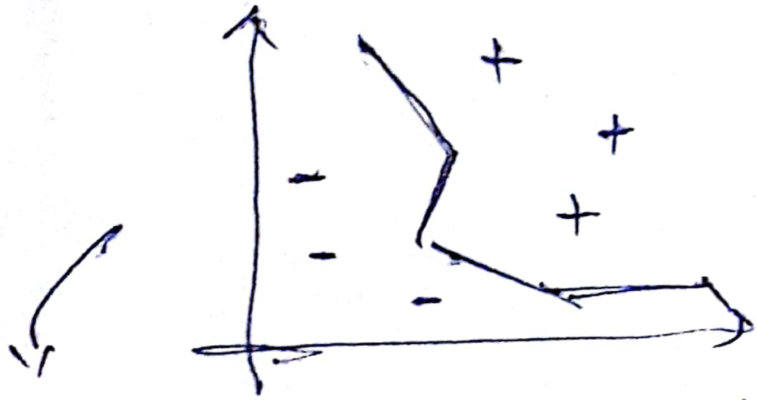
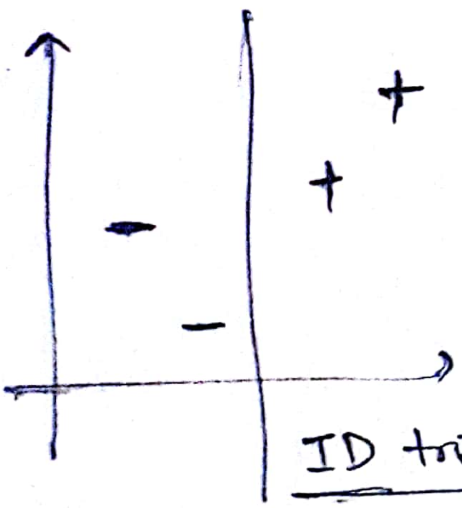
Support Vector machines

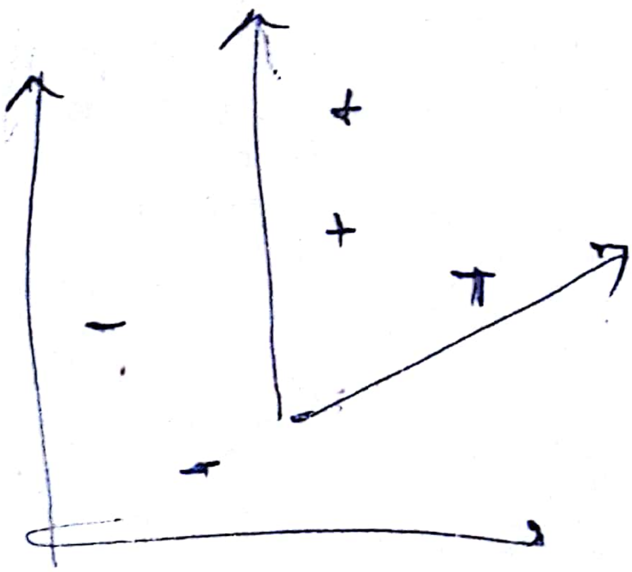
Suppose we have a dataset like below and we wish to obtain a decision boundary separating +ve and -ve samples.



