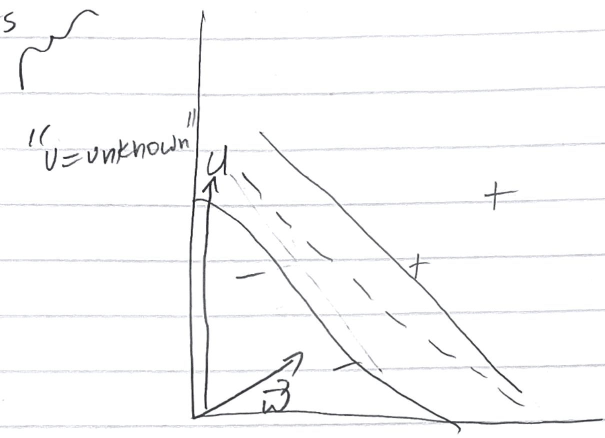
Support Vector Machines MIT OCW

Gutter = area in between the two streets



"This w is a vector of any length that you'd like as long as it's perpendicular to the d.

- we don't know if the O is on the left or right-side of the street !

$\vec{w} \cdot \vec{u} \geq c$ if the dot product between $\vec{w}$ and $\vec{u}$ is big enough then we know that $\vec{u}$ must be a positive sample.

$\Rightarrow$ $\vec{w} \cdot \vec{u} + b \geq 0$ if true then it's a positive sample

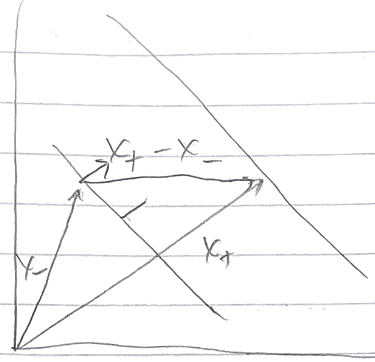
"take $\vec{w}$ and dot product w/ positive sample"

$\vec{w} \cdot \vec{x}_+ + b \geq 1$ and $\vec{w} \cdot \vec{x}_- + b \leq -1$

Introduce y_i s.t. $y_i = \pm 1$ for positive and negative samples.

$y_i (x_i \vec{w} + b) \geq 1$ and $y_i (x_i \vec{w} + 1) \geq 1$

$\Rightarrow$ $y_i (x_i \vec{w} + b) - 1 \geq 0$ also $y_i (x_i \cdot \vec{w} + b) - 1 = 0$ for values in the gutter.



"if there was a unit vector, then you'd dot product the two together and that'd be the width of the street"

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

But we also have $y_i(x_i \cdot \bar{w} + b) - 1 \geq 0$

$$\text{width} = (\underbrace{\bar{x}_+}_{1-b} - \underbrace{\bar{x}_-}_{1+b}) \cdot \frac{\bar{w}}{\|w\|} = \frac{2}{\|w\|} \Rightarrow \text{width} = \frac{2}{\|w\|}$$

Goal now is to find: max $\frac{2}{\|\bar{w}\|}$

or min $\|\bar{w}\|$ or min $\frac{1}{2}\|\bar{w}\|^2$

Lagrange Multipliers

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

$$\frac{\partial L}{\partial \bar{w}} = \bar{w} - \sum \alpha_i y_i \bar{x}_i = 0 \Rightarrow \bar{w} = \sum_i \alpha_i y_i \bar{x}_i$$

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

$$L = \frac{1}{2}\left(\sum \alpha_i y_i \bar{x}_i\right)\left(\sum \alpha_j y_j \bar{x}_j\right) - \sum \alpha_i y_i \bar{x}_i \cdot \left(\sum \alpha_j y_j \bar{x}_j\right)$$

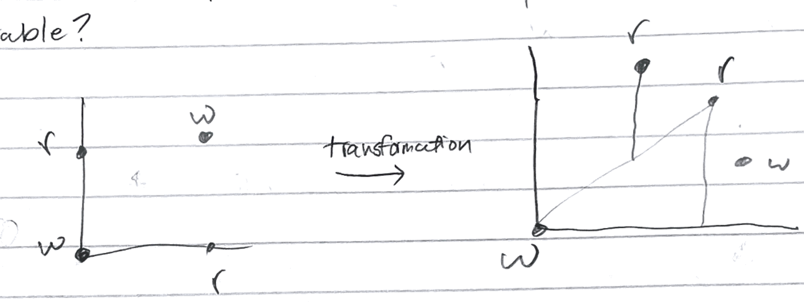
$$- \sum \alpha_i y_i b + \sum \alpha_i$$

$$= 0$$

$$\boxed{L = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \boxed{\bar{x}_i \cdot \bar{x}_j}}$$

☆ The maximization only depends on dot products. ☆

What happens if the points in the space aren't linearly separable?



