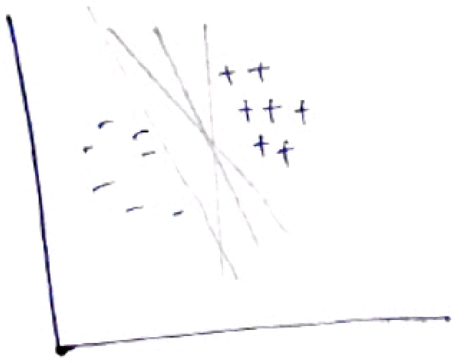
SVM

Prof: Patrick Winston.

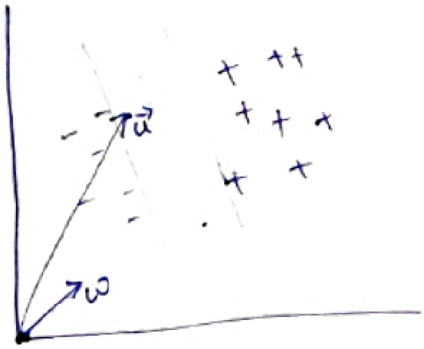
Vladimir Vapnik

Let's say we have positive and negative data samples.



there are infinitely many lines we can draw to separate them. But how do we decide the best line?

→ widest street (margin) possible.



Let $\vec{w}$ be a vector perpendicular to the street.

Let $\vec{u}$ be an unknown, we would like to figure out whether $\vec{u}$ is in the left side of the street? or right side of the street?

Let us project $\vec{u}$ down onto the perpendicular of the street ($\vec{w}$) which would tell us whether it is in left or right side.

$$\vec{w} \cdot \vec{u} \geq c$$ (bigger the projection is more it crosses the street)

$\boxed{\text{decision rule.} \quad \vec{w} \cdot \vec{u} + b \geq 0 \quad \text{then } +}$ where $c = -b$

$\vec{w} \cdot \vec{x}_+ + b \geq 1$ -① $\because$ $\vec{x}_+$ is a positive sample.

$\vec{w} \cdot \vec{x}_- + b \leq -1$ -② $\because$ $\vec{x}_-$ is a negative sample.

for mathematical convenience,

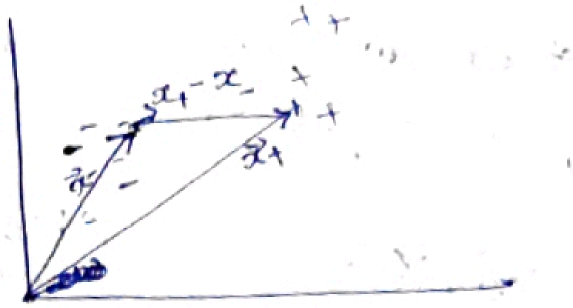
$\exists$ y_i such that $y_i = \begin{cases} +1, & \text{for + sample.} \\ -1, & \text{for − sample.} \end{cases}$

① × y_i.

$$\left.\begin{aligned} &y_i (\vec{w} \cdot \vec{x}_i + b) \geq 1 \\ \text{② × } y_i \quad &y_i (\vec{w} \cdot \vec{x}_i + b) \geq 1 \end{aligned}\right\} \text{ same. ☺}$$

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

$$\boxed{y_i (\vec{w} \cdot \vec{x}_i + b) - 1 = 0}$$ for x in gutter (middle) —— ③



let vector to a positive sample be x_+

" " " negative " " x_-

difference of those vector $x_+ - x_-$

$$\text{width of street} = (x_+ - x_-) \cdot \frac{\vec{\omega}}{\|\omega\|}$$

$$= x_+ \cdot \frac{\vec{\omega}}{\|\omega\|} - x_- \cdot \frac{\vec{\omega}}{\|\omega\|}$$

ω is a normal to the street.

$$\text{~~margin~~} = \frac{1-b}{\|\omega\|} - \frac{-(1+b)}{\|\omega\|}$$

$$= \frac{1-b+1+b}{\|\omega\|}$$

$$= \frac{2}{\|\omega\|}$$ (using equation 3)

we would like to maximize the width of the street.

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

$$\Rightarrow \min \|\omega\|$$

$$\Rightarrow \min \frac{1}{2}\|\omega\|^2$$ (mathematical convenience)

to find extremum of functions with equality constraints we need to use Lagrange multiplier.

$$L(x,\lambda) = f(x) - \lambda g(x)$$

L = Lagrangian
λ = Lagrange multiplier
$g(x)$ = equality constraint
$f(x)$ = function
x = Integer.

that would give us a new expression, which we can maximize or minimize without thinking about the constraints any more.

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

$\alpha_i = \lambda$

to find min or max, set derivative to zero.

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

$$\Rightarrow \boxed{\vec{w} = \sum_i \alpha_i y_i \vec{x}_i}$$

IMP.

decision vector is a linear sum of all the samples.

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

$$\Rightarrow \boxed{\sum \alpha_i y_i = 0}$$

$$L = \frac{1}{2}\left(\sum \alpha_i y \bar{x}_i\right)\left(\sum \alpha_j y_j \bar{x}_j\right) - \left(\sum \alpha_i y_i x_i\right)\cdot\left(\sum \alpha_j y_j x_j\right) - \overbrace{\sum \alpha_i y_i b}^{=0} + \sum \alpha_i$$

$$= \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \, x_i \cdot x_j$$

(depends on this)

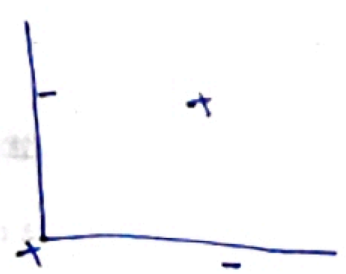
And now the decision rule is:

$$\sum \alpha_i y_i \vec{x}_i \cdot \vec{u} + b \geq 0 \text{ then } +.$$

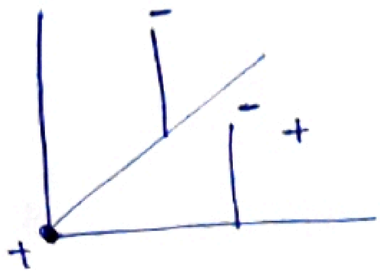
Summary

the whole thing depends on the dot product of all samples ($\vec{x}$) with the unknown ($\vec{u}$)

Till now we were talking about linearly separable data, But lets look at the following



this is not linearly separable so we take to a space where it is ~~separa~~ linearly separable.



transformation ϕ

$\phi(\bar{x})$

$\phi(x_i) \cdot \phi(\cancel{x}_j)$ to max.

$\phi(x_i) \cdot \phi(u)$

$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$

K is kernal function.

we need K and no need of ϕ.

different Kernels.

① choice 1.

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

② choice 2

$e^{-\frac{\| x_i - x_j \|}{\sigma}}$