decision boundary obtained by using KNN

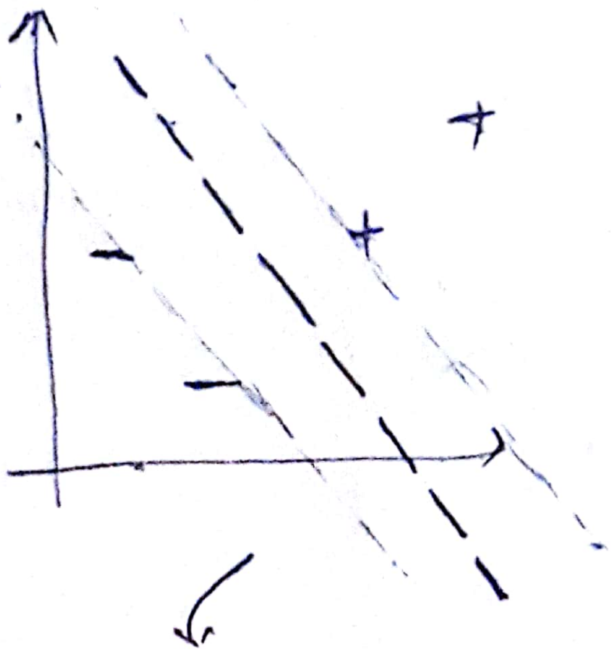


ID trees



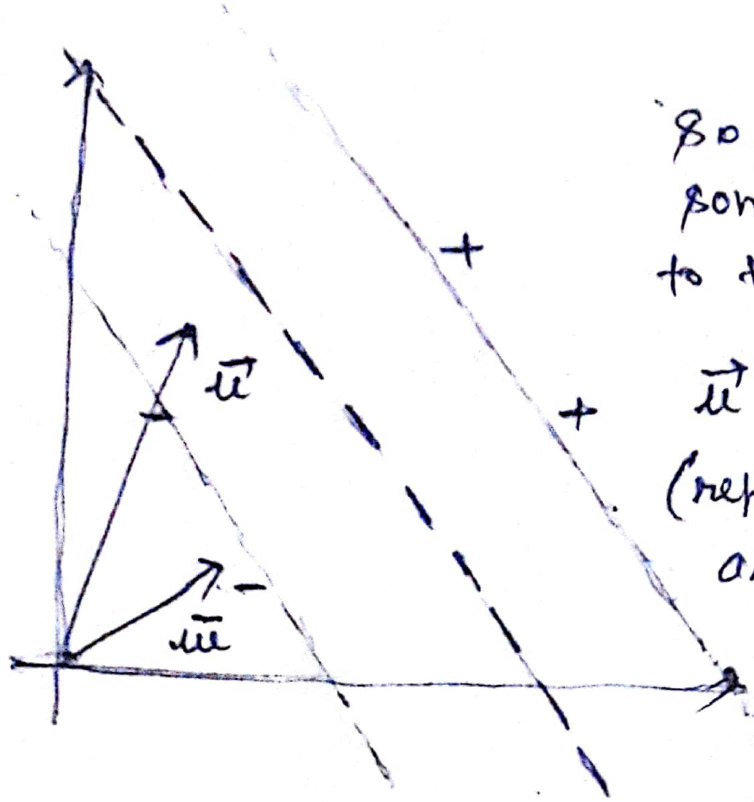
Neural Nets.

But how would a SVM work with such a dataset?



Vladimir Vapnik, introduced SVM's by saying that we draw a line which is farthest from each of the class.

It tries to put a line in between so that the separation b/w the line and the street is as wide as possible.



So, $\vec{w}$ is vector of some length perpendicular to the street.

$\vec{u}$ is some vector (represents a vector to an unknown in the space).

Now we will try and take projection of $\vec{u}$ on the vector $\vec{w}$, and it's length will tell us if the point $\vec{u}$ is in -ve class or +ve class. (3)

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

$$\boxed{b = -c}$$

$$\boxed{\vec{w} \cdot \vec{u} + b \geq 0 \quad \text{Then +ve}}$$ (1)

DECISION RULE

but right now we don't know what $\vec{w}$ to use or which b to use. (there are many $\vec{w}$ possible right now).

* So what we are going to do now is lay down enough constraints so that we'll eventually be able to figure out $\vec{w}$ and b.

Scanned by CamScanner

Now we are going to start out laying some constraints.

$$\boxed{\overline{w}.\overline{x_+} + b \geq 1}$$

We are going to insist that if a sample is a positive sample, then our decision function is going to give a value of 1 or greater than 1.

$$\boxed{\overline{w}.\overline{x_-} + b \leq -1}$$

Similarly, if some sample is a negative sample, we are going to insist that

Now above we have two equations but, carrying around these equations is kinda pain. So I'll try to introduce a variable (y_i) to make life little easier.

$$y_i = \begin{cases} +1 & \text{for +ve samples} \\ -1 & \text{for -ve samples} \end{cases}$$

Multiplying y_i to both the equations.

