FoDA : Support Vector

L27 : Machines (SVMs)

& Kernels

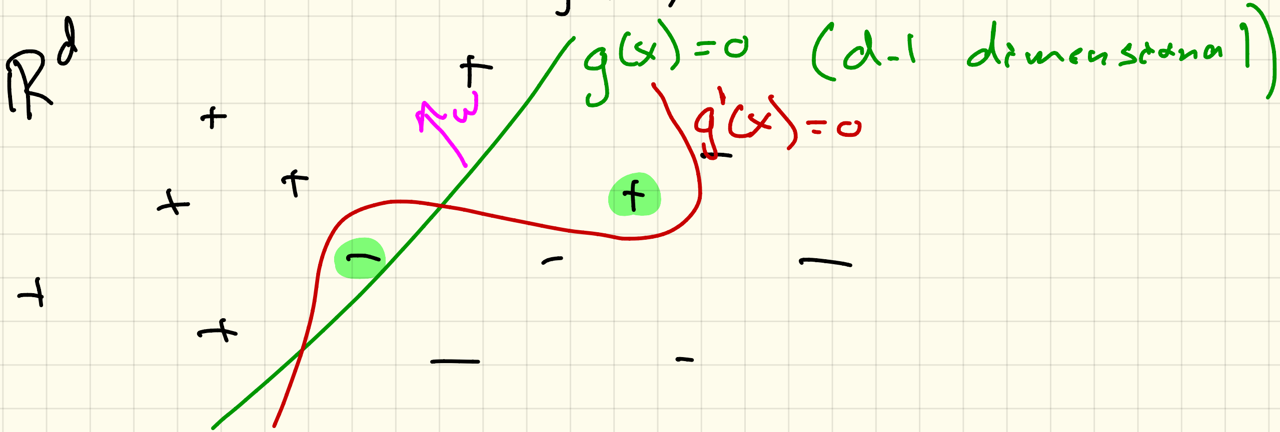
Classification

Input $X \subset \mathbb{R}^d$ labels $y \in \{-1, +1\}^n$

$x_1, \dots, x_n$

goal function $g : \mathbb{R}^d \to \mathbb{R}$

so $g(x_i) > 0$ iff $y_i = +1$

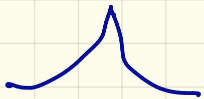


linear $\quad g(x) = \langle x, w \rangle + b \quad \leftarrow$ keep

linear $\langle x, w \rangle = \sum_{i=1}^{d} x_i \cdot w_i$

replace $\langle x, w \rangle$ w/ $\langle x, w \rangle_K$

Gaussian $\quad K(x, w) = \exp\left(-\|w - x\|^2 / \sigma^2\right)$ 

Laplace $\quad K(x, w) = \exp\left(-\|w - x\| / \sigma\right)$ 

Polynomial $\quad K(x, w) = \left(\langle w, x \rangle + 1\right)^{r}$

Kernel Expansion

map $p \in \mathbb{R}^d$ to point q in $\mathbb{R}^{O(d^r)}$

$p \to q = (q_1 = p_1, \; q_2 = p_2, \; q_3 = p_1^2, \; q_4 = p_2^2, \; q_5 = p_1 \cdot p_2)$

