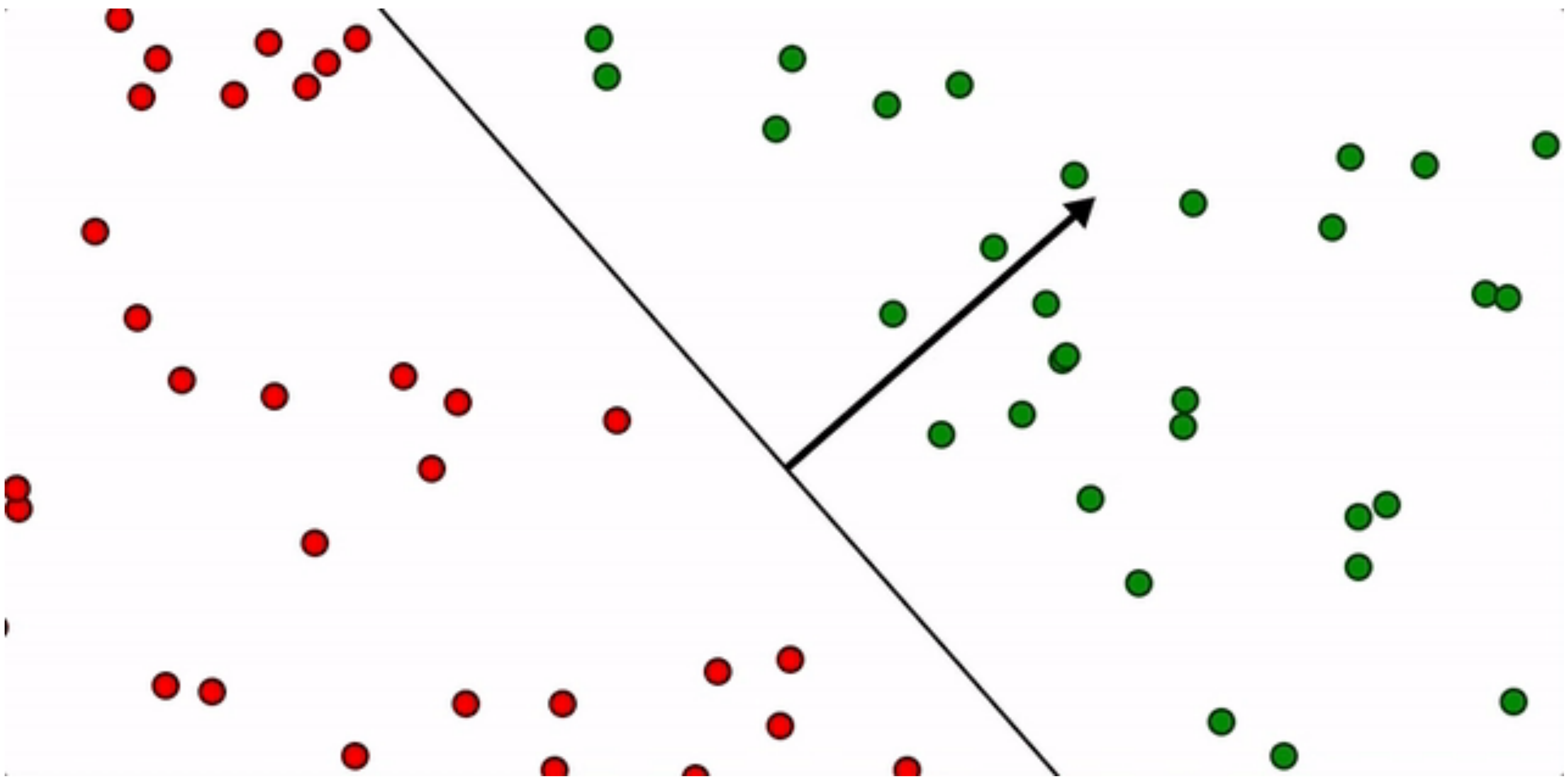
P is the length or is the magnitude of the projection of the vector V onto the vector U.

P can either be positive or negative.

