FoDA L27

Classification (Non-Linear)

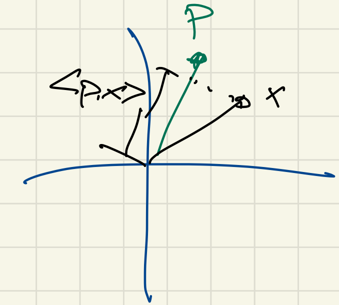
SVMs & Kernels

Dec 1, 2022

Linear Models

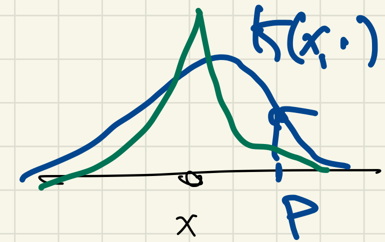
dot product

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



$$\langle p, x \rangle = \sum_{i=1}^{d} p_i \cdot x_i$$

similarity function



Kernel

- Gaussian Kernel $K(x,p) = \exp\left(-\frac{\|x-p\|^2}{\sigma^2}\right)$
- Laplace Kernel $K(x,p) = \exp\left(-\frac{\|x-p\|}{\sigma}\right)$
- Polynomial Kernel $K(x,p) = (\langle x,p \rangle + c)^r$

Perceptron Algo

$\mathrm{sign}(g_w(x) = \langle w, x\rangle)$ $\qquad T = \left(\frac{1}{\gamma^*}\right)^2$

$w = y_1 x_1$ $\qquad \alpha = (1, 0, 0, \dots 0)$

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

α_i = # times found mistake w/ (x_i, y_i)

