ॐ AMMA

# Support Vector Machine

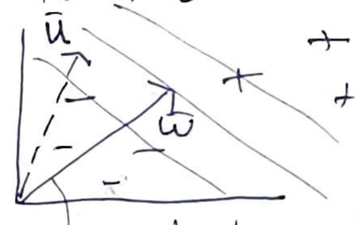
This is a sophisticated idea.
How do you divide positive examples from the negative examples?

You want to draw a line between them?
Which one is the one?

The decision boundary is to put it in a straight line in such a way that the separation between positive and negative example is as wide a possible.


Think about how do you make a decision boundary.
What should be the decision rule?

Imagine a vector perpendicular $\bar{w}$ to the median.
We don't know any thing about its length here




Perpendicular vector.

Given an unknown vector $\bar{u}$,
what we are really interested is,
whether $\bar{u}$ is on the left or right side of the decision boundary.
Hence, we want to project $\bar{u}$ on $\bar{w}$. Then we will have a number that is proportional to this in the direction of $\bar{w}$.
i.e. dot product of $\bar{u}$ and $\bar{w}$ is a number which fall on this side or that side of the decision boundary.

So, what we can do is take

$$\bar{w} \cdot \bar{u} \geq c$$

Remember, dot product is taking the projection of $\bar{u}$ on $\bar{w}$.

Without loss of generality, $\rightarrow b = -c$

(1) $\boxed{\bar{w} \cdot \bar{u} + b \geq 0 \quad \text{then +ve sample else -ve sample.}}$

Decision Rule.

At this point, we don't know what $\bar{w}$ and $b$ we have to use. All we know is $\bar{w}$ is perpendicular to the median line.

Note: There are lots of $\bar{w}$ perpendicular to the median since $\bar{w}$ could be of any length.

The next step is to lay on some additional constraints to fix a particular $w$ and a particular $b$.

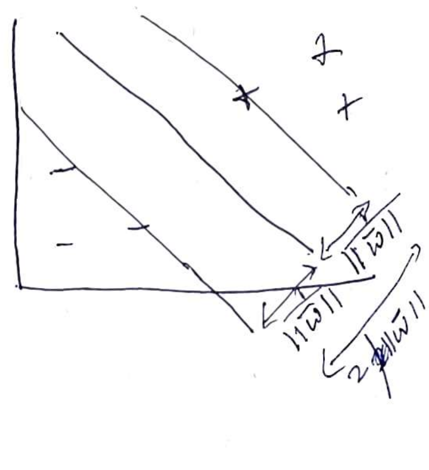
If we look at $\bar{w} \cdot \bar{u} + b \geq 0$, to make a decision, we want

$$\bar{w} \cdot \bar{x}_+ + b \geq 1 \quad \text{for + sample } x_+$$

Likewise

$$\bar{w} \cdot \bar{x}_- + b \leq 1 \quad \text{for - sample. } x_-$$

∴ There is a separation on the distance here between -1 and +1 for all samples



For mathematical convenience, we are going to combine the above two equations.

We know, $y_i = +1$ if $i^{th}$ sample is +ve

$= -1$ if $i^{th}$ sample is -ve

Multiplying $y_i$ to both the above equations

SAME

$y_i(\bar{w}_i \cdot \bar{x}_i + b) \geq 1$ positive sample $y_i = +1$

$y_i(\bar{w}_i \cdot \bar{x}_i + b) \geq 1$ negative sample $y_i = -1$

The inequality reverses

Now we can say.

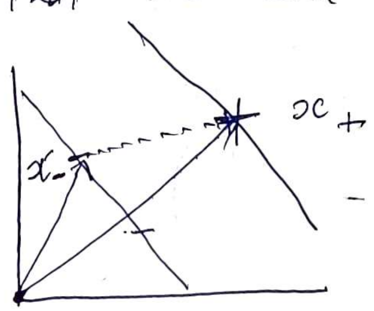
$y_i(\bar{w}_i \cdot \bar{x}_i + b) - 1 \geq 0$ for all samples

For those points on the margins (i.e. support vectors)

(2) $\boxed{y_i(\bar{w}_i \cdot \bar{x}_i + b) - 1 = 0}$

We want to arrange the line such that the positive and negative samples are separated as much as possible.

To this end, we try to figure out how to express the distance between the margins so, that we can maximize this distance.




What is the width ~~the~~ between the margins?