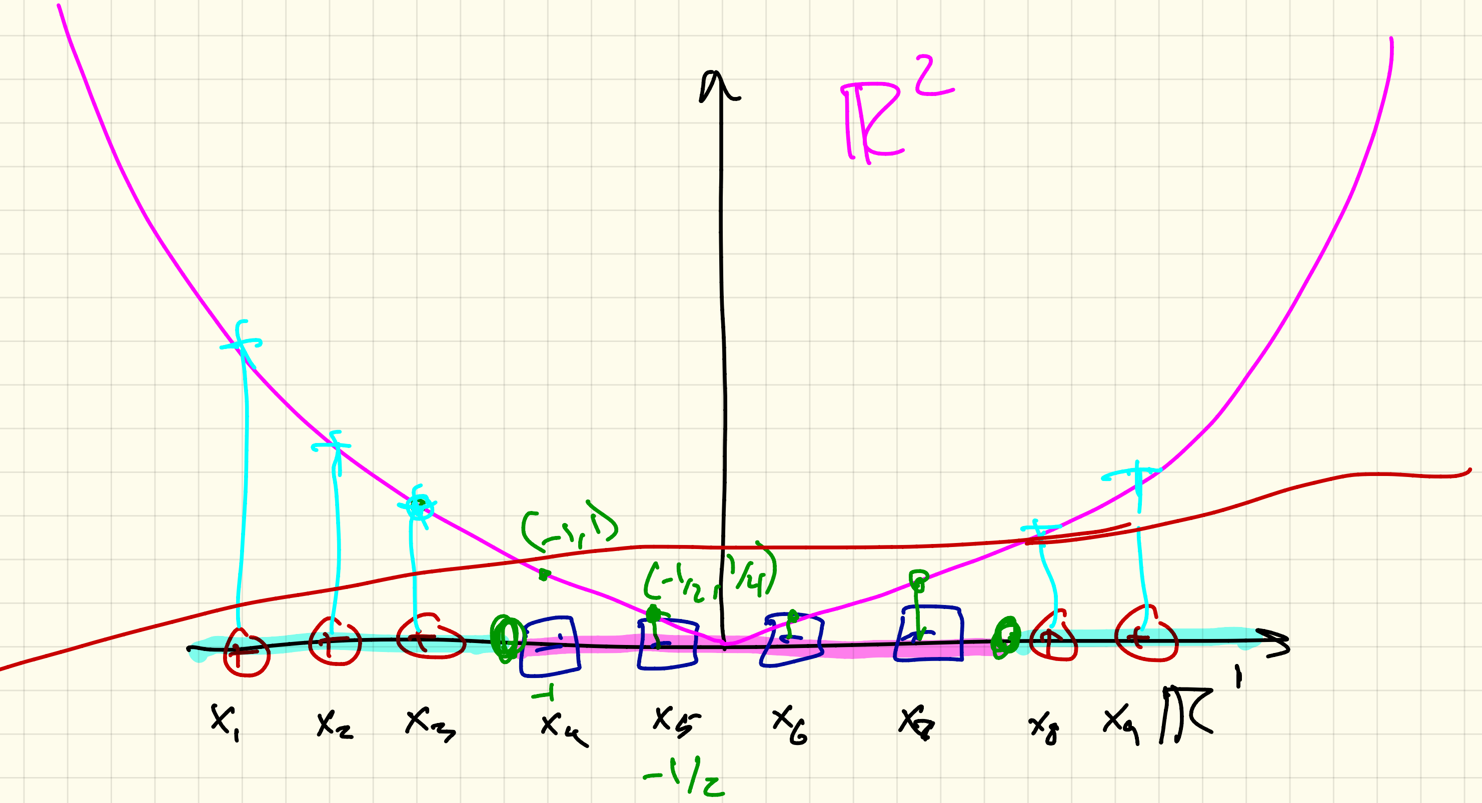
$p = (p_1, p_2) \in \mathbb{R}^2$ $\qquad \in \mathbb{R}^5$

$\alpha_1 \quad \alpha_2 \quad \alpha_5$

<u>model</u> $\alpha \in \mathbb{R}^5$ $\quad \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_5)$

Ways to expand Gaussian / Laplace Kernels to $\mathbb{R}^m$ also

exactly $m = \infty$, approximately $m \geq 100$

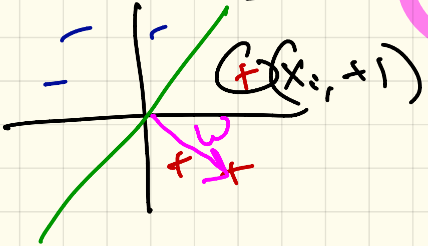


$$p = x \Longrightarrow g = (x, x^2) \in \mathbb{R}^2$$

Kernel Perceptron: Dual, Mistake Counter

model

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$



α_i = # times we added $y_i x_i$

$w \in \mathbb{R}^d$

$d < n$

mistakes on (x_i, y_i)

$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{R}^n$

$\# \alpha_i \neq 0 \leq \left(1/\gamma^*\right)^2$

New data point $p \in \mathbb{R}^d$

$$g(p) = \langle w, p \rangle = \left\langle \sum_{i=1}^{n} \alpha_i y_i x_i, p \right\rangle = \sum_{i=1}^{n} \alpha_i y_i \overbrace{\langle x_i, p \rangle}^{K(x_i, p)}$$

