COL774
Machine Learning
Sep 28, 2021

Linear SVM
$$\begin{cases} \min_{w,b,\xi} \frac{1}{2}w^Tw + C\sum_{i=1}^{m}\xi_i \\ y^{(i)}(w^Tx^{(i)}+b) \geq 1-\xi_i \\ \xi_i \geq 0 \end{cases}$$

Last class.... SVMs:-

Decision boundary:- non-linear.

quadratic separator

$(c_1, c_2)$

+ve $\{(x_1-c_1)^2 + (x_2-c_2)^2 \leq r^2$

$1\{(x_1-c_1)^2 + (x_2-c_2)^2 - r^2 \leq 0\}$ : $\left.\begin{matrix} -ve \; \frac{1}{5} +ve \\ \geq 0 \; -ve \end{matrix}\right\}$ $(x_1-c_1)^2 + (x_2-c_2)^2$

$\{x_1^2 - 2x_1c_1 + c_1^2 + x_2^2 - 2x_2c_2 + c_2^2 \leq r^2$ +ve

separator:- $x_1^2 - 2x_1c_1 + c_1^2 + x_2^2 - 2x_2c_2 + c_2^2 = r^2$

Transform into a new space where:-

$x_1^2 \quad x_2^2 \quad x_1 \quad x_2$ ] Quadratic transformation

$z_1 \quad z_2 \quad z_3 \quad z_4$

$\Rightarrow$ Decision boundary is linear in

$z_1 \quad z_2 \quad z_3 \quad z_4$

$z_1 - 2c_1z_3 + z_2 - 2c_2z_4 + c_2^2 = r^2$

$x \equiv (x_1 \ldots x_n)$ ] non linear

Transform $\hookrightarrow \phi(x) \equiv [(\phi(x))_1, (\phi(x))_2 \ldots (\phi(x))_N)$

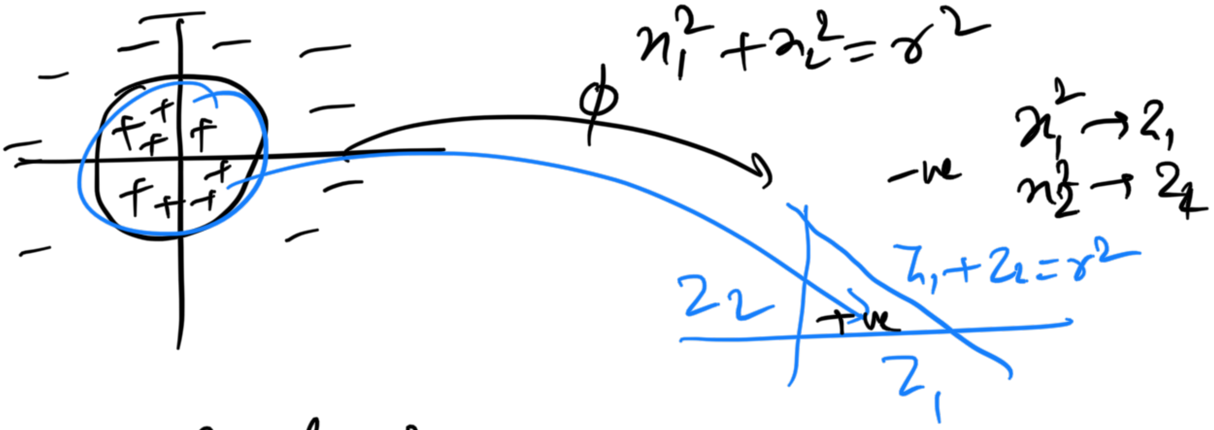
$N \gg n$

$N = \ldots$ — Linear separator in the new feature space

Illustration:-

$c_1 = 0 \quad c_2 = 0$

$x_1^2 + x_2^2 = r^2$




How to large?

$(x_1 \, -- \, x_n)$

$\downarrow \phi$ [polynomial transformation of degree $d$]

$((\phi(x))_1 \, -- \, (\phi(x))_N)$ → monomials of degree $\le d$

e.g. $x_1^{d-2} x_3 x_4$, $x_1^{d-3}$, $--$ $x_1 x_2^{d-5} x_3^4$ ] $x_1^d \cancel{x_2}$

Hint:- $\boxed{x_1} \, -- \, x_n$

$n$ | | | | | | |

$d$ bars

$\frac{(n+d-1)!}{(n-1)!\, d!}$

What is the size of this feature space?

$\left[ {}^{n+d-1}C_d \right]$ :- $O(n^d)$ :- exponential in $d$

${}^{n+d}C_d$

Example:-

$n=3$

$d=2$

$\phi(x) = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ x_1 x_3 \\ x_2 x_1 \end{bmatrix}$ $\phi(z) =$ ↓ similarly

