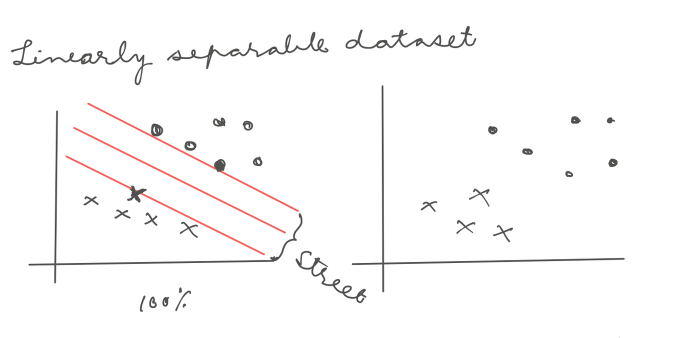
Support Vector Machine

↳ Widest street approach
↳ Pre-deep ML → most popular

Linearly separable dataset

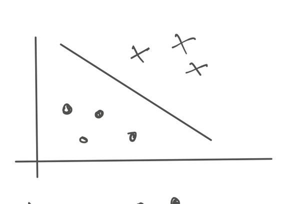


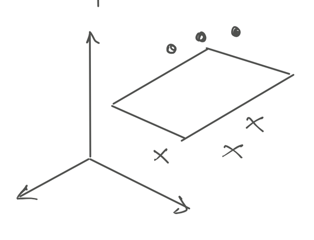
Let me find the decision boundary that maximizes the street.

max S ⟵ math

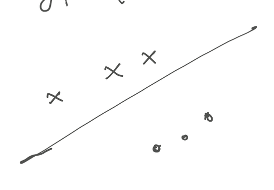
Plane.
↳ hyperplane

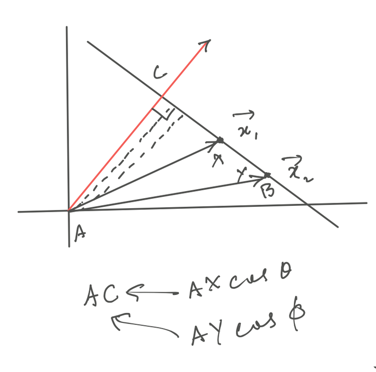
1D o o o o ☆ × × × point

2D  line

3D  plane

>3D hyperplane





$AC \leftarrow AX\cos\theta$
$\leftarrow AY\cos\phi$

Dot product → $\vec{A}\cdot\vec{B} = \|A\|\,\|B\|\cos\theta$

$$\frac{\vec{A}\cdot\vec{B}}{\|A\|} = \vec{B}\cdot\frac{\vec{A}}{\|A\|}$$

Unit vector → Direction
→ Mag $=1$

Vector $\vec{w}$ which normal to our plane

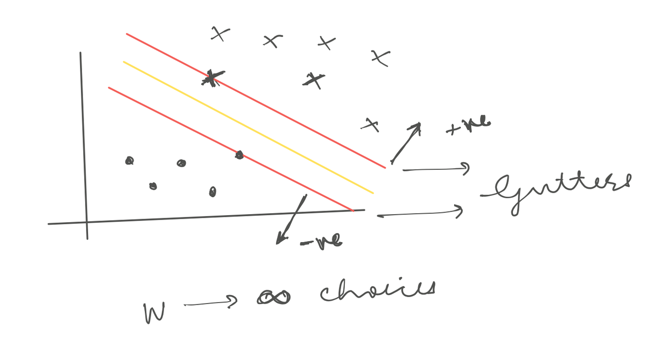
$$\frac{\vec{W}\vec{x}}{\|W\|} = c$$

$\Rightarrow \vec{W}\cdot\vec{x} = \|W\|c$

$\Rightarrow \vec{W}\cdot\vec{x} - \|W\|c = 0$

$\Rightarrow \boxed{\vec{W}\cdot\vec{x} - b = 0}$

$\forall x$ lie on the plane



$W \to \infty$ choices

$W,\ 2W,\ 3W$

$\boxed{Wx_+ + b \geq 1}$ & $\boxed{Wx_- + b \leq -1}$

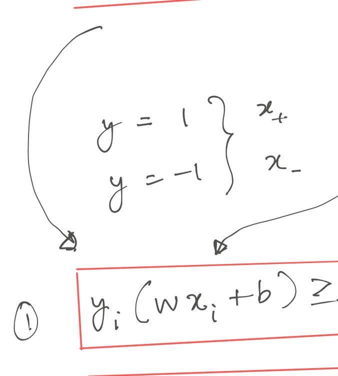
$-(Wx_- + b) \geq 1$

$\left.\begin{matrix} y=1 \\ y=-1 \end{matrix}\right\}\begin{matrix} x_+ \\ x_- \end{matrix}$