Kernel Perceptron $K(x, p)$

$$g(p) = \sum_{i=1}^{n} \alpha_i \, y_i \, K(x_i, p)$$

mistake counter

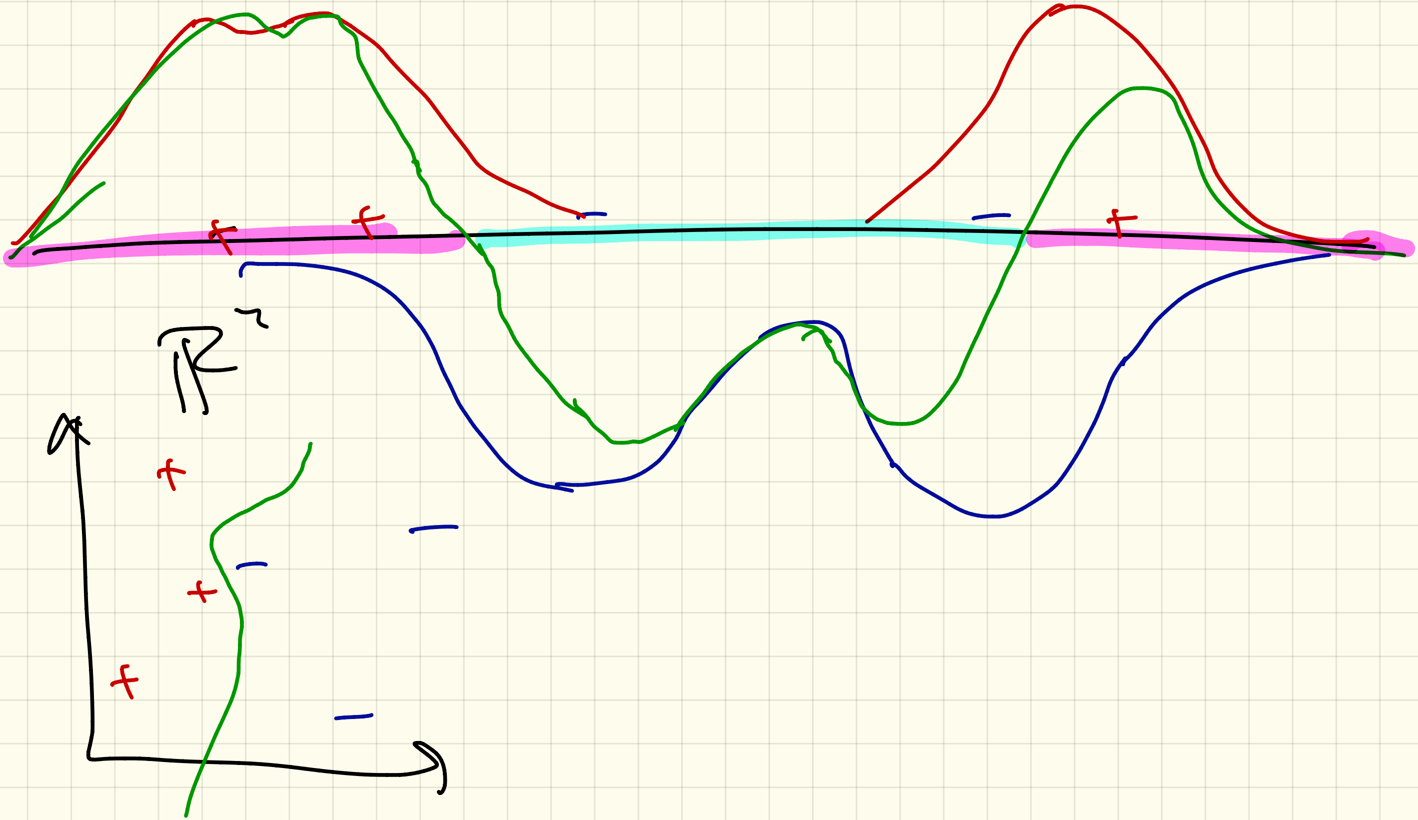
0. $\alpha = (0, 0, \ldots 0) \in \mathbb{R}^n$

$\alpha^{(1)} = (1, 0, \ldots 0)$ for $(x_1, y_1 = +1)$

1. repeat

if exists (x_i, y_i): $\operatorname{sign}(g(x_i)) \neq y_i$

then $\alpha_i \mathrel{+}= 1$


$\mathbb{R}^2$

Support Vector Machine

$$w = \sum_{j=1}^{k} \alpha_j \overbrace{y_j K(x_j, \cdot)}^{g_j \in \mathbb{R}^m}$$

