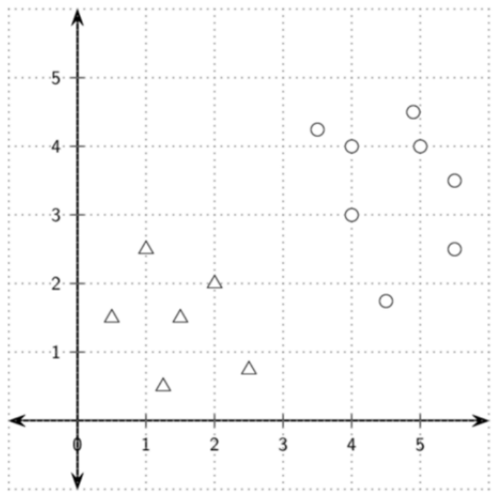
Support Vector Machine SVM

SVM Introduction.pdf

Maths behind SVM:



Linear SVM :

Suppose the following dataset is given :

X_i	X_i1	X_i2	Y_i
X_1	3.5	4.25	+1
X_2	4	3	+1
X_3	4	4	+1
X_4	4.5	1.75	+1
X_5	4.9	4.5	+1
X_6	5	4	+1
X_7	5.5	2.5	+1
X_8	5.5	3.5	+1
X_9	0.5	1.5	-1
X_{10}	1	2.5	-1
X_{11}	1.25	0.5	-1
X_{12}	1.5	1.5	-1
X_{13}	2	2	-1
X_{14}	2.5	0.75	-1

Now, we need to find the supporting vectors and coefficients of variables. According to our course, these things will be already provided in the question:

X_i	X_i1	X_i2	Y_i	α_i
X_1	3.5	4.25	+1	0.0437
X_2	4	3	+1	0.2162
X_4	4.5	1.75	+1	0.1427
X_{13}	2	2	-1	0.3589
X_{14}	2.5	0.75	-1	0.0437

Support Vectors have $\alpha_i > 0$

Hence, there are total fine supporting vectors.

The weight vector and bias are:

$$w = \sum \alpha_i y_i x_i$$

$$= 0.0437\begin{pmatrix} 3.5 \\ 4.25 \end{pmatrix} + 0.2162\begin{pmatrix} 4 \\ 3 \end{pmatrix} + 0.1427\begin{pmatrix} 4.5 \\ 1.75 \end{pmatrix}$$

$$- 0.3589 \begin{pmatrix} 2 \\ 2 \end{pmatrix} - 0.0437 \begin{pmatrix} 2.5 \\ 0.75 \end{pmatrix}$$

$$= \begin{pmatrix} 0.15295 \\ 0.18573 \end{pmatrix} + \begin{pmatrix} 0.8648 \\ 0.6486 \end{pmatrix} + \begin{pmatrix} 0.64215 \\ 0.24973 \end{pmatrix}$$

$$- \begin{pmatrix} 0.7178 \\ 0.7178 \end{pmatrix} - \begin{pmatrix} 0.10925 \\ 0.03278 \end{pmatrix}$$

$$w = \begin{pmatrix} 0.83285 \\ 0.33348 \end{pmatrix}$$

$$b = \sum \left(y_i - (w^T x_i) \right) / \text{Number of Support Vectors}$$

$$= \sum \left(y_i - (w_1 x_1 + w_2 x_2) \right) / \text{Number of Support Vectors}$$

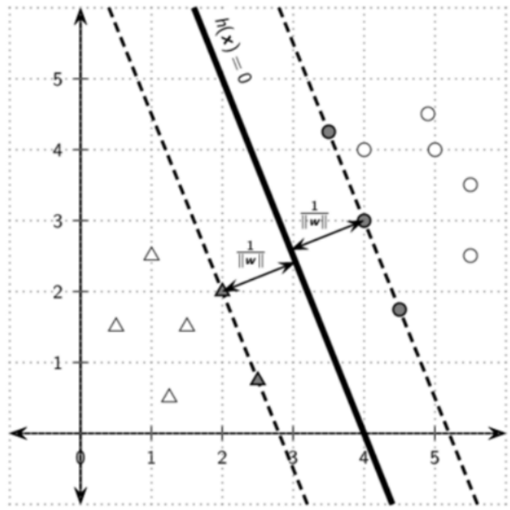
$$= \Big[\left(+1 - (0.833 \times 3.5 + 0.333 \times 4.25) \right)$$
$$+ \left(+1 - (0.833 \times 4 + 0.333 \times 3) \right)$$
$$+ \left(+1 - (0.833 \times 4.5 + 0.333 \times 1.75) \right)$$
$$+ \left(-1 - (0.833 \times 2 + 0.333 \times 2) \right)$$
$$+ \left(-1 - (0.833 \times 2.5 + 0.333 \times 0.75) \right) \Big] / 5$$

