FoDA L27

Support Vector Machines (SVMs), Kernels, and non-linear classification

Perceptron Algorithm

Input $X, y \in \mathbb{R}^{d}$

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

so, $y_i \langle w, x_i \rangle > 0$

$\text{sign}(\langle w, x_i \rangle) = \text{sign}(y_i)$

$w = y_i x_i$ for any $(x_i, y_i) \in X, y$

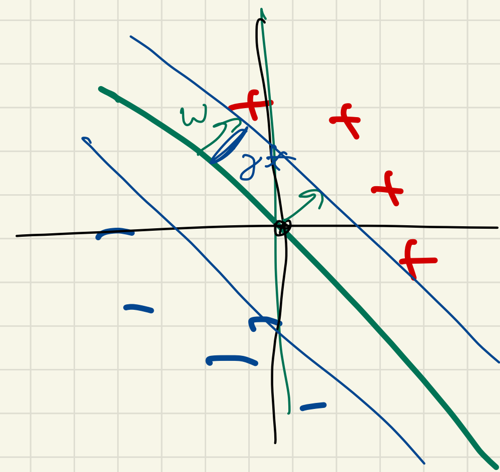
repeat

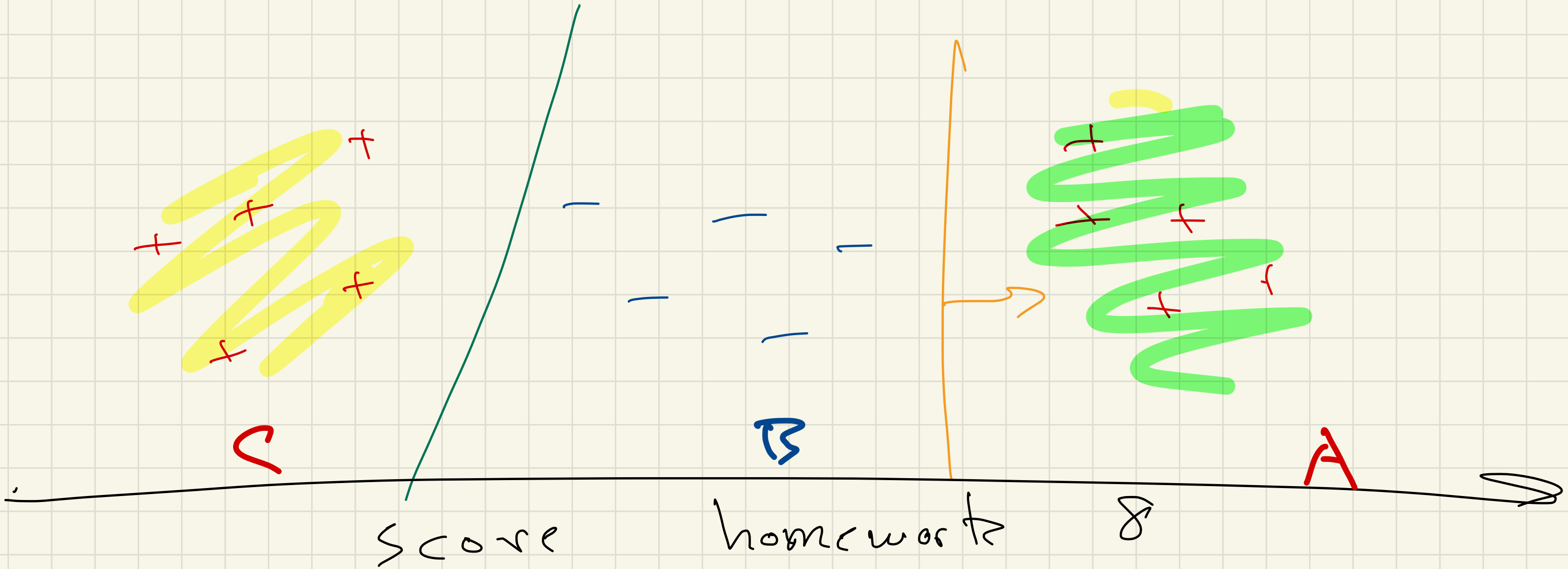
if any x_i, y_i s.o.