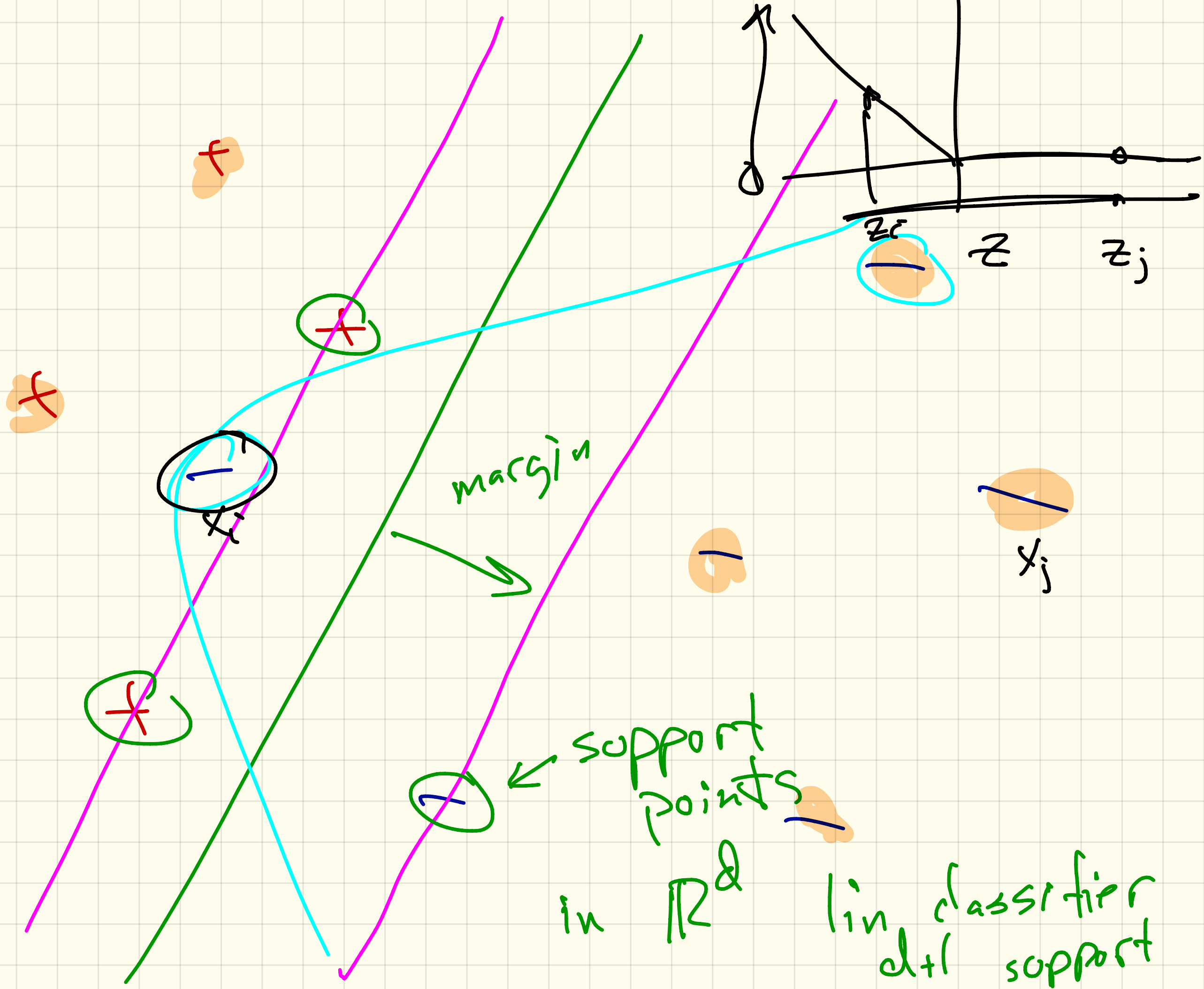
in mistake counter $\xrightarrow{}$ $\alpha_j \in [0, 1, 2, \dots]$

$\xrightarrow{w/\ GD}$ $\alpha_j \in \mathbb{R}$

$$g(p) = \sum_{j=1}^{k} \underbrace{\alpha_j}_{\in \mathbb{R}} \cancel{y_j} K(x_j, p)$$

subset $S \subset X$

$S = \{s_1, s_2, \dots s_k\}$ ← support vectors



margin
support points
x_i
x_j
z_i
z
z_j
in $\mathbb{R}^d$
lin classifier
$d+1$ support

K SVM

1. Identify support Points

$S \subset X \qquad s_1 \dots s_k$

- Kernel perceptron
- Choose $S = X$
- ↳ $S = X$ + SGD

2. $g(p) = \sum_{j=1}^{k} \alpha_j K(s_j, p)$

optimize over $\alpha \in \mathbb{R}^k$

$z_i = y_i \, g(x_i) = y_i \sum_{j=1}^{k} \alpha_j K(s_j, x_i)$