$$= \frac{-16.6574}{5}$$

$$= -3.3315$$

The optimal hyperplane is :

$$h(x) = \begin{pmatrix} 0.833 \\ 0.333 \end{pmatrix}^T x - 3.332 = 0$$


$h(x) = 0$
$\frac{1}{\|w\|}$
$\frac{1}{\|w\|}$
0
1
2
3
4
5

Distance of a point to the Hyperplane:

Suppose we have the following optimal hyperplane equation:

$$h(x) = \begin{pmatrix} 5 \\ 2 \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 20 = 0$$

We can find the distance of point using this formula:

$$\delta = yr = \frac{y\,h(x)}{\|W\|}$$

Where y is the class label and when $h(x_i) \leq 0$, the class is -1 and when $h(x_i) > 0$ the class is $+1$.

For the origin $x = 0$, the directed distance is:

$$r = \frac{h(0)}{\|W\|} = \frac{W^T 0 + b}{\|W\|} = \frac{b}{\|W\|}$$

For the above equation:

$$h(x) = (5 \quad 2)\begin{pmatrix} 0 \\ 0 \end{pmatrix} - 20 = -20$$

As $h(x) < 0$

So $y = -1$

Now

$$\delta = yr = -1 \times \frac{b}{\|W\|} = \frac{-1 \times -20}{\sqrt{5^2 + 2^2}} = \frac{20}{\sqrt{29}} = 3.71$$

Margin and Support Vectors:

The margin is the minimum distance of a point from the hyperplane. All points that achieve the minimum distance are called support vectors for the hyperplane.

$$\delta^* = \frac{y^*(w^T x^* + b)}{\|w\|}$$

Canonical Hyperplane:

To obtain the unique or canonical hyperplane, we choose the scalar,

$$s = \frac{1}{y^*(w^T x^* + b)}$$

So that the absolute distance of a support vector from the hyperplane is 1 i.e. margin is

$$\delta^* = \frac{1}{\|w\|}$$

For the canonical hyperplane, for each support vector, we have $y_i^* h(x_i^*) = 1$, and for any other point it will be greater than 1.

For example,

$$h(x) = \begin{pmatrix} 5 \\ 2 \end{pmatrix}^T x - 20 = 0$$

Given $x^* = (2,2)^T$, $y^* = -1$

$$s = \frac{1}{y^* h(x^*)} = \frac{1}{-1\left(\begin{pmatrix} 5 & 2 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \end{pmatrix} - 20\right)} = \frac{1}{6}$$

$$w = \frac{1}{6}\begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 2/6 \end{pmatrix}$$

$$= \begin{pmatrix} 0.833 \\ 0.333 \end{pmatrix}$$

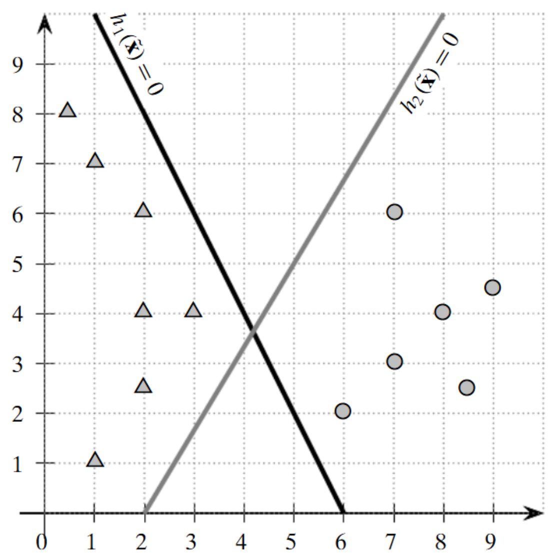
$$b = \frac{-20}{6}$$

$$= -3.33$$

$$h(x) = \begin{pmatrix} 0.833 \\ 0.333 \end{pmatrix}^T x - 3.33$$

As the given point is a support vector,

$$\delta^* = \frac{1}{\|w\|} = \frac{1}{\sqrt{0.833^2 + 0.333^2}} = 1.1147$$



a) Find the equations for the two hyperplanes h_1 and h_2:

