Further talking about SVM hyperplane

1. Equation of a hyperplane

An equation of a line is:

$$y = ax + b$$

The equation of a hyperplane is an inner/dot/scalar product of two vector:

$$W^T X = 0$$

in 2D,

$$W = \begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix}, \quad X = \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}, \quad \text{and, } W^T X = y - ax - b$$

It is easier to work in more than two dimensions with this notation. The vector W will always be normal to the hyperplane.

Or, a hyperplane can also be represented as:

$$WX + b = 0$$

where,

$$W = \begin{pmatrix} -a \\ 1 \end{pmatrix}^T, \qquad X = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{and, } WX + b = y - ax + b$$

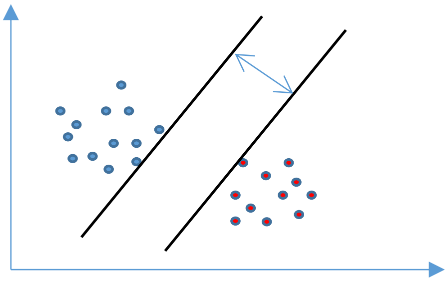
2. Finding the optimal hyperplane

In SVM, finding the optimal hyperplane is to find the hyperplane with the biggest margin. Then, how can we find the biggest margin? This problem can be defined as:

Given a dataset D

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p,\, y_i \in \{-1, 1\}\}_{i=1}^n$$

select two hyperplanes which separate the data with no points between them and maximize their distance (the margin).



Now, let us see how to find the maximum distance (the margin).

Assuming there is a hyperplane H separating the dataset D and satisfying:

$$WX + b = 0$$

Select two other a hyperplane H_1 and H_0:

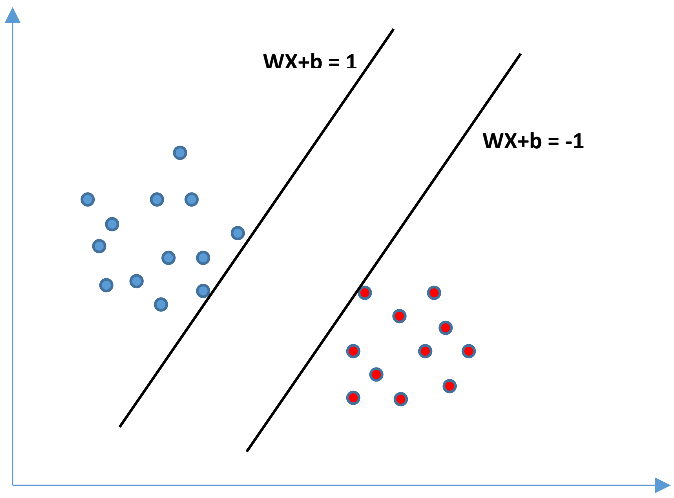
$$WX + b = \delta$$

$$WX + b = -\delta$$

In order to simplify the problem, we can set $\delta = 1$, so hyperplane H_1 and H_0 are:

$$WX + b = 1$$

$$WX + b = -1$$



To be sure there is no pints between them:

For each vector X_i, either:

$WX_i + b \geq 1$ for X_i having the class 1 , i.e., $y_i = 1$;

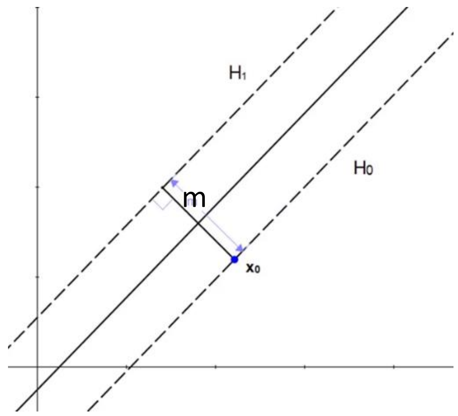
Or

$WX_i + b \leq -1$ for X_i having the class -1, i.e., $y_i = -1$

These two constraints can be combined into a unique:

$y_i(WX_i + b) \geq 1$ for all $1 \leq i \leq n$

Next, we calculate the distance between the two hyperplanes:



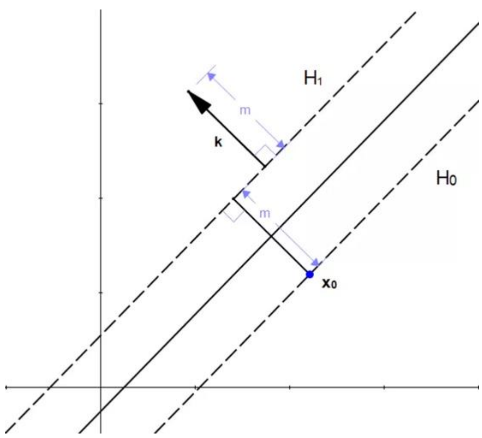
m: the perpendicular distance from $\mathbf{X_0}$ to the hyperplane $\mathbf{H_1}$

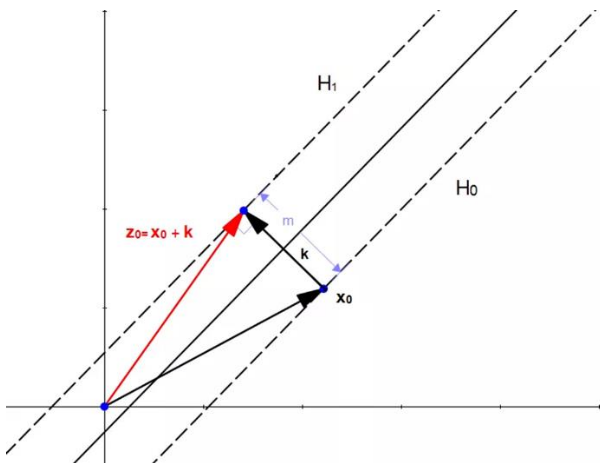
As illustrated in the figure above, the distance m is the perpendicular distance from a point on the hyperplane H_0, e.g., point X_0, to the hyperplane H_1, define a vector K, which is perpendicular to H_1 and has the length of m. Moving the start point of vector K to point X_0, the end point of vector K, i.e., point Z_0, will be on hyperplane H_1 , so:

$$WZ_0 + b = 1$$

as $Z_0 = X_0 + K$, then:

$$W \cdot (X_0 + K) + b = 1$$





Vector W and K have the same direction, so vector K can be written as $m\frac{W}{\|W\|}$, put it in the formula above, we have:

$$W \cdot (X_0 + m\frac{W}{\|W\|}) + b = 1$$

$$W \cdot X_0 + m\frac{W \cdot W}{\|W\|} + b = 1$$

$$W \cdot X_0 + m\frac{\|W\|^2}{\|W\|} + b = 1$$

$$W \cdot X_0 + m\|W\| + b = 1$$

$$W \cdot X_0 + b = 1 - m\|W\|$$

as X_0 is on H_0, then $W \cdot X_0 + b = -1$, so:

$$-1 = 1 - m\|W\|$$

$$m\|W\| = 2$$

$$m = \frac{2}{\|W\|}$$

Finally, we find out that maximizing the margin is the same as minimizing the norm of W. this give us the following optimization problem:

Minimize $\|W\|$ in (W, b),

subject to

$$y_i(WX_i + b) \geq 1 \quad \text{for all } 1 \leq i \leq n$$

Solve this problem, then we will have the equation of the optimal hyperplane!