① $\boxed{y_i(Wx_i + b) \geq 1}$ $\forall i$ not in the street

② $\boxed{y_i(Wx_i + b) = 1}$ $\forall i$ on the gutter

Goal : Max. width of street
↳ Max. gap between gutters



$\vec{x}_+ - \vec{x}_- = \vec{y}$
$\vec{x}_- + \vec{y} = x_+$

Project $(\vec{x}_+ - \vec{x}_-)$ on $\vec{w}$.

③ $\boxed{(\vec{x}_+ - \vec{x}_-)\cdot\frac{\vec{W}}{\|W\|} = \text{Street width}}$

② → $y_i(\vec{w}\vec{x}_i + b) = 1$ ⟵ on gutters

+ve class	-ve class
$+1(\vec{W}\vec{x}_+ + b) = 1$	$-1(\vec{W}\vec{x}_- + b) = 1$
$\boxed{\vec{W}\vec{x}_+ = 1-b}$	$-\vec{W}\vec{x}_- - b = 1$
Wx_+	$\boxed{\vec{W}\vec{x}_- = -1-b}$
	Wx_-

Substituting in ③

$$(\vec{x}_+ - \vec{x}_-)\frac{\vec{W}}{\|W\|}$$

$$= \frac{\vec{W}\vec{x}_+ - \vec{W}\vec{x}_-}{\|W\|}$$

$$= \frac{(1-b)-(-1-b)}{\|W\|}$$

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

$$\boxed{= \frac{2}{\|W\|}\ \Big\}\ \text{Street width}}$$

Max

SVM → Max $\frac{2}{\|W\|}$ st. $\boxed{\begin{matrix} y_i(\vec{w}\vec{x}_i + b) = 1 \leftarrow \text{on gutters} \\ y_i(\vec{w}\vec{x}_i + b) \geq 1 \leftarrow \text{outside the street} \end{matrix}}$

$\simeq$ min $\|W\|$ st.

$\simeq$ min $\frac{1}{2}\|W\|^2$ st.

$\simeq$ $\boxed{\min \frac{1}{2}\vec{W}\cdot\vec{W}}$ st.

Constrained Optimization problem
→ Quadratic obj
→ Linear constraints } QP } QP solver

Lagrange multipliers

OT → friends

$f(w)$
$\left.\begin{matrix} g_i(w) \leq 0 \\ h_i(w) = 0 \end{matrix}\right\}$

$$\mathcal{L}(w,\alpha,\beta) = f(w) + \sum_i \alpha_i g_i(w) + \sum_i \beta_i h_i(w)$$

↳ constraints

$$\theta_p = \max_{\alpha,\beta;\alpha\geq 0} \mathcal{L}(w,\alpha,\beta) = \begin{cases} f(w), & \text{when const. are satisfied} \\ \infty, & \text{otherwise} \end{cases}$$

Satisfied
$g_i(w) \leq 0\ \forall i$
↳ -ve
↳ $\alpha = 0$

$h_i(w) = 0\ \forall i$

Not satisfied
$\exists i,\ g_i(w) > 0$
↳ $\alpha_i = \infty$

$\exists i,\ h_i(w) \neq 0$
↳ $\beta_i \in \{-\infty, \infty\}$

Primal problem

$$f(w) = \max_{\alpha,\beta;\alpha\geq 0} \mathcal{L}(w,\alpha,\beta)$$

① $\min_w f(w) = \min_w \max_{\alpha,\beta;\alpha\geq 0} \mathcal{L} = p^*$

Dual problem

$$\theta_D = \min_w \mathcal{L}(w,\alpha,\beta)$$

② $\max_{\alpha,\beta;\alpha\geq 0} \theta_D = \max_{\alpha,\beta;\alpha\geq 0} \min_w \mathcal{L}(w,\alpha,\beta) = d^*$

OP → $\max_x \min_y f(x,y) \leq \min_y \max_x f(x,y)$
↳ $d^* \leq p^*$

$$\max_{\alpha,\beta;\alpha\geq 0} \min_w \mathcal{L} \leq \min_w \max_{\alpha,\beta;\alpha\geq 0} \mathcal{L}$$