For h_1, we are selecting two points on the hyperplane:

(6,0) and (1,10),

The slope is given as,

$$m = \frac{q_2 - p_2}{q_1 - p_1} = \frac{10-0}{1-6} = -\frac{10}{5} = -2$$

Using (6,0) as the point, we get the equation of line as:

$$\frac{x_2 - 0}{x_1 - 6} = -2$$

$$x_2 = -2x_1 + 12$$

$$\Rightarrow 2x_1 + x_2 - 12 = 0$$

Hence,

$$h(x) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}^T X - 12 = 0$$

For h_2, we are selecting two points on the hyperplane:

(2,0) and (8,10),

The slope is given as,

$$m = \frac{q_2 - p_2}{q_1 - p_1} = \frac{10-0}{8-2} = \frac{10}{6} = \frac{5}{3}$$

Using (2,0) as the point, we get the equation of line as:

$$\frac{x_2 - 0}{x_1 - 2} = \frac{5}{3} \Rightarrow 3x_2 = 5x_1 - 10 \Rightarrow 5x_1 - 3x_2 - 10 = 0$$

Hence,

$$h(x) = \begin{pmatrix} 5 \\ -3 \end{pmatrix}^T X - 10 = 0$$

b) Show all the support vectors for h_1 and h_2:

For finding support vectors for h_1:

As, $\gamma = yr = \frac{y\,h(x)}{\|W\|}$ where $\|W\| = \sqrt{2^2+1^2} = \sqrt{5}$

For $(0.5, 8)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}0.5\\8\end{pmatrix} - 12 = -3 \Rightarrow \frac{(-1)(-3)}{\sqrt{5}} = 1.3416$$

For $(1, 7)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}1\\7\end{pmatrix} - 12 = -3 \Rightarrow \frac{(-1)(-3)}{\sqrt{5}} = 1.3416$$

For $(1, 1)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}1\\1\end{pmatrix} - 12 = -9 \Rightarrow \frac{(-1)(-9)}{\sqrt{5}} = 4.0249$$

For $(2, 6)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}2\\6\end{pmatrix} - 12 = -2 \Rightarrow \frac{(-1)(-2)}{\sqrt{5}} = 0.8944$$

For $(2, 4)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}2\\4\end{pmatrix} - 12 = -4 \Rightarrow \frac{(-1)(-4)}{\sqrt{5}} = 1.7889$$

For $(2, 2.5)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}2\\2.5\end{pmatrix} - 12 = -5.5 \Rightarrow \frac{(-1)(-5.5)}{\sqrt{5}} = 2.4597$$

For $(3, 4)$:

$$h(x) = \begin{pmatrix}2\\1\end{pmatrix}^T \begin{pmatrix}3\\4\end{pmatrix} - 12 = -2 \Rightarrow \frac{(-1)(-2)}{\sqrt{5}} = 0.8944$$

Now, For the points on the otherside of hyperplane.

For $(6,2)$:

$h(x) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}^T \begin{pmatrix} 6 \\ 2 \end{pmatrix} - 12 = 2 \Rightarrow \dfrac{(1)(2)}{\sqrt{5}} = 0.8944$

In this way we calculate all the distances and found that 0.8944 is the minimum distance.

Hence support vectors for h_1 are:

$(2,6)$, $(3,4)$ and $(6,2)$.

In the similar way we will calculate the distances for h_2.

c) Which of the two hyperplanes is better is better at seperating the two classes based on the margin computation:

For h1:

margin = 2 × Distance of Supporting vectors to hyperplane

= 2 × 0.8944 = 1.7888

For h2:

margin = 2 × Distance of Supporting vectors to hyperplane

= 2 × 1.2006 = 2.4014

Hence according to above calculations h_2 is better than h_1.

Note: For canonical hyperplane we will use this formula to calculate the distance between margins: m = 2 / ||w||.