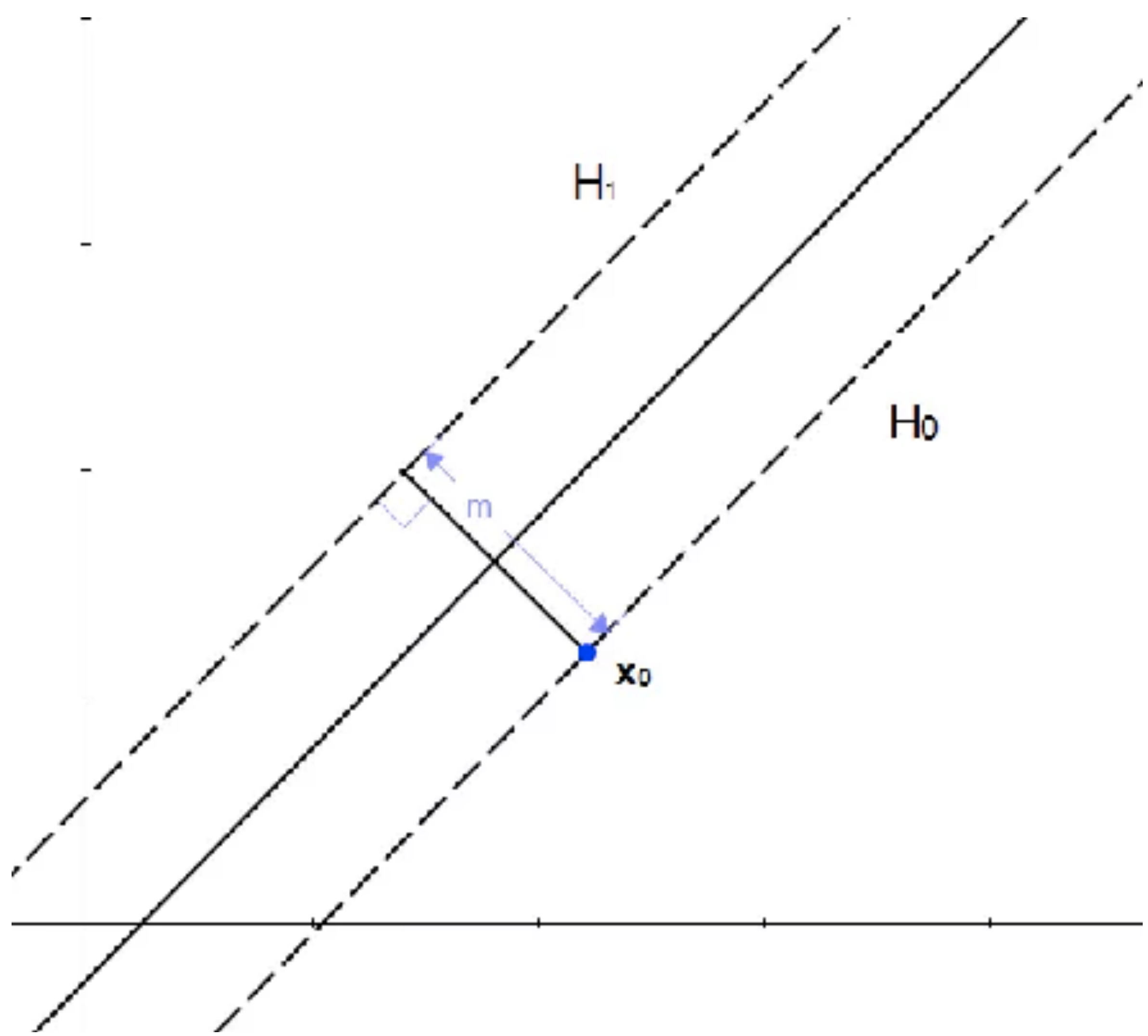


So if the angle between U and V is greater than ninety degrees. Then if we project V onto U, what we get is a projection P will be negative

Animation



Distance between 2 Hyper planes (m)



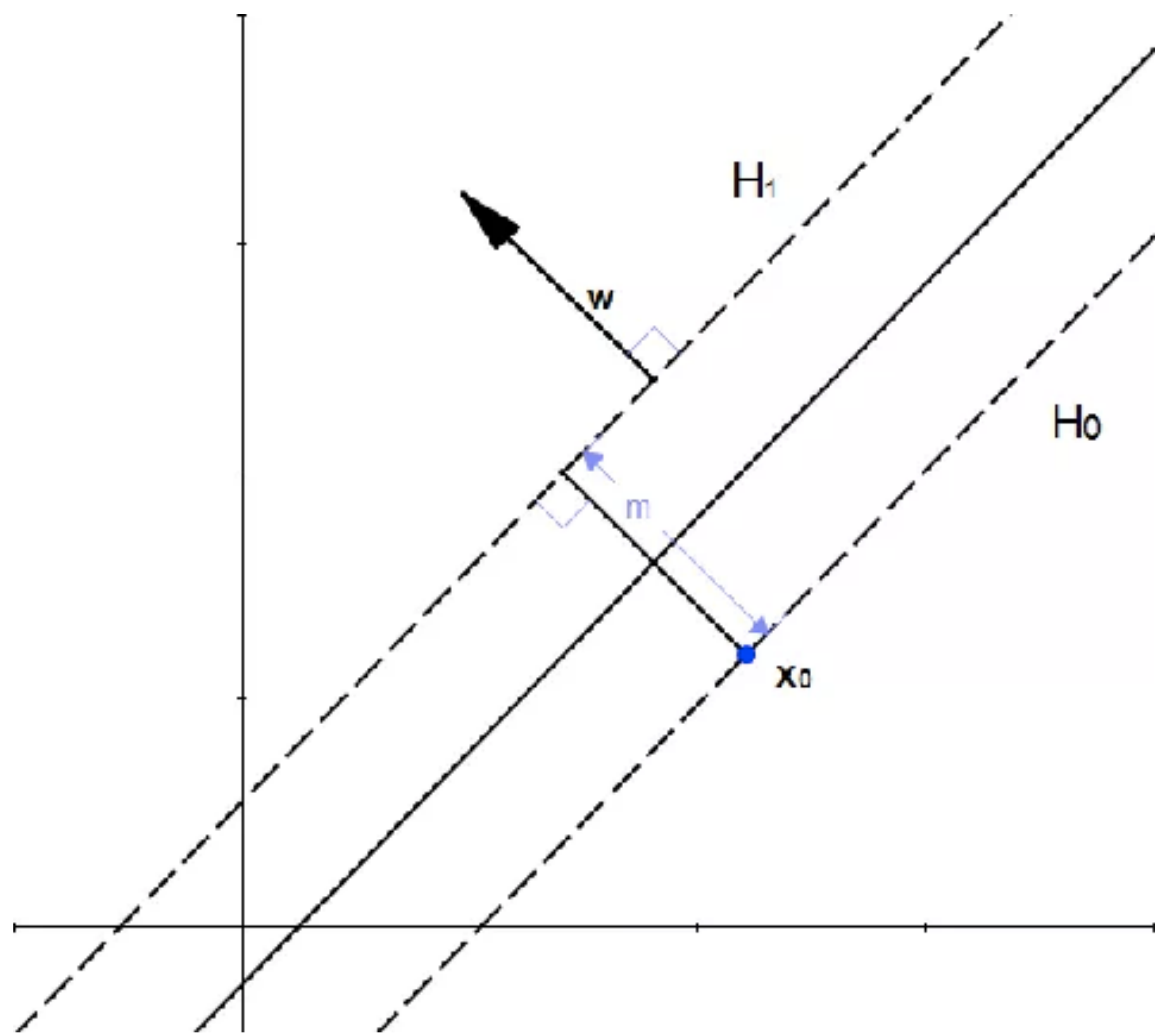
How will you find m, the perpendicular distance from x0 in hyperplane in H0 to the hyperplane H1.