Conditions → True → $d^* = p^*$

(i) $f(w)$ and $g_i(w)$ are convex

(ii) $h_i(w)$ are affine

$$\underbrace{\underbrace{W^T x}_{\text{Linear}} + b}_{\text{Affine}}$$

(iii) $g_i(w)$ is strictly feasible
↳ $\exists w,\ g_i(w) < 0\ \forall i$

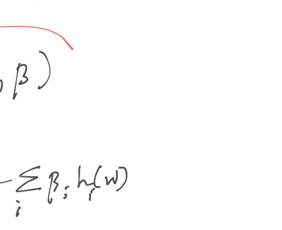
Check

✓ i $f(w) = \frac{1}{2}w^T w$ ⟵ Linear ⟵ Convex ✓
$g_i(w) = -y_i(w^T x + b) + 1$ ⟵ Affine

✓ ii $h_i(w)$ ⟵ affine
↳ we don't have $h_i(w)$

✓ iii $g_i(w) < 0\ \forall i, \exists w$

choose a tighter street



$\left.\begin{matrix} y_i(w^T x + b) + 1 \geq 0 \\ -y_i(w^T x + b) - 1 \leq 0 \end{matrix}\right\}$

from i, ii, iii
↳ $d^* = p^*$
↳ $\max_{\alpha,\beta,\alpha\geq 0} \min_w \mathcal{L}(w,\alpha,\beta)$

we can minimize without any constraints

$$\min_w \mathcal{L} = \min_w f(w) + \sum_i \alpha_i g_i(w) + \sum_i \beta_i h_i(w)$$

$$= \min_w \frac{1}{2}w^T w + \sum_i \alpha_i\left(-y_i(w^T x + b) + 1\right)$$

$$= \min_w \frac{1}{2}w^T w - \sum_i \alpha_i y_i w^T x_i - \sum_i \alpha_i y_i b + \sum_i \alpha_i$$

To minimize
↳ $\frac{\partial \mathcal{L}}{\partial w} = 0$

$\Rightarrow w - \sum_i \alpha_i y_i x_i = 0$

$\Rightarrow \boxed{w = \sum_i \alpha_i y_i x_i}$ — ③

↳ $\frac{\partial \mathcal{L}}{\partial b} = 0$

$\boxed{\sum_i \alpha_i y_i = 0}$ — ④

Plugging ③ & ④
↳ $\min_w \mathcal{L} = \frac{1}{2}\left(\sum_i \alpha_i y_i x_i\right)^T\left(\sum_j \alpha_j y_j x_j\right)$

$$- \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j x_j\right)^T x_i$$

$$- 0 \quad \longleftarrow \text{from 4}$$

$$+ \sum_i \alpha$$

$$\Rightarrow \min_w \mathcal{L} = \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i^T x_j - \sum_{i,j}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_i \alpha$$

$$\Rightarrow \min_w \mathcal{L} = \sum_i \alpha - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

Now $\max_{\alpha,\alpha\geq 0}\left(\min_w \mathcal{L}\right)$ max this wrt. α, $\alpha \geq 0$ easier
↳ Quadratic Problem

From 3 → $\boxed{w = \sum_i \alpha_i y_i x_i}$ Too many α_i's?

From i, ii, iii → KKT conditions

$\boxed{\forall i\quad \alpha_i g_i(w) = 0}$

when $g_i(w) < 0 \to \alpha_i = 0$

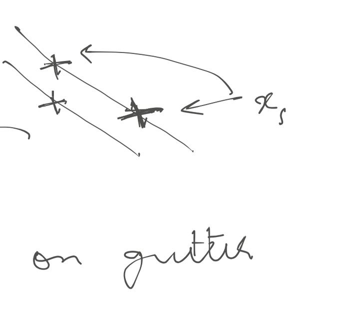
when $g_i(w) = 0 \to \alpha_i > 0$

$g_i(w) < 0$ → you are not on the gutter

$g_i(w) = 0$ → you are on the gutters

⟹ only points on the gutter define w.

set : $\{\alpha_1\ \alpha_2\ \alpha_3\}$



$b = ?$

For $x_s \to g_i(w) = 0$ on gutter

$\Rightarrow -y_s(w^T x_s + b) + 1 = 0$

$\Rightarrow y_s(w^T x_s + b) = 1$

$\Rightarrow w^T x_s + b = y_s$

$\Rightarrow \boxed{b = y_s - w^T x_s}$