Module 6 : 4 Lectures 7/9 – 7/18

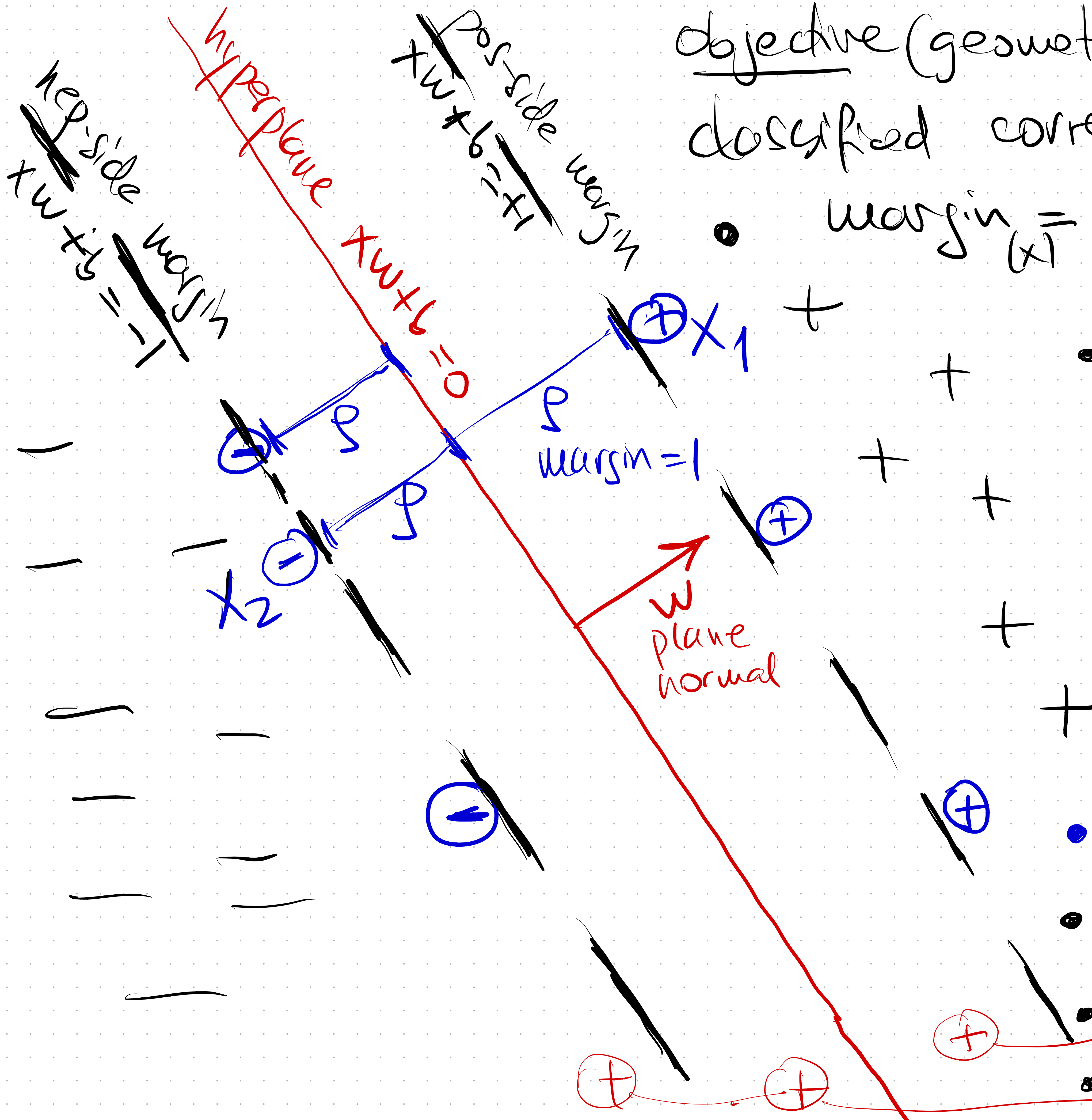
- Kernels
- Kernelization / dualization of ML algorithms
- $\boxed{\text{SVM}}$, Sequential Minim Optimization
- active learning

Support Vector Machine : linear classifier

binary labels $y \in \{\pm 1\}$

$$h(x) = xw + \underbrace{b}_{\text{bias}}$$

$\boxed{\text{objective}}$: increase geometrical margin to closest points



<u>objective</u> (geometry): want all points classified correctly

- $\text{margin}_{(x)} = (xw+b)\cdot \boxed{y} > 0$ (y: label)

CONFIDENCE

- margin far from plane

$\text{margin}(x) \geq 1$
(both sides)

<u>blue</u>: points closest to margin = 1

points type

- support $(xw+b)y = 1$
- loose (correct) $(xw+b)y > 1$
- tight (correct) $1 > (xw+b)y > 0$
- mistake $(xw+b)y < 0$

(x_2) $x_2 w + b = \boxed{-1}$

(x_1) : $x_1 w + b = \boxed{+1}$

fix scale

$\Rightarrow (x_1 - x_2)\, w = 1 - (-1) = 2$

geometry : $2\rho = \|x_1 - x_2\|$

$\Rightarrow \|(x_1 - x_2)\, w\| = 2$

$2\rho \cdot \|w\| = 2$

$\rho = \dfrac{1}{\|w\|}$

ISOLATE geom for support vectors

want margin ρ (to support vectors) as high as possible and equal on both sides.

SVM problem : find separation params (w, b) OBJ : Quad

(constrained optimization)

$$\max \rho \iff \min \frac{1}{2}\|w\| \iff \boxed{\min \frac{1}{2}\|w\|^2}$$

subject to : ~~constraints = linear~~

$$\boxed{(x_i w + b)\, y_i \geq 1} \quad \text{for all points } (x^i, y_i)$$

y_i: $+1$ or -1

all points have margin ≥ 1

Support vectors = points with margin ≤ 1

$(xw + b) = 1$

Solution part 1: Constrained optimization $\Rightarrow$ Lagrangian Multiplier

$$L(w,b,\alpha) = \underbrace{\frac{1}{2}\|w\|^2}_{\text{OBJ}} - \sum_{i=1}^{N} \boxed{\alpha_i}\underbrace{[y_i(x_i w + b) - 1]}_{\text{constraint}}$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i y_i x_i^T \overset{\text{want}}{=} 0$$

$$\Rightarrow w = \sum_{i=1}^{N} \alpha_i y_i x_i^T$$

= linear combination of datapoints with coef $(\alpha_i)_{i=1:N}$

w

PRIMARY VARIABLES

$h(x_j) = x_j w$

α

$\boxed{\text{DUALITY}}$

DUAL VARIABLE

$$\overset{1\times1}{h(x_j)} = x_j w = x_j \sum_{i=1}^{N} \alpha_i y_i x_i^T$$

$$= \sum_{i=1}^{M} \alpha_i y_i \boxed{\underset{1\times D}{x_j} \cdot \underset{D\times1}{x_i^T}}$$

K_{ij}

__NEXT__: rewrite $L()$ as function of α, not w

- solve for α vars with Quad Optimization