You might be tempted to think that if we add m to x0 we will get another point, and this point will be on the other hyperplane !

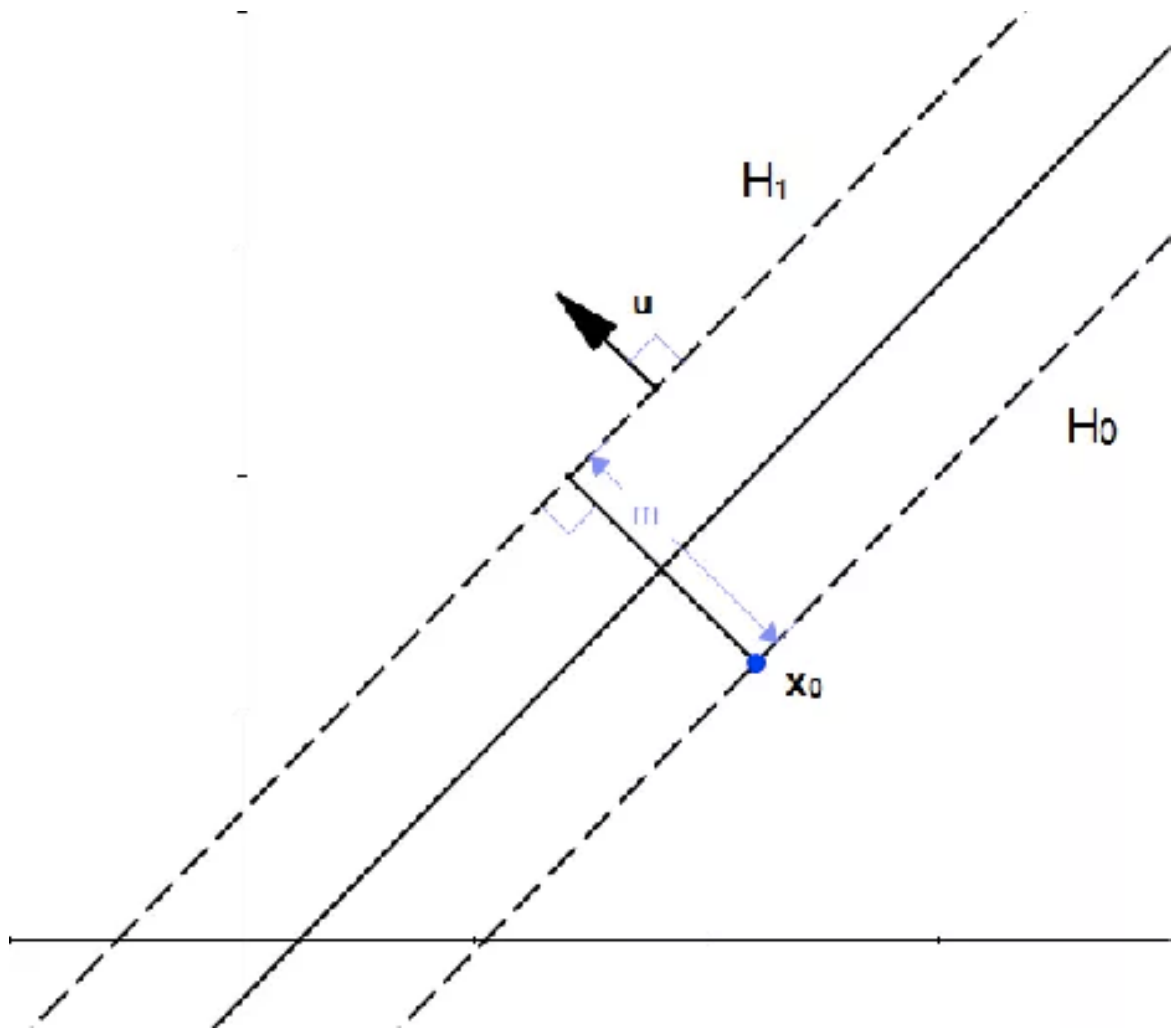
But **it does not work**, because **m is a *scalar*, and x0 is a *vector* and adding a scalar with a vector is not possible.**

However, we know that adding two vectors is possible, so if we transform m into a vector we will be able to do an addition.

