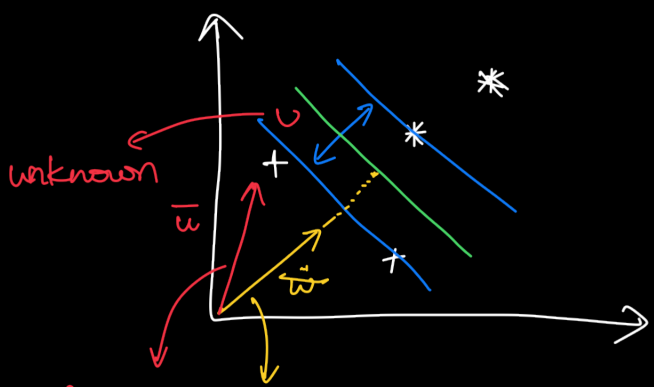
Support vector machine

Decision boundaries



vector pointing to unknown

vector which perpendicular to the median

Draw straight line

But How?

Draw wide as possible.

Now we don't known whether it is "*" or "+"

we project that vector down on the one that's $\perp^r$ to the street.

$\bar{w} \cdot \bar{u} \geq c$ → constant $c = -b$

DECISION RULE

$$\boxed{\bar{w} \cdot \bar{u} + b \geq 0 \text{ then it is "*" else it is "+"}}$$

But we don't which b to use and which $\bar{w}$ to use either. (1)

because there can be multi $\vec{w}$ which can be drawn from with many lengths.

adding addition constraints

$$\bar{w}\cdot\bar{x}_+ + b \geq 1 \quad \text{—— ②}$$

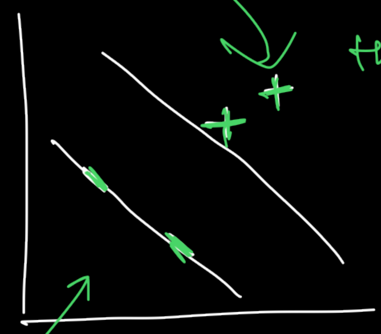
↓

positive samples

$$\bar{w}\cdot\bar{x}_- + b \leq -1 \quad \text{—— ③}$$

↓

negative samples.



we introduce a term y_i for mathematical conversion.

y_i such that $y_i = +1$ for + samples.

$y_i = -1$ for − samples.

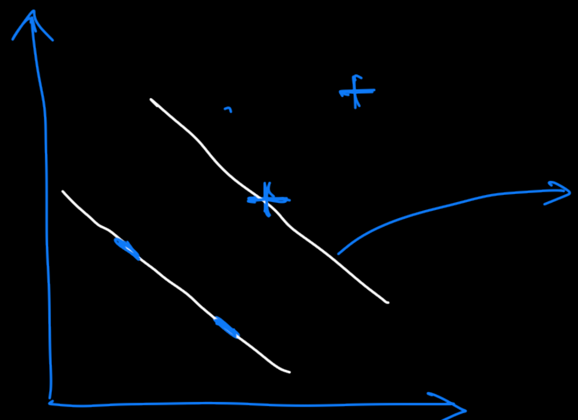
② ⇒ $$y_i(\bar{x}_i\bar{w} + b) \geq 1$$

③ ⇒ $$y_i(\bar{x}_i\bar{w} + b) \geq 1$$

both became same equations.

$$\Downarrow$$

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

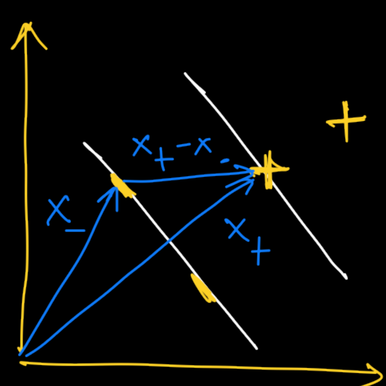


this line will be gutter

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

(4)

Now to express the distance the gutter.



$\bar{x}_+$ = vector of x_+ in gutter

$\bar{x}_-$ = vector of x_- in gutter.

$$\boxed{\text{width} = (\bar{x}_+ - \bar{x}_-) \cdot \frac{\bar{w}}{\|\bar{w}\|}}$$

Scalar

(5)

(5) $\Rightarrow$ $$\text{width} = \frac{\bar{w}\cdot\bar{x}_+ - \bar{w}\cdot\bar{x}_-}{\|\bar{w}\|}$$

From (2) $\Rightarrow$ $$+1(\bar{w}\cdot\bar{x}_+ - b) - 1 = 0$$

$$\bar{w}\cdot\bar{x}_+ - b - 1 = 0$$

$$\boxed{\bar{w}\cdot\bar{x}_+ = b+1}$$

From (3) $\Rightarrow$ $$-1(\bar{w}\,\bar{x}_- - b) - 1 = 0$$

$$-\bar{w}\,\bar{x}_- + b - 1 = 0$$

$$\boxed{\bar{w}\cdot\bar{x}_- = b - 1}$$

Sub in (5)

$$\text{width} = \frac{\cancel{b}+1 - \cancel{b}+1}{\|w\|}$$

$$\boxed{\text{width} = \frac{2}{\|w\|}} \longrightarrow (6)$$

we need to maximize the line

$$\max \frac{2}{||w||}$$

So we need to max $\frac{1}{||w||}$ or

min $||w||$

or

$$\boxed{\frac{1}{2}||w||^2}$$

→ But why $\frac{1}{2}$ and squared

for mathematical convension

7

7 is an expression and 2 & 3 are constraints and we need to find maximum.