repeat

find some (x_i, y_i) s.t. $\boxed{g_w(x_i) \neq y_i}$ mistake

$w \leftarrow w + y_i x_i$

$\alpha \leftarrow \alpha + (0 \dots 0, \overset{i}{1}, 0\, 0\, 0)$

new point $p \in \mathbb{R}^d$

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

$w \in \mathbb{R}^d$ $\qquad \alpha \in \mathbb{R}^n$

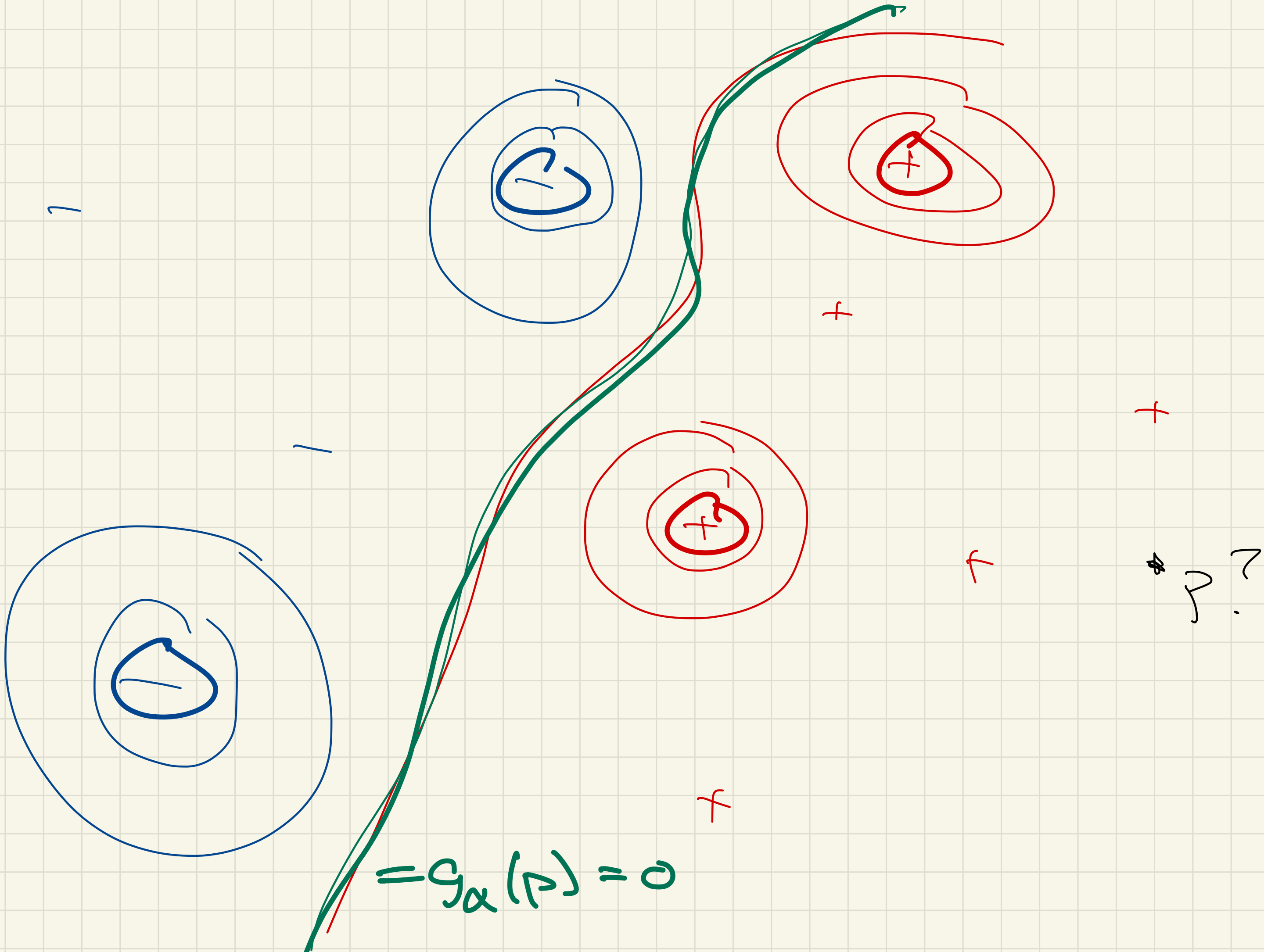
$\alpha = (\alpha_1, \dots \alpha_n)$

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

replace

$\boxed{K(x_i, p)}$

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


$= g_\alpha(p) = 0$
p ?