We can't add a scalar to a vector, but we know if we multiply a scalar with a vector we will get another vector.

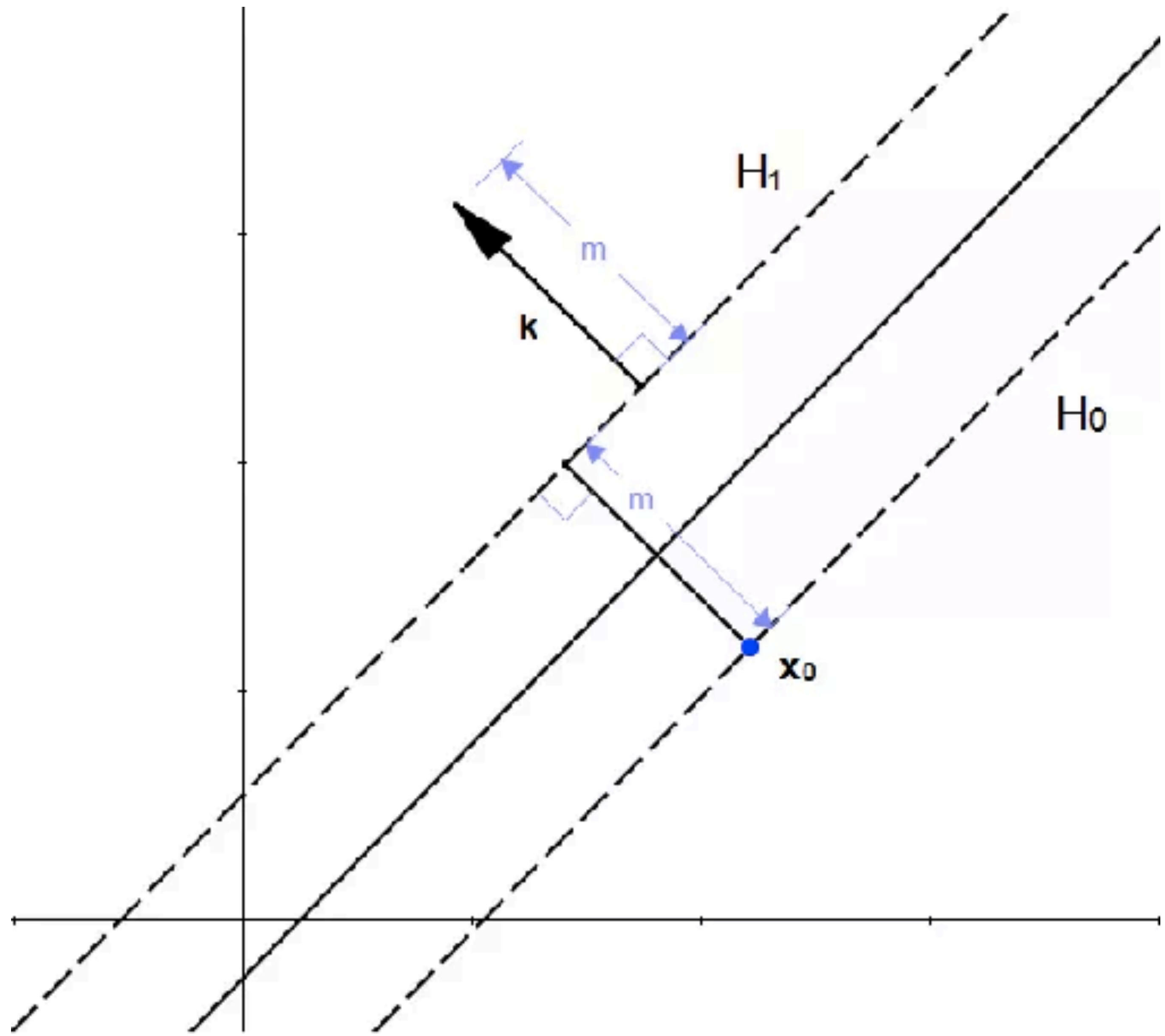


we already know a vector perpendicular to H1, that is W (because H1 = W. X + b = 1)



Let's define u = W / llWll   uint vector of W
As it is a unit vector llull=1 and it has the same direction as W. so it is also perpendicular to the hyperplane.

