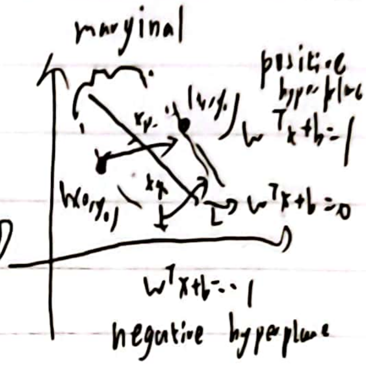
SVM:

1° lead in: WTF hyperplane seperate data points into 2 classes



In math, set 2 points (x_1,y_1) (x_0,y_0), max L

Then: $\vec{w}^T\cdot(\vec{x}_1-\vec{x}_0)=2 \cdots\cdots ①$

$\vec{x}_1=(x_1,y_1)-(0,0)$

$\vec{x}_0=(x_0,y_0)-(0,0)$

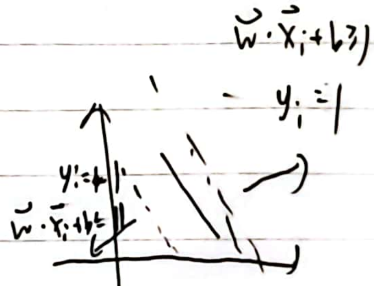
Find 2 points $\vec{x}_p$, $\vec{x}_q$, $\vec{w}^T(\vec{x}_p-\vec{x}_q)=0 \cdots\cdots ②$

①: $||\vec{x}_1-\vec{x}_0||\cos\theta=L$

$||\vec{x}_1-\vec{x}_0||\,||\vec{w}||\cos\theta=2 \Rightarrow L=\frac{2}{||\vec{w}||}$

$\Rightarrow \max L \Leftrightarrow \min||\vec{w}||$

In figure right: let $y_i=\pm1$ indicate the points position in plane



optimal problem:

$\min||\vec{w}||$, s.t. $y_i\cdot(\vec{w}\cdot\vec{x}_i+b)\geq1$, i count out all datas s

make inequality to equality

$$\Rightarrow \min f(w)=\frac{||\vec{w}||^2}{2} \quad (f:\mathbb{R}^n\to\mathbb{R})$$

$$\text{s.t. } g_i(w,b)=y_i*(\vec{w}\cdot\vec{x}_i+b)-1=p_i^2,\ p_i\geq0$$

2° calculate L.

$$\Rightarrow L(w,b,\lambda_i,p_i)=\frac{||\vec{w}||^2}{2}-\sum_{i=1}^{s}\lambda_i*[y_i*(\vec{w}\cdot\vec{x}_i+b)-1-p_i^2]$$

$$\begin{cases}\dfrac{\partial L}{\partial(w,b,\lambda_i,p_i)}=0 \\ \lambda\geq0\end{cases} \Rightarrow \begin{cases}\vec{w}-\sum_{i=1}^{s}\lambda_i y_i\vec{x}_i=0 \cdots\cdots ③ \\ -\sum_{i=1}^{s}\lambda_i y_i=0 \cdots\cdots ④ \\ y_i*(\vec{w}\cdot\vec{x}_i+b)-1-p_i^2=0 \cdots ⑤ \\ 2\lambda_i p_i=0 \Rightarrow \lambda_i p_i^2=0 \cdots ⑥\end{cases}$$

⑤→⑥: $\lambda_i[y_i(\vec{w}\cdot\vec{x}_i+b)-1]=0$

Since $y_i(\vec{w}\cdot\vec{x}_i+b)\geq1$

1° $\lambda_i=0$, $y_i*(\vec{w}\cdot\vec{x}_i+b)-1=0$

Lemma: $\lambda_i \geq 0$

If $\lambda_i < 0 \Rightarrow -\sum_{i=1}^{s} \lambda_i [y_i(\vec{w}\cdot\vec{x}_i + b) - 1 - p_i^2] < 0$

Then larger the $y_i \cdot (\vec{w}\cdot\vec{x}_i + b)$, smaller the $\|\vec{w}\|$ which violate the constraint. $\square$

Simpler: $\nabla f \in \mathrm{Span}\{\nabla g_i\}$, and 

$\Rightarrow \lambda_i \geq 0$

3* translate to dual problem, set optimal w^*, b^*

$q(\lambda_i) \triangleq \min[L(w,b,\lambda_i)] \leq f(\vec{w}) - \sum_{i=1}^{s} \lambda_i g_i(\vec{w}^*, b^*)$

Since $\lambda_i \geq 0$, $g_i(\vec{w}^*, b^*) \geq 0 \Rightarrow \lambda_i \cdot g_i(\vec{w}^*, b^*) \geq 0$

So $q(\lambda_i) \leq f(\vec{w}^*) - \sum_{i=1}^{s} \lambda_i \cdot g_i(\vec{w}^*, b^*) \leq f(\vec{w}^*) \leq f(w)$

$q(\lambda_i^*) < f(\vec{w})$ weak dual.