Polynomial Kernel

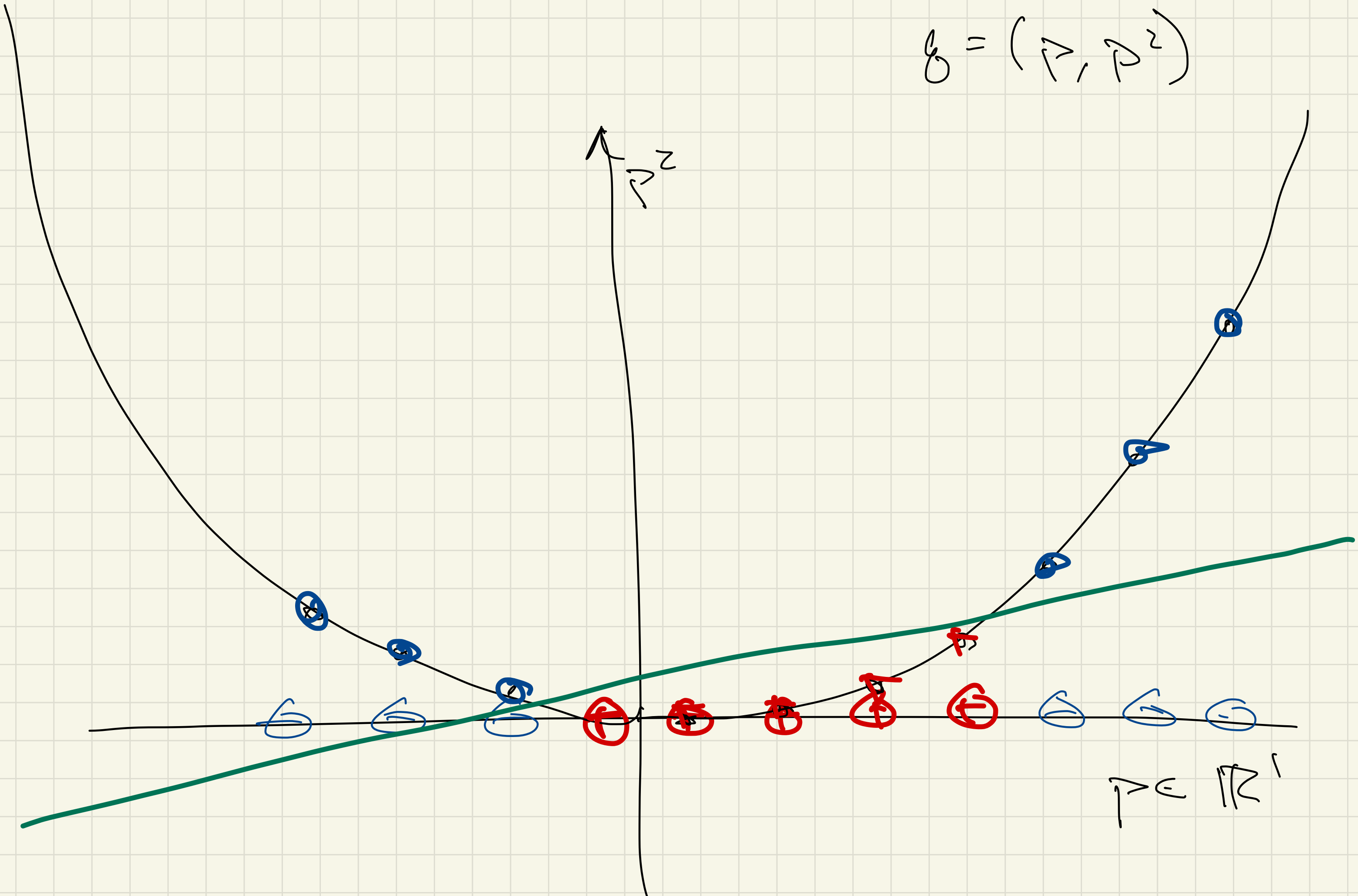
$$K(x,p) = (\langle x, p \rangle + c)^r$$

represent all polynomials of x, p of power at most r

$p \to$ polynomial expansion

$p \in \mathbb{R}^1 \quad v \in \mathbb{R}^{r+1} \quad v = (1, p, p^2, \dots, p^r)$

$$p = (p_1, p_2) \in \mathbb{R}^2 \longrightarrow q = (1, p_x, p_x^2, p_y, p_y^2, p_x \cdot p_y) \in \mathbb{R}^6$$



$\phi = (p, p^2)$
p^2
$p \in \mathbb{R}^1$

Polynomial Classification via Lifting

1. $\forall\, x_i \in X$ map to $q_i \in \mathbb{R}^{O(d^r)}$ ex. $(1, x_i, x_i^2, \ldots)$

2. Build Linear Classifier on (Q, y)
   $g : \mathbb{R}^{O(d^r)} \to \mathbb{R}$

3. Apply to new $p \in \mathbb{R}^d$
   3a. map $p \to \mathbb{R}^{O(r^d)} \to q_p$
   3b. $\text{sign}(g(q_p))$

Support Vector Machines (SVMs)

Goal $\quad g_\alpha(p) = \sum_{i=1}^{n} \alpha_i K(x_i, p)$

learn $\alpha \in \mathbb{R}^n$

how misclassified

$$z_j = y_j\, g_\alpha(x_j) = y_j \sum_{i=1}^{n} \alpha_i \underbrace{K(x_i, x_j)}_{n^2}$$

$$\min_{\alpha} \sum_{i=1}^{n} \ell(z_j)$$

$|S| = k$

subset $S \subset X$

$\leftarrow$ support vectors

$$g_\alpha(p) = \sum_{x_i \in S} \alpha_i K(x_i, p)$$

$K(x_i, x_j)$

$k \cdot n$