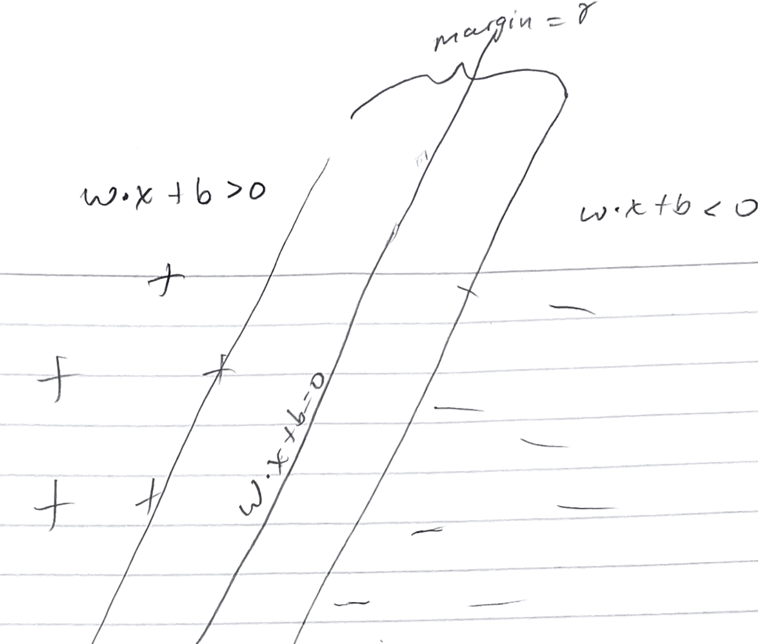
This transformation takes you from a space you're in, to a space where they're linearly separable. This is called $\phi(\bar{x})$.

So just do $\phi(\bar{x}_1) \cdot \phi(\bar{x}_2)$ which is the dot-product of the transformation of two vectors. Need to maximize: $\phi(\bar{x}_1) \cdot \phi(\bar{x}_2)$

Let's say we have a kernel function $K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$ so you don't need the transformation; you just need the kernel function.

Popular kernel functions:

1) $(\bar{u} \cdot \bar{v} + 1)^n$ 2) $e^{-\frac{\|x_i - x_j\|}{\sigma}}$



$y_j \in \{-1, +1\}$ – class it's in

margin := Distance of closest examples from the line / hyperplane. margin $= \gamma = \frac{2a}{\|w\|}$

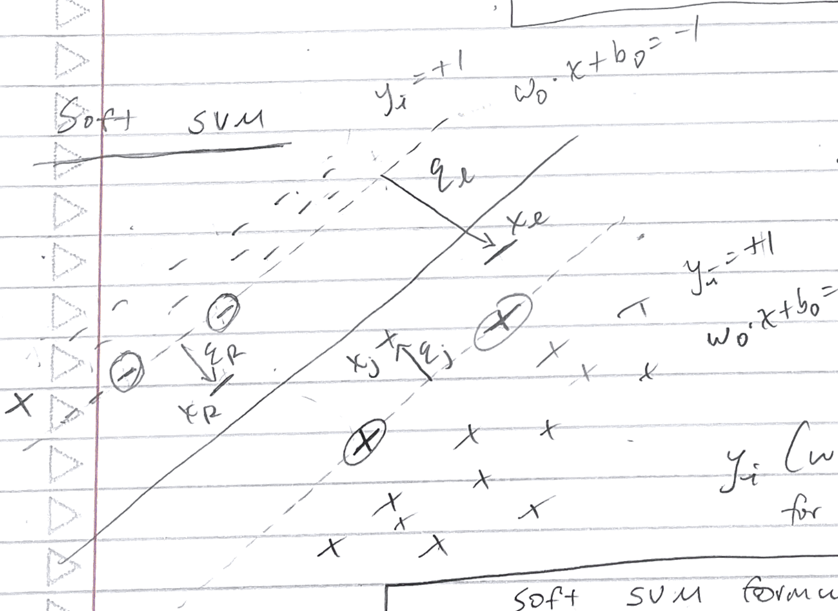
$$\boxed{\max_{w,b} \gamma = \frac{2a}{\|w\|} \quad \text{s.t.} \quad (w \cdot x_j + b)\, y_j \geq a \quad \forall_j}$$

1.) Solve by quadratic programming

2.) Only need to store the support vectors to predict labels of new points.

3.) For all support vectors x_j, $(w \cdot x_j + b)\, y_j = 1$

Soft SVM



$y_i (w_0 \cdot x_i + b_0) \geq 1 - \xi_i$

for point x_i

Soft SVM formula:

$w_0, b_0 = \arg\min_{w,b} \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^{n} \xi_i$

Such that: $y_i (w \cdot x_i + b) \geq 1 - \xi_i$ for all i

$\xi_i \geq 0$