$y_i(\overline{x_i}\cdot\overline{w}+b) \geq 1\cdot y_i \Rightarrow$ here $y_i = +1$

$\therefore$ $\boxed{y_i(\overline{x_i}\cdot\overline{w}+b) \geq 1}$

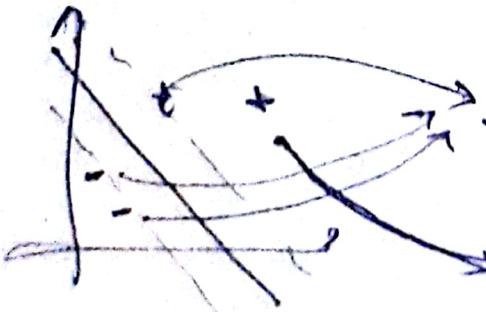
$y_i(\overline{x_i}\cdot\overline{w}+b) \leq -1\cdot y_i \Rightarrow$ here $y_i = -1$

$\therefore$ $\boxed{y_i(\overline{x_i}\cdot\overline{w}+b) \geq 1}$

Now, our purpose of introducing this new variable y_i has been fullfilled, both of our equations are now same.

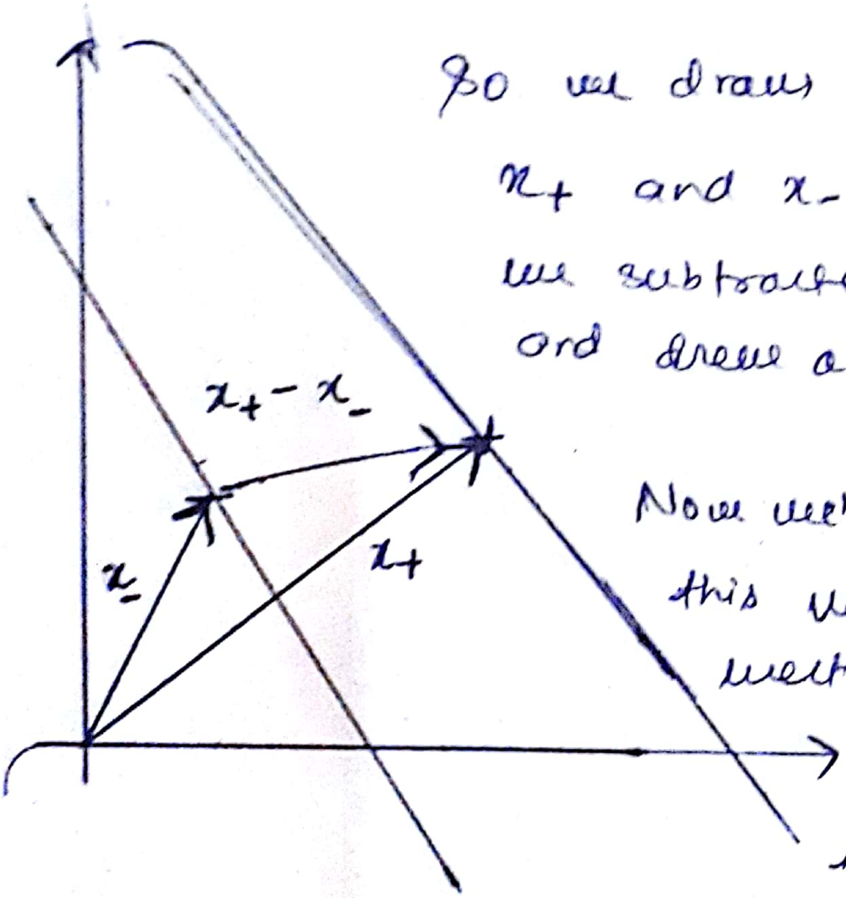
$$\boxed{y_i(\overline{x_i}\cdot\overline{w}+b) - 1 \geq 0} \quad (2)$$

It's always going to be greater than 0, but for $\overline{x_i}$ in the gutter, it's going to be exactly 0.



these points are in the gutter, so it's going to be 0 for them.

greater than 1 for this.



So we draw a few vectors, x_+ and x_- vector and then we subtracted those vectors and drew a $(x_+ - x_-)$ vector.