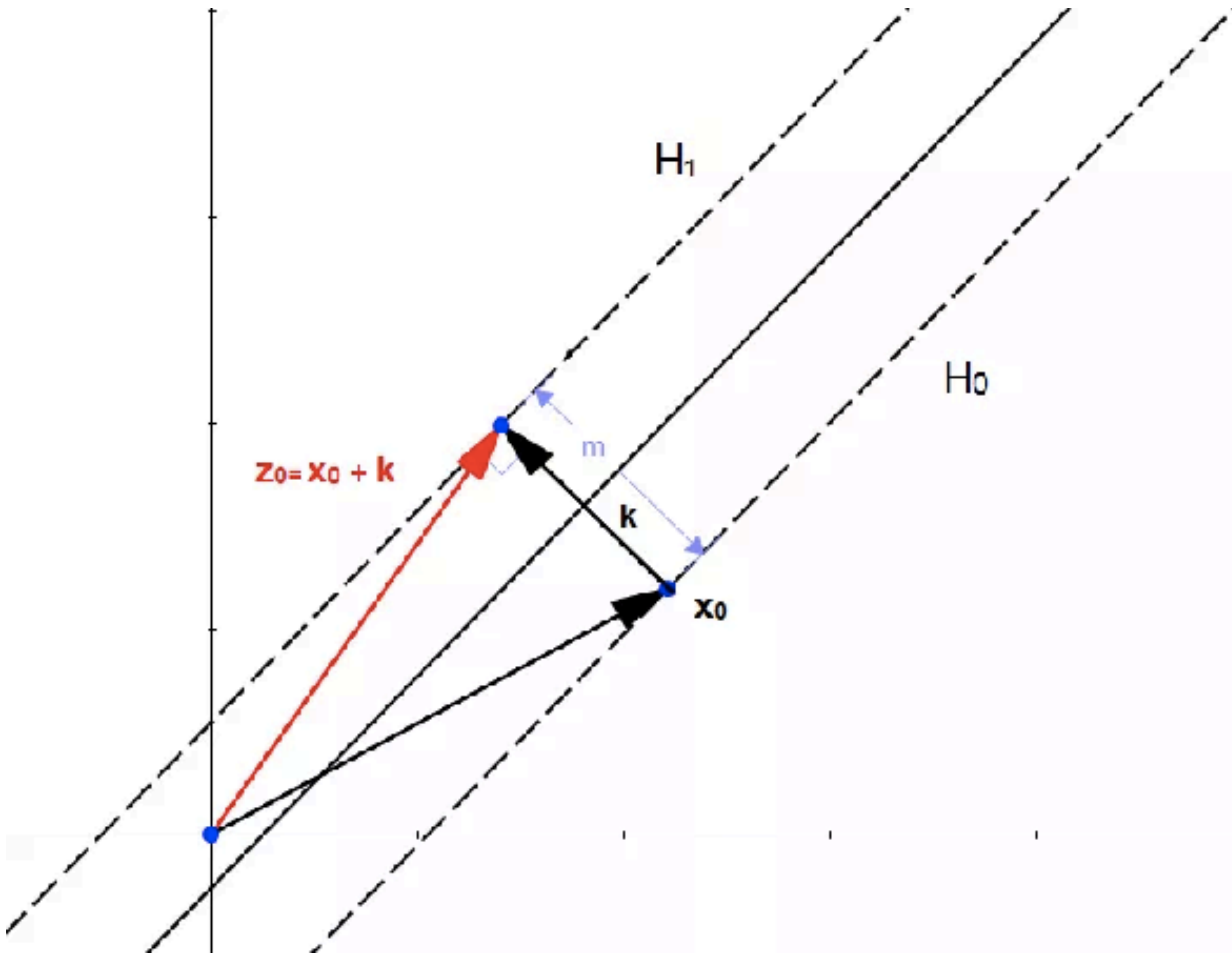
If we multiply U by m we get k = mu



note :
||k|| = m

$$\mathbf{k} = m\mathbf{u} = m\frac{\mathbf{w}}{\|\mathbf{w}\|}$$

We transformed our scalar m into a vector k which we can use to perform an addition with the vector X0. If we start from the point x0 and add k we find that the point Z0 = X0 + k is in the hyperplane H1



The fact that Z0 is in H1 means that W . Z0 + b = 1

We can replace Z0 by X0 + k because that is how we constructed it.

$$\mathbf{w} \cdot (\mathbf{x}_0 + \mathbf{k}) + b = 1$$

$$\mathbf{k} = m\mathbf{u} = m\frac{\mathbf{w}}{\|\mathbf{w}\|}$$

$$\mathbf{w} \cdot (\mathbf{x}_0 + m\frac{\mathbf{w}}{\|\mathbf{w}\|}) + b = 1$$

$$\mathbf{w} \cdot \mathbf{x}_0 + m\frac{\mathbf{w} \cdot \mathbf{w}}{\|\mathbf{w}\|} + b = 1$$

$$\mathbf{w}\cdot\mathbf{x}_0+m\frac{||\mathbf{w}||^2}{||\mathbf{w}||}+b=1$$

$$\mathbf{w}\cdot\mathbf{x}_0+m||\mathbf{w}||+b=1$$

$$\mathbf{w}\cdot\mathbf{x}_0+b=1-m||\mathbf{w}||$$

As X0 is in H0 then w.X0 + b = -1

$$-1=1-m||\mathbf{w}||$$

$$m||\mathbf{w}||=2$$

$$m=\frac{2}{||\mathbf{w}||}$$

We now have a formula to compute the margin

m = 2 / llWll

When $\|\mathbf{w}\| = 1$ then $m = 2$

When $\|\mathbf{w}\| = 2$ then $m = 1$

When $\|\mathbf{w}\| = 4$ then $m = \frac{1}{2}$

the bigger the norm is, the smaller the margin become.

we will choose the hyperplane with the smallest $\|\mathbf{w}\|$ because it is the one which will have the biggest margin.