$y_i \langle x_i, w \rangle < 0$

then update $w = w + y_i x_i$

return $w / \|w\|$





− if student earns a B

+ if student earns not a B

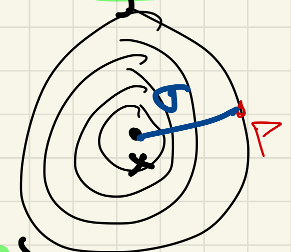
Generalized Inner Products

↳ Kernels

$p, x \in \mathbb{R}^d$

linear inner product
dot product

$$\langle p, x \rangle = \sum_{j=1}^{d} p_j x_j$$



Kernel

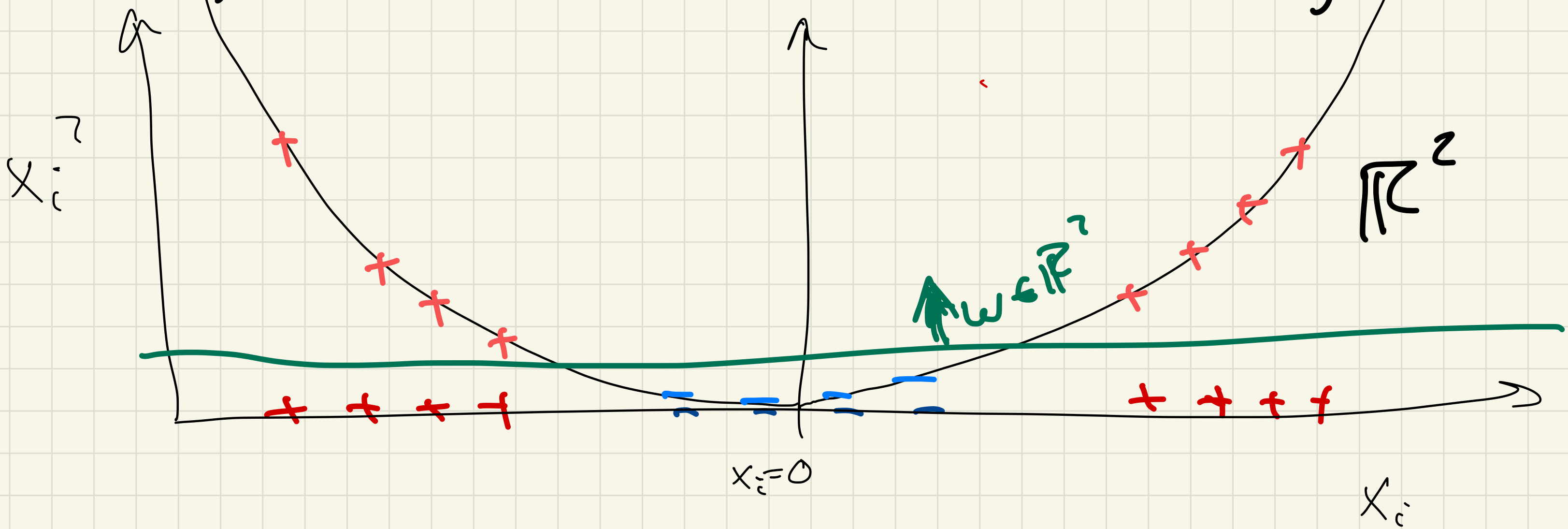
$\langle p,x\rangle_G = K(p,x) = \exp\left(-\|x-p\|^2/\sigma^2\right)$ Gaussian

$\langle p,x\rangle_L = K(p,x) = \exp\left(-\|x-p\|/\sigma\right)$ Laplace

RBF

$\langle p,x\rangle_P = K(p,x) = \left(\langle p,x\rangle + c\right)^r$ polynomial

Polynomial Classifier $x \in \mathbb{R}$ $y \in \{-1,+1\}$



$$x_i \longrightarrow g_i = (x_i, x_i^2)$$

Kernel Perceptron

change classifier : mistake counter

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

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

$\alpha_i = \#$ updates w/ x_i, y_i

mostly $\alpha_i = 0$

predictor function

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

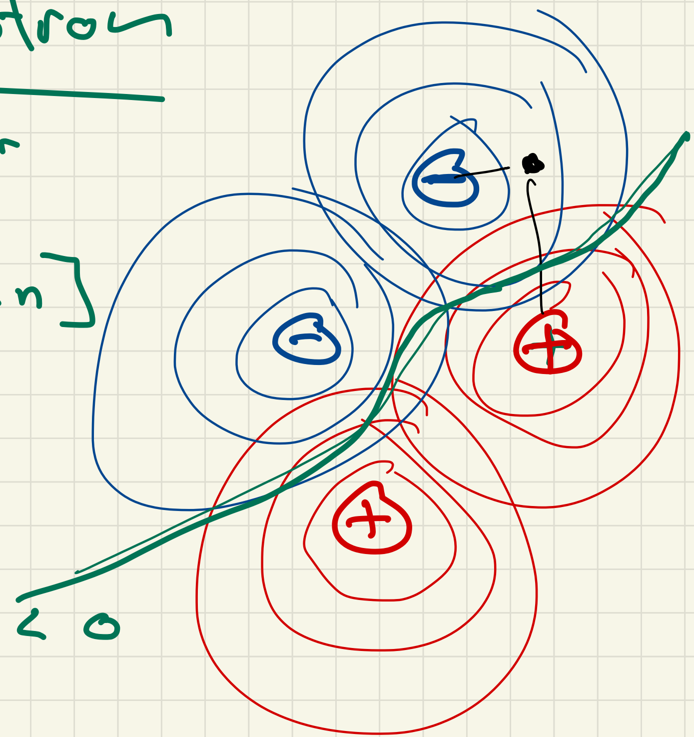
$$= \sum_{i=1}^{n} \alpha_i \, y_i \langle x_i , p \rangle$$

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

Kernel Perceptron

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

Choose $\alpha_i = 1 \quad i \in [1 \dots n]$



repeat

if some $x_i, y_i \in X, y$

$$\sum_{j=1}^{n} \alpha_j y_j \underbrace{\langle x_j, x_i \rangle}_{K(x_j, x_i)} < 0$$

then $\alpha_i = \alpha_i + 1$

Return $g(\cdot) = \sum_{j=1}^{n} \alpha_j y_j \underbrace{\langle x_j, \cdot \rangle}_{K(x_j, \cdot)}$

$p = (x, z)$ data point

Polynomial Expansion

$$g = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 z + \alpha_4 z^2 + \alpha_5 x z$$

$$\alpha \in \mathbb{R}^6$$

degree $r \Rightarrow \dim(g) = O(d^r)$

$$\langle x, w \rangle_r = \langle (1, x, x^2, z, z^2, xz), \alpha \rangle$$