If we need to find minima or maxima of a function subject to constraint then we have to use lagrange multiplier.

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

expression

L (*)

summation of all constants

multipliers

constraints

$$\frac{\partial L}{\partial \bar{w}} = \bar{w} - \sum \alpha_i y_i x_i = 0$$

$$\Rightarrow \boxed{\bar{w} = \sum \alpha_i y_i \bar{x}_i} \longrightarrow (8)$$

$\bar{w}$ is the linear sum of all samples.

$$\frac{\partial L}{\partial b} = 0 - \sum \alpha y_i = 0$$

$$\Rightarrow \boxed{\sum \alpha y_i = 0} \longrightarrow (9)$$

Sub 8 & 9 in $*$

$$L = \frac{1}{2}(\Sigma \alpha_i y_i \bar{x}_i)(\Sigma \alpha_j y_j \bar{x}_j) - (\Sigma \alpha_i y_i \bar{x}_i)(\Sigma \alpha_j y_j \bar{x}_j) - (\Sigma \alpha_i y_i b) - \Sigma \alpha_i$$

$$L = \frac{1}{2}(\Sigma\Sigma \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j) - (\Sigma\Sigma \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j) - b \cancel{\Sigma \alpha_i y_i}^{0} - \Sigma \alpha_i$$

$$\boxed{L = \Sigma \alpha_i - \frac{1}{2}(\Sigma\Sigma \alpha_i \alpha_j y_i y_j \boxed{\bar{x}_i \bar{x}_j})}$$

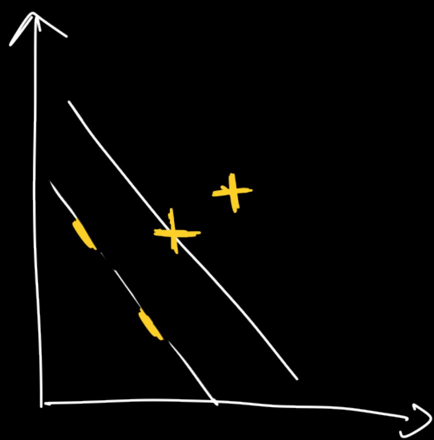
We got this expression to get the dependences

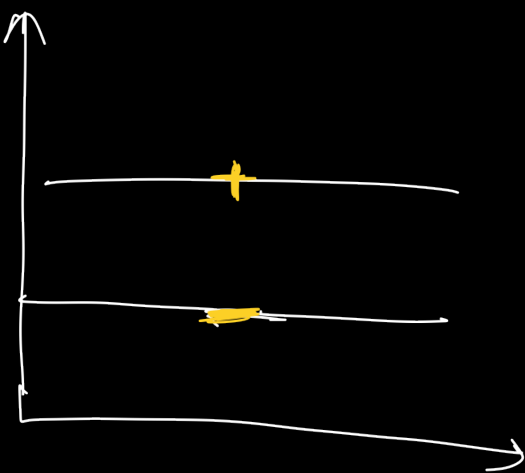
optimization only depends on the dot product of $\bar{x}_i$ and $\bar{x}_j$

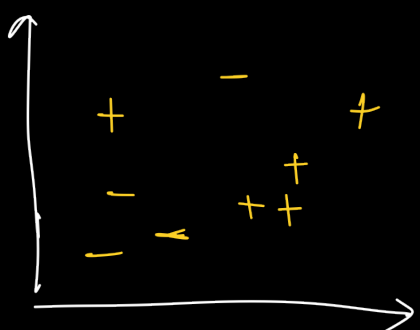
$$\bar{w}\cdot\bar{u} + b \geq 0 \Rightarrow \sum \alpha_i y_i \boxed{\bar{x}_i \, \bar{u}} + b \geq 0$$

then +

depends on only the dot product



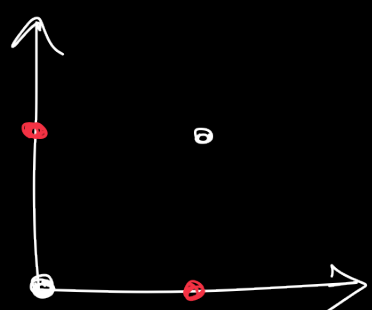




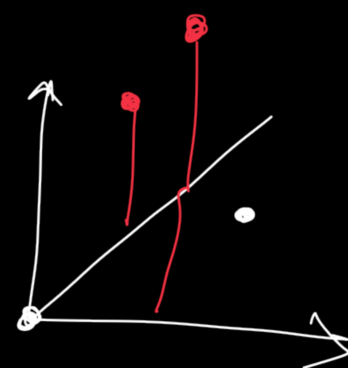
Now it get confused because it is not linearily inseparable.

But then

Suppose



transform $\Rightarrow$ to another space



$\phi(\bar{x}) \rightarrow$ is the transformation function.

and we also know the optimization only depends on the dot product of the sample vectors.

so, $\phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$ to max.

and to recognize we need.

$$\phi(\bar{x}_i) \cdot \phi(\bar{u})$$

$$\boxed{K(\bar{x}_i, \bar{x}_j) = \phi(x_j) \cdot \phi(\bar{x}_j)}$$

Kernel function.

<u>popular Kernels</u>

① linear Kernel

$$(\bar{u} \cdot \bar{v} + 1)^n$$

② $$e^{-\frac{\|\bar{x}_i - \bar{x}_j\|}{\sigma}}$$