Given a hyperplane H0 separating the dataset and satisfying

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

We can select two others hyperplanes H1 and H2 which also separate the data and have the following equations :

$$\mathbf{w} \cdot \mathbf{x} + b = \delta$$

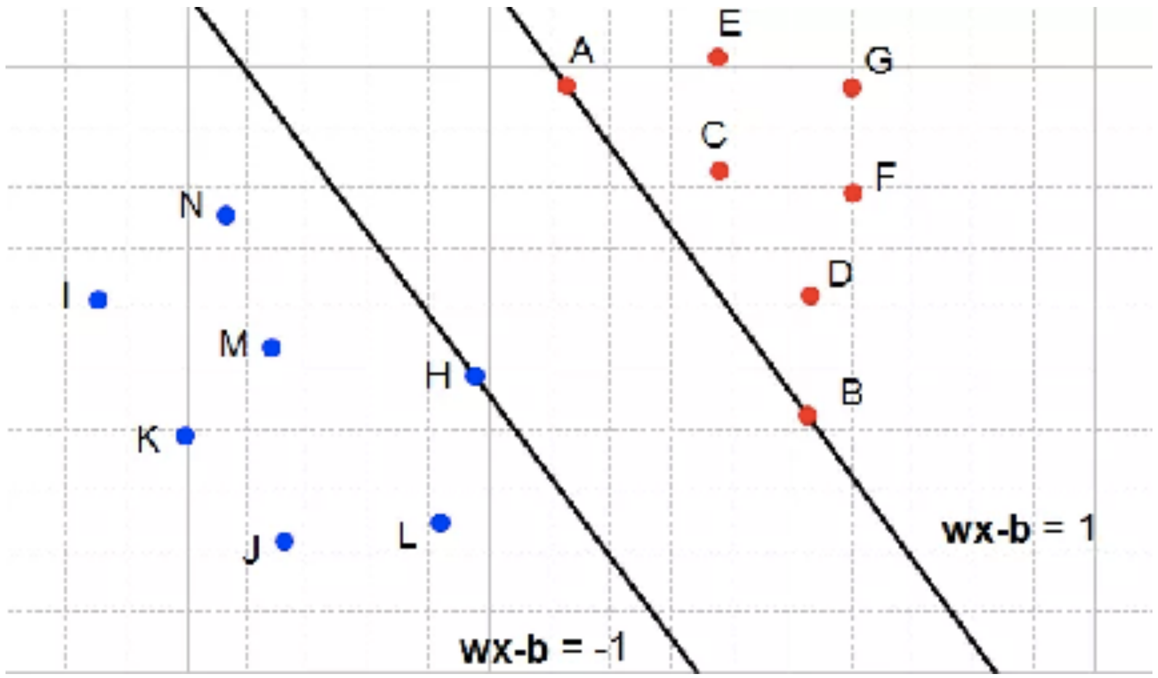
$$\mathbf{w} \cdot \mathbf{x} + b = -\delta$$

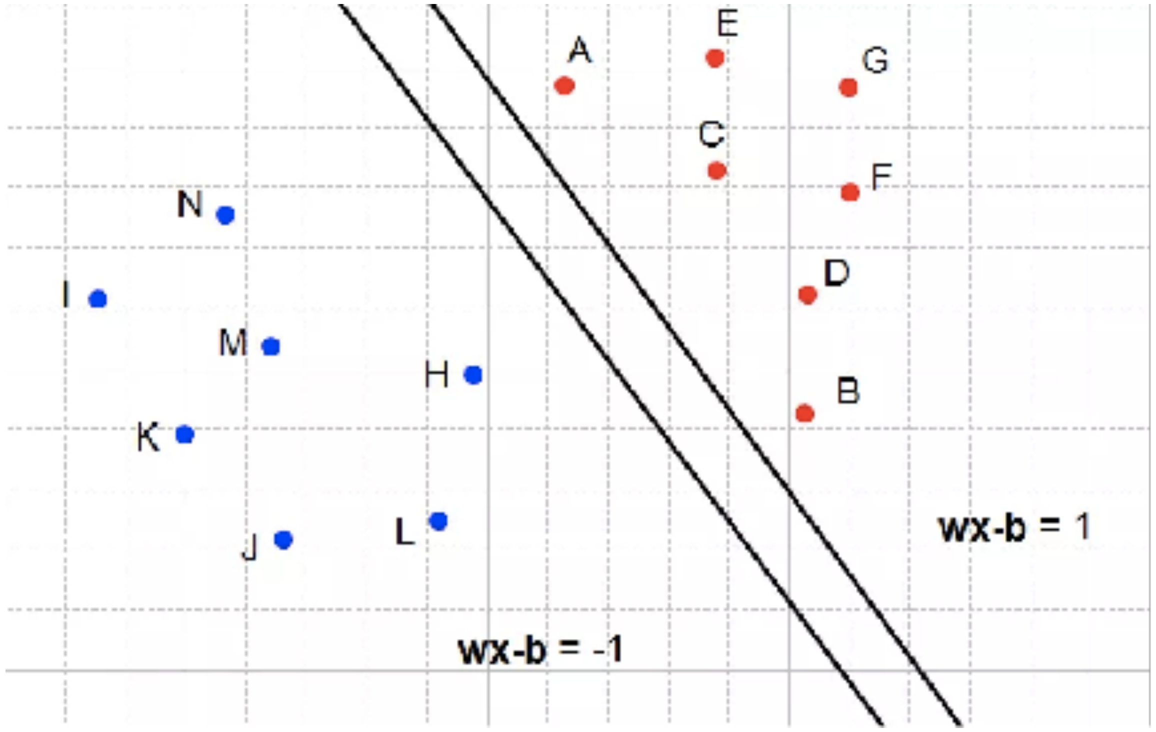
so that H0 is equidistant from H1 and H2