$(x_1, x_2, x_3) \longrightarrow \begin{bmatrix} x_2x_2 \\ x_2x_3 \\ x_3x_1 \\ x_3x_2 \\ x_3x_3 \\ \sqrt{2c}\, x_3 \\ \sqrt{2c}\, x_2 \\ \sqrt{2c}\, x_1 \\ c \end{bmatrix}$ $\quad Z = (z_1, z_2, z_3) \quad \begin{bmatrix} \vdots \\ \sqrt{2c}\, z_3 \\ \sqrt{2c}\, z_2 \\ \sqrt{2c}\, z_1 \\ \end{bmatrix}$

$$\left[\underline{\phi(x)^T . \phi(z)}\right] = ? \qquad O(N)$$

Claim:- $\phi(x)^T\phi(z) = \underbrace{(x^Tz + c)^2}_{\text{inner product under quadratic transformation}}$

$$= (x^Tz)^2 + 2x^Tzc + c^2$$

$$= \left(\sum_{j=1}^{n} x_j z_j\right)^2 + (\sqrt{2c})^2 \sum_{j=1}^{n} x_j z_j + c^2$$

$$\left[\sum_j (x_j z_j) \sum_k (x_k z_k)\right]$$

$$\Downarrow$$

$$(x_j z_j x_k z_k) \equiv \underbrace{(x_j x_k)}\,\underbrace{(z_j z_k)}$$

$\Rightarrow$ To compute:-

$$\phi(x)^T\phi(z) = (x^Tz + c)^2$$

$$\Rightarrow O(n+2)$$

$\Rightarrow$ To get a degree $d$ polynomial transformation & compute the inner product $\underbrace{\phi(x)^T\phi(z)}$

$O(\cancel{N})$

Claim:- $\underline{\phi(x)^T\phi(z)} = \underline{(x^Tz+c)^d}$

Directly:- $O(N) \equiv O(n^d)$

using "conceptual" kernel rep:- $O(\underbrace{n+d}))$ ]

Kernel function: $K: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

if $K(x,z)$ can be computed efficiently $\Leftarrow$ [$K(x,z) = \underbrace{\phi(x)^T \cdot \phi(z)}$

[where $\phi$ is some feature transformation over $x$ ($z$).

SVM Dual:- Soft Margin Linear SVM

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} [x^{(i)T} x^{(j)}]$$

$\phi(x^{(i)})^T \cdot \phi(x^{(j)})$

$$0 \le \alpha_i \le C \qquad \sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

$\Rightarrow$ entire computation can be done in terms of $\phi(x^{(i)})^T \phi(x^{(j)})$ $\forall i, j$

$\Leftrightarrow K(x^{(i)}, x^{(j)})$

$$w = \sum_{i=1}^{m} \alpha_i y^{(i)} \underbrace{\phi(x^{(i)})}$$

$$b = -1/2 \left[ \min_{i: y^{(i)}=1} \underline{w^T \phi(x^{(i)})} + \max_{i: y^{(i)}=-1} \underline{w^T \phi(x^{(i)})} \right]$$

Prediction time:- $\underbrace{\quad\quad\quad}$ can be computed using inner product $\phi(x^{(i)}) \cdot \phi(x^{(j)})$

$x$:- $[w^T \cdot \underline{\phi(x)} + b]$

:- $\sum_{i=1}^{m} \alpha_i y^{(i)} \cdot \left[ \underbrace{(\phi(x^{(i)}))^T \cdot \phi(x)}_{\text{inner product}} \right]$

$\Rightarrow$ Never need to represent $w$ [$\phi$] explicitly.

Implicitly :- At prediction time, do compute using inner product

:- $w \equiv$ store $\underbrace{\alpha_i, x^{(i)}, y^{(i)}}_{\text{support vectors}}$ $\underline{\alpha_i \neq 0}$

$\Rightarrow$ $K(x^{(i)}, x^{(j)}) \equiv \phi(x^{(i)})^T \phi(x^{(j)})$

Efficiently we can have an efficient implementation of non-linear SVM.

kernel functi

Kernel function

① $K(x, z) = (x^T z + c)^d$ — $\phi(x)$:- degree $d$ polynomial

② $K(x, z) = e^{-\frac{\|x - z\|^2}{2\sigma^2}}$ → hyperparameter

[$\phi(x)$:- "Gaussian" Transformation

Gaussian kernel $\phi(x)^T \phi(z)$

RBF kernel

$x \in \mathbb{R}^n$

Mercer's:-

Let $K: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a "kernel" function

$$\underline{\{x^{(1)} \; \dots \; x^{(m)}\}}$$

$K^M$ :- $\mathbb{R}^{m \times m}$

$$K^M_{ij} = K(x^{(i)}, x^{(j)})$$

$$\begin{array}{cc} & j^{\text{th}} \\ i^{\text{th}} & \begin{bmatrix} & & \\ & & K(x^{(i)}, x^{(j)}) \\ & & \end{bmatrix} \end{array}$$

(A) If $\exists \; \phi: \mathbb{R}^n \to \mathbb{R}^N$ s.t

$$\underbrace{\phi(x^{(i)})^T \cdot \phi(x^{(j)})} = K(x^{(i)}, x^{(j)})$$

then (i) $K^M$ is symmetric

(ii) $K^M$ is +ve semi-definite

$$\forall z \; z^T K^M z \geq$$

(B) $\underline{K: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}}$ $\quad \underline{\forall \{x^{(i)}\}_{i=1}^m}$

$K^M$:- kernel matrix

if (i) $K^M$ is symmetric

② $K^{M}$ is +ve semi-definite

$\Rightarrow \exists \phi!$ $\quad K(x^{(i)}, x^{(j)}) = \phi(x^{(i)})^T \phi(x^{(j)})$

$\Downarrow$