$\bar{w} \cdot \bar{x}_+ - \bar{w} \cdot \bar{x}_-$

We take the difference between the two vectors $x_+$ and $x_-$.

$$\text{width} = \bar{w} \cdot (\bar{x}_+ - \bar{x}_-)$$
(or distance)

By dividing by the norm $\|w\|$ we can make this unit vector.

$$\text{width} = \frac{\bar{w}}{\|\bar{w}\|} (\bar{x}_+ - \bar{x}_-) \quad \text{——— ③}$$

Look at equation ②. It says

$$y_i(\bar{w}_i \cdot \bar{x}_i + b) - 1 = 0$$ for samples on the margins (support vectors)

From ~~hence~~ we can derive

$$1(\bar{w} \cdot \bar{x}_+ + b) - 1 = 0$$ for positive sample $x_+$

$$\Rightarrow x_+ = \frac{1 - b}{\bar{w}} \quad \text{——— 4a}$$

and

$$-1(\bar{w} \cdot \bar{x}_- + b) - 1 = 0$$ for negative sample $x_-$

$$\Rightarrow x_- = -\frac{1+b}{\bar{w}} \quad \text{——— 4b}$$

Substituting 4a and 4b in ③

$$\text{width} = \frac{\bar{w}}{\|\bar{w}\|} \left( \frac{1-b}{\bar{w}} - -\frac{1+b}{\bar{w}} \right)$$

$$= \frac{\cancel{\bar{w}}}{\|\bar{w}\|} \left( \frac{1 - \cancel{b} + 1 + \cancel{b}}{\cancel{\bar{w}}} \right) = \frac{2}{\|\bar{w}\|}$$

We want to maximize width $= \frac{2}{||w||}$

$\Rightarrow$ Maximize $\frac{1}{||w||}$

$\Rightarrow$ Minimize $||w||$

$\Rightarrow$ Minimize $\frac{1}{2} ||w||^2$ (for mathematical convenience)

Hence our goal is to

(5) $\boxed{\text{Minimize } \frac{1}{2} ||w||^2 \quad \text{subject to the constraint } y_i (\bar{w}_i \cdot \bar{x}_i + b) - 1 = 0}$

This is a constrained optimization problem.

We want to convert the constrained optimization problem into a form such that the derivative test of an unconstrained problem can still be applied.

The relationship between the gradient of the function and gradients of the constraints rather naturally leads to a reformulation of the original problem, known as the Lagrangian function or Lagrangian.

In the general case, Lagrangian is defined as

$$\mathcal{L}(x, \lambda) = f(x) + \lambda \cdot g(x)$$

↳ Lagrange multiplier

(In short Lagrangian combines the objective and the constraint(s) into a single equation)

Applying Lagrangian to ⑤

$$\boxed{L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\bar{w}_i \cdot \bar{x}_i + b) - 1]}$$

n = # samples

$\alpha_i \rightarrow$ Lagrange multipliers.

Now, we have got to find the derivatives and set them to zero. $\frac{\partial L}{\partial w} = 0$ + $\frac{\partial L}{\partial b} = 0$.

$$\frac{\partial L}{\partial w} = \bar{w} - \sum_{i=1}^{n} \alpha_i y_i \bar{x}_i = 0$$

$$\Rightarrow \boxed{\bar{w} = \sum_{i=1}^{n} \alpha_i y_i \bar{x}_i} \quad ⑥$$

This tells us that $\bar{w}$ is a linear combination of the samples.

Similarly,

$$\frac{\partial L}{\partial b} = 0 - \sum_{n=1}^{n} \alpha_i y_i = 0$$

$$\Rightarrow \sum \alpha_i y_i = 0$$

To find $b$; substitute $\bar{w}$ in equation ②

$$y_i(\bar{w}_i \bar{x}_i + b) - 1 = 0$$

$$y_i\left(\left[\sum_{j=1}^{n} \alpha_j y_j x_j\right] \cdot x_i + b\right) - 1 = 0$$

Note: We used a different index $j$ while sub. for $\bar{w}$ (from ⑥) since $i$ was present in the equation ②

Multiplying both sides by $y_i$:

$$y_i \cdot y_i \left( \left[ \sum \alpha_j y_j x_j \right] \cdot x_i + b \right) - y_i = 0$$

$$\Rightarrow 1 \left( \sum_{j=1}^{n} \alpha_j y_j x_j \cdot x_i + b \right) = y_i$$

$$\Rightarrow b = y_i - \sum \alpha_j y_j x_j \cdot x_i$$

Generally, the average is taken

i.e.

$$\boxed{b = \frac{\sum_j \left( y_i - \sum \alpha_j y_j x_j \cdot x_i \right)}{N_j}} \quad (7)$$

Now, that $\overline{w}$ and $b$ are known, the optimal hyperplane can be given by

$$\boxed{\overline{w} \cdot \overline{x} + b = 0}$$