However, here the variable δ is not necessary. So we can set δ=1 to simplify the problem. (or any other value)

$$\mathbf{w} \cdot \mathbf{x} + b = 1$$

$$\mathbf{w} \cdot \mathbf{x} + b = -1$$





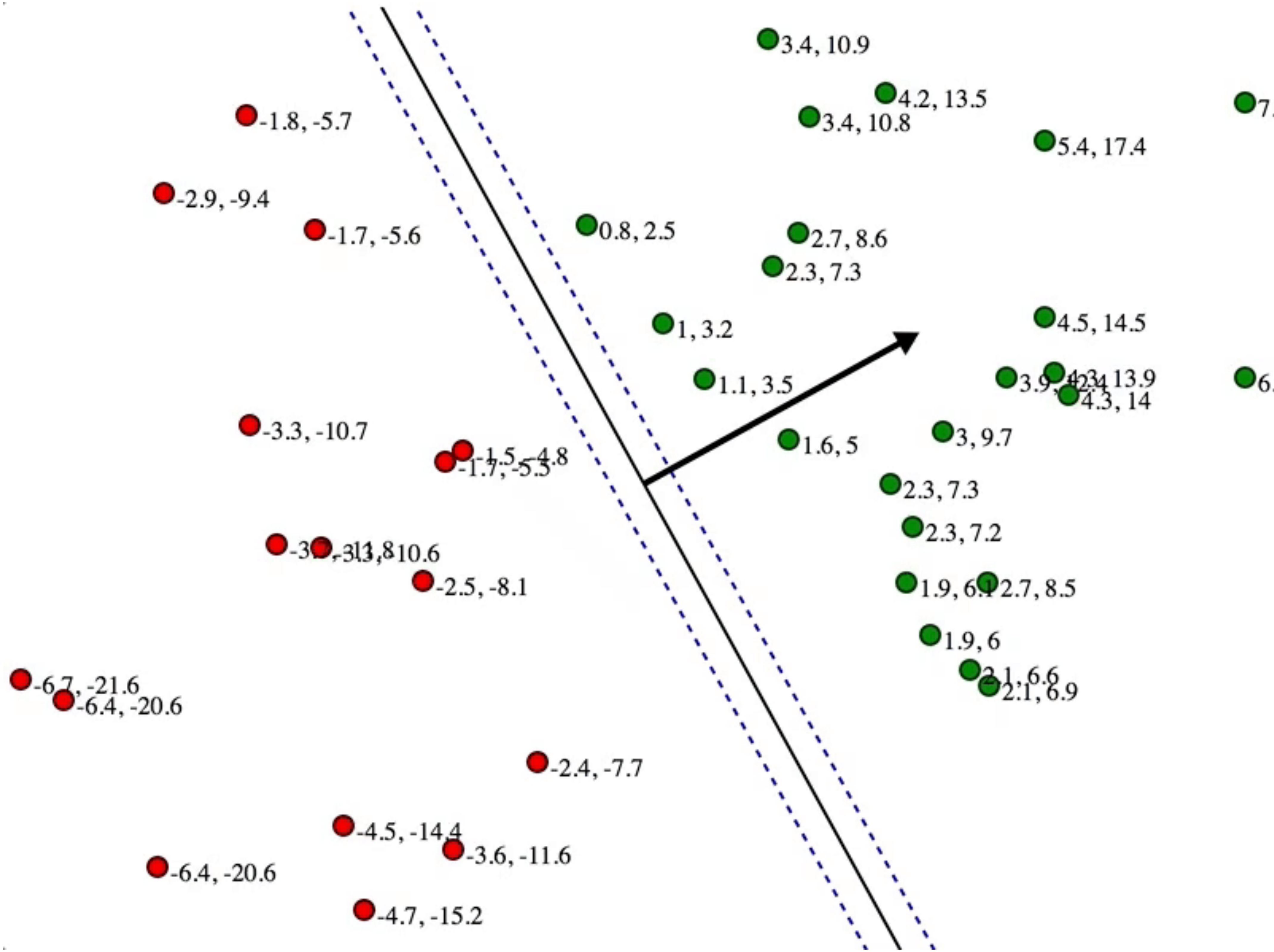
We won't select *any* hyperplane, we will only select those who meet the two following **constraints**:

$$\mathbf{w} \cdot \mathbf{x_i} + b \geq 1 \text{ for } \mathbf{x_i} \text{ having the class } 1$$

or

$$\mathbf{w} \cdot \mathbf{x_i} + b \leq -1 \text{ for } \mathbf{x_i} \text{ having the class } -1$$

Animation



3.4, 10.9
4.2, 13.5
3.4, 10.8
5.4, 17.4
-1.8, -5.7
-2.9, -9.4
-1.7, -5.6
0.8, 2.5
2.7, 8.6
2.3, 7.3
4.5, 14.5
1, 3.2
1.1, 3.5
4.3, 13.9
4.3, 14
-3.3, -10.7
3, 9.7
1.6, 5
-1.5, -4.8
-1.7, -5.5
2.3, 7.3
2.3, 7.2
-2.5, -8.1
2.7, 8.5
1.9, 6
2.1, 6.6
2.1, 6.9
-6.7, -21.6
-6.4, -20.6
-2.4, -7.7
-4.5, -14.4
-3.6, -11.6
-6.4, -20.6
-4.7, -15.2

```python
X = np.array([
    [-2, 4],
    [4, 1],
    [1, 6],
    [2, 4],
    [6, 2]
])

y = np.array([-1,-1,1,1,1])
```

b - 11.12
w1 - 1.56
w2 - 3.17

First sample (–2,4), supposed to be negative:
$-2*1.56+4*3.17-11.12=\text{sign}(-1{,}56)=-1$