Now we'll take projection of this vector on a unit vector in a direction $\perp$ to the street i.e. in the direction of vector $\overline{w}$.

$$\boxed{\text{WIDTH} = (\overline{x}_+ - \overline{x}_-) \cdot \left(\frac{\overline{w}}{\|w\|}\right)}$$

Now, let's look at the equation (2),

$$\boxed{y_i(\overline{x_i} \cdot \overline{w} + b) - 1 = 0}$$

for a +ve sample, $\underline{\overline{w} \cdot \overline{x_i} = 1 - b}$.

for a -ve sample, $\underline{-\overline{w} \cdot \overline{x_i} = 1 + b}$.

Putting these back into the above equation,

$$\text{WIDTH} = \frac{(\overline{x}_+ \cdot \overline{w}) - (\overline{x}_- \cdot w)}{||w||} = \frac{(1-b) + (1+b)}{||w||}$$

$$\boxed{\text{WIDTH} = \frac{2}{||w||}}$$ (7)

So, Now what we're trying to do is, we are trying to maximize the width of the street.

∴ what we'll like to do is $\boxed{\text{MINIMIZE } ||w||}$

And to minimize $||w||$, we can also say that, we'll like to $\boxed{\text{MINIMIZE} \left(\frac{1}{2} ||w||^2 \right)}$

(*) And we did that because it's mathematically convinient.

Now if we want to find minimum, extremum of something and we have some constraints that we'd like to honour, we'll use <u>Lagrange's Multipliers</u>

Scanned by CamScanner

Now, we are going to use Lagrange multipliers to find the extremum of our function. (8)

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

And to find the extremum i.e. minimum/maximum we'll find its partial derivatives and set them to 0,

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

$$\Rightarrow \boxed{\bar{w} = \sum_i \alpha_i \, y_i \, x_i}$$

(3)

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i \, y_i = 0$$

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

So we found the derivatives of L w.r.t to whatever might vary. i.e. w and b ~~and~~ and equaled it to 0.

And then we obtain these above two relations.

(9)

Now we have obtained the value of $\overline{w}$ and b which will give us the extremum, so we'll plug these values of $\overline{w}$ and b into the Lagrangian.

$$L = \frac{1}{2}\|w\|^2 - \sum \alpha_i \left[y_i (\overline{w} \cdot \overline{x_i} + b) - 1 \right]$$

$$L = \frac{1}{2}\left(\sum_i \alpha_i y_i \overline{x_i}\right) \cdot \left(\sum_j \alpha_j y_j \overline{x_j}\right) - \sum \alpha_i \cdot y_i \cdot b + \sum \alpha_i$$

$$- \sum_i \alpha_i \cdot y_i \cdot \overline{x_i} \left(\sum_j \alpha_j y_j \overline{x_j}\right)$$

(1) (2) (3) (4)

(2) $-\sum \alpha_i y_i b = -b \underbrace{\sum \alpha_i y_i}_{0} = 0$

Adding (1) and (4)

$$L = -\frac{1}{2} \sum_i \sum_j \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot (\overline{x_i} \cdot \overline{x_j}) + \sum_i \alpha_i$$

$$\boxed{L = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot (\overline{x_i} \cdot \overline{x_j})} \quad (4)$$

Scanned by CamScanner

We went through all this trouble to realize that what's the dependence of this width.
As it turns out, it depends on the sample vectors $\overline{x_i}$ and $\overline{x_j}$ and the optimization depends only on the dot product of the pairs of samples, $(\overline{x_i} \cdot \overline{x_j})$.

Now as we plugged the $\overline{u}$ into the Lagrangian, we are also going to plug into the decision rule, (1) to see what happens.

$$\overline{w} \cdot \overline{u} + b \geq 0 \longrightarrow \text{then the}$$

$$\sum_i \alpha_i y_i \overline{x_i} \cdot \overline{u} + b \geq 0$$

$$\boxed{\sum_i \alpha_i y_i (\overline{x_i} \cdot \overline{u}) + b \geq 0} \longrightarrow \text{then the}$$

So all the math is being dependent on the dot product of $(\overline{x_i})$ and $(\overline{u})$.