$q(\lambda_i^*) = f \Rightarrow$ strong dual

WTF λ_i^* s.t. $q(\lambda_i) \leq q(\lambda_i^*) \leq f(\vec{w}^*) \leq f(\vec{w})$

$[q(\lambda_i^*) \to f(\vec{w}^*)]$

dual problem: $\max q(\lambda_i) = \max[\min L(w,b,\lambda_i)]$

s.t. $\lambda_i \geq 0$, $i \in [\![1, s]\!]$.

Lemma: $q(\lambda_i^*) = f(\vec{w}^*)$, problem and dual problem optimal isamoterially

proof: $f(w) \geq q(\lambda_i^*) = f(\vec{w}^*) \Rightarrow f(\vec{w}^*) \leq f(w)$

$f(w) \geq q(\lambda_i^*) = f(\vec{w}^*) \geq q(\lambda_i) \Rightarrow q(\lambda_i^*) \geq q(\lambda_i)$ $\square$

eg1: original problem: $\min f(x) = x^2$ s.t. $x - 1 \geq 0$

construct $q(\lambda) = \min[L(x,\lambda)] = \min[x^2 - \lambda(x-1)]$, s.t. $\lambda \geq 0$

$\frac{\partial L}{\partial x} = 0 \Rightarrow x = \frac{\lambda}{2} \Rightarrow q(\lambda) = -\frac{\lambda^2}{4} + \lambda$

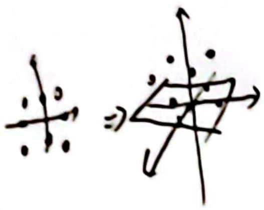
$\max[q(\lambda)] = 1$ when $\lambda = 2$, $x = 1$

use dual problem to simplify original problem

$$\max q(\lambda) = \max\left(\sum_{i=1}^{s}\lambda_i - \frac{1}{2}\sum_{i=1}^{s}\sum_{j=1}^{s}\lambda_i\lambda_j y_i y_j (\vec{x}_i\cdot\vec{x}_j)\right)$$

s.t. $\lambda_i \geq 0$, $i\in[1,s]$

Solve $\lambda_i \xRightarrow{①} \vec{w} \Rightarrow y_i(\vec{w}\cdot\vec{x}_i+b)-1=0$, sol b.



4th kernel trick: some problem can solve in higher dimension by SVM: $x_i \to T_j(x_i)$

$$\max q(\lambda_i) = \max\left(\sum_{i=1}^{s}\lambda_i - \frac{1}{2}\sum_{i=1}^{s}\sum_{j=1}^{s}\lambda_i\lambda_j y_i y_j T(\vec{x}_i)\cdot T(\vec{x}_j)\right)$$

s.t. $\lambda_i \geq 0$; $i\in[1,s]$

1° $T(\vec{x}_i)$, $T(\vec{x}_j)$ first, $\sum_{i=1}^{n} T(x_i)T(x_j)$

2° $T(\vec{x}_i)\cdot T(\vec{x}_j) = K(\vec{x}_i,\vec{x}_j)$, $\sum_{i=1}^{n}\sum_{j=1}^{n} K(\vec{x}_i,\vec{x}_j)$

normally: $K(\vec{x}_i,\vec{x}_j) = (1+\vec{x}_i\cdot\vec{x}_j)^d$

use T, s.t. $\vec{x} \in (x_1,x_2) \xrightarrow{T} T(\vec{x}) = (a_1x_1, a_1x_2, a_1x_1x_2, a_2x_1^2, a_2x_2^2, \cdots, a_nx_1^n, a_nx_2^n, \cdots \infty)$

$T: \mathbb{R}^2 \to \mathbb{R}^\infty$

1° can't be done

2°: RBF function: $K(\vec{x}_i,\vec{x}_j) = e^{-\gamma\|\vec{x}_i-\vec{x}_j\|^2}$

$K(\vec{x}_i,\vec{x}_j) = C\, e^{\langle \vec{x}_i, \vec{x}_j\rangle}$, $(C = \exp[-\frac{1}{2}(\|\vec{x}_i\|^2+\|\vec{x}_j\|^2)])$

$K(\vec{x}_i,\vec{x}_j) = C\sum_{n=0}^{\infty}\frac{(\vec{x}_i\cdot\vec{x}_j)^n}{n!} = C\sum_{n=0}^{\infty}\frac{K_{poly(n)}(\vec{x}_i,\vec{x}_j)}{n!}$