Second sample (4,1), supposed to be negative:
$4*1.56+1*3.17-11.12=\text{sign}(-1{,}71)=-1$

Third sample (1,6), supposed to be positive:
$1*1.56+6*3.17-11.12=\text{sign}(9{,}46)=+1$

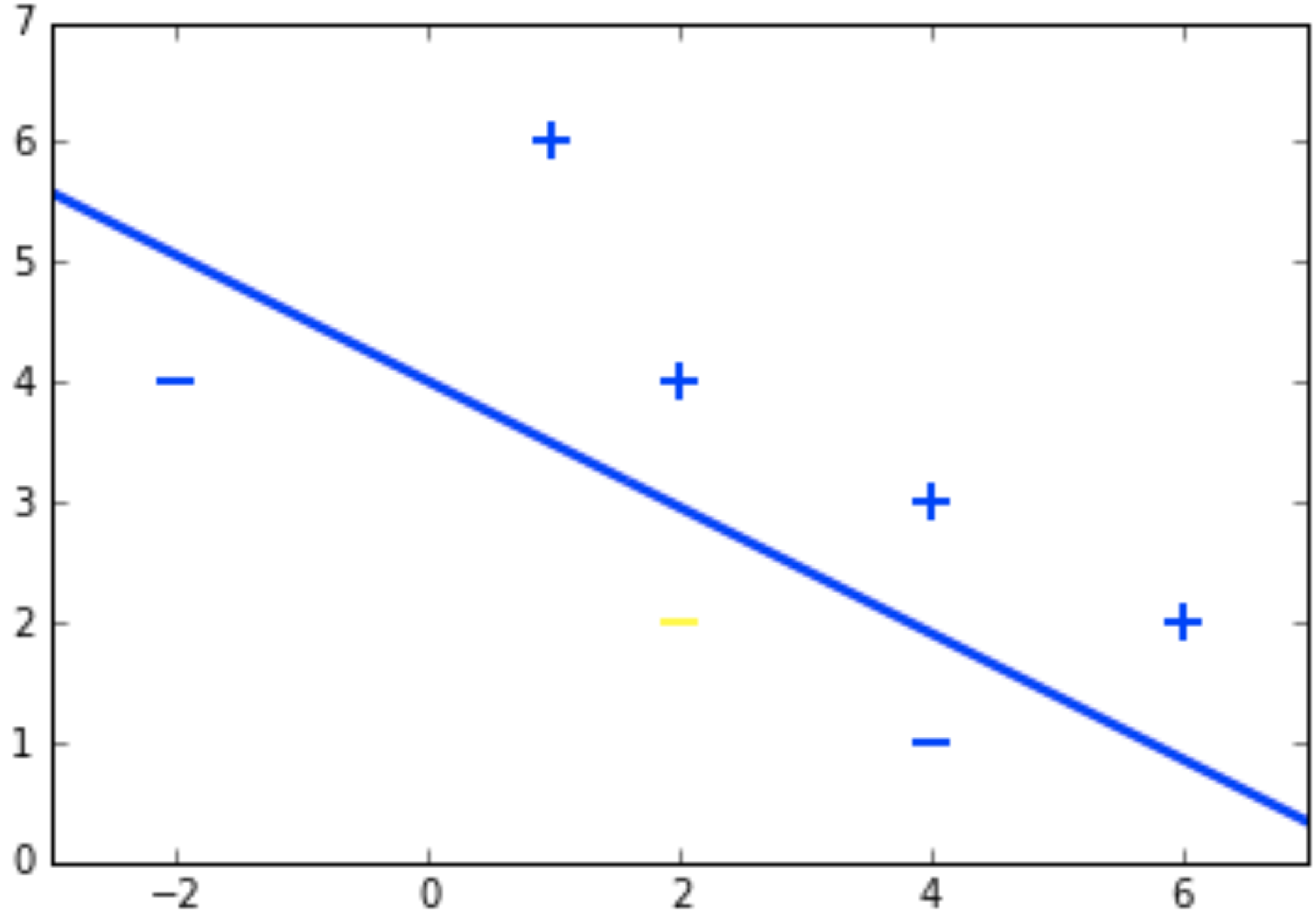
Fourth sample (2,4), supposed to be positive:
$2*1.56+4*3.17-11.12=\text{sign}(4{,}68)=+1$

Fifth sample (6,2), supposed to be positive:
$6*1.56+2*3.17-11.12=\text{sign}(4{,}58)=+1$

Lets define two test samples now, to check how well our perceptron generalizes to unseen data:

First test sample (2,2), supposed to be negative:
$2 * 1.56 + 2 * 3.17 - 11.12 = \text{sign}(-1{,}66) = -1$

Second test sample (4,3), supposed to be positive:
$4 * 1.56 + 3 * 3.17 - 11.12 = \text{sign}(4{,}63) = +1$



7
6
5
4
3
2
1
0
−2
0
2
4
6