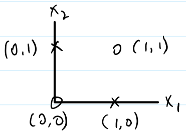
P36 Background

Data : Linear divisible — Allow mistakes — Nonlinear

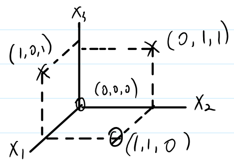
PLA — Pocket Algorithm — $\phi(x)$ + PLA

Hard-Margin SVM — Soft-Margin SVM — $\phi(x)$ + Hard-Margin SVM (kernal SVM)

Non linear - Transformation: $\underset{\text{input space}}{\mathcal{X}} \xrightarrow{\phi(x)} \underset{\text{feature space}}{\mathcal{Z}}$



设 $X_3 = (X_1 - X_2)^2$



XOR Problem : Linear indivisible
异或问题

Problems caused by non-linearity : ① High-dimensional transformation due to nonlinearity (Dimensional upgrading process $\phi(x)$)
② Duality donates dot product $\phi(x_i)^T\phi(x_j)$

②: Hard-Margin SVM

$$\begin{cases} \lambda = \arg\min\limits_{\lambda} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j (x_i^T x_j) - \sum_{i=1}^{N}\lambda_i \\ s.t.\ \lambda_i \geq 0,\ i = 1\cdots N \\ \sum_{i=1}^{N}\lambda_i y_i = 0 \end{cases}$$

$x_i^T x_j \rightarrow \phi(x_i)^T\phi(x_j)$

$\therefore$ Complete dimensional upgrading and dot product simultaneously by kernel fuction

$k(x,x') = \phi(x)^T\phi(x') = \langle\phi(x), \phi(x')\rangle$

P37 Positive-Definite kernel

Kernel: $k: \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$ $\forall x, x' \in \mathcal{X}$ ($\mathcal{X}$ indicates space) Then $k(x,x')$ is kernel function

positive-definite kernel ①: $k: \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$, $\forall x, x' \in \mathcal{X}$, $\exists k(x,x')$。If $\exists \phi: \mathcal{X} \mapsto \mathbb{R}^n$, $\phi \in \mathcal{H}$ (Hilbert space)
s.t. $k(x,x') = \langle\phi(x)^T\phi(x')\rangle$, than $k(x,x')$ is positive-definite kernel function

②: $k: \mathcal{X}\times\mathcal{X} \mapsto \mathbb{R}$, $\forall x, x' \in \mathcal{X}$, 有 $k(x,x')$。If $k(x,x')$ satisfies the following two properties :
① symmetry ② positive-definite Then $k(x,x')$ is called positive-definite function

① symmetry $\iff k(x,x') = k(x',x)$

② positive-definite $\iff$ Any n elements, $x_1, x_2 \cdots x_N \in \mathcal{X}$, the corresponding Gram matrix is positive semidefinite
i.e. $\forall y \in \mathcal{X}$, $y^T k y \geq 0$

\# Gram matrix : $k = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_N) \\ \vdots & & \\ k(x_N,x_1) & \cdots & k(x_N,x_N) \end{bmatrix}$

$\therefore k(x,x') = \langle\phi(x), \phi(x')\rangle \iff$ Gram matrix is positive semidefinite

\# Hilbert Space is $\underset{①}{\underline{\text{complete}}}$ $\underset{③}{\underline{\text{linear space}}}$ with infinite number of dimensions and $\underset{②}{\underline{\text{inner product}}}$
(希尔伯特空间)

① Closed under limit : $\{k_n\} \in \mathcal{H}$ $\lim\limits_{n\to\infty} k_n = k \in \mathcal{H}$

② $\begin{cases} \text{Symmetry} \\ \text{Positive-definite} \Rightarrow f, g \in \mathcal{H} \\ \text{Linear} \quad (f, g \text{ are functions}) \end{cases}$ $\begin{cases} \langle f, g\rangle = \langle g, f\rangle \\ \langle f, f\rangle \geq 0 \text{, equation holds} \iff f = 0 \\ \langle r_1 f_1 + r_2 f_2, g\rangle = r_1\langle f_1, g\rangle + r_2\langle f_2, g\rangle \end{cases}$

P38 Proof of Necessity

Known $k(x,x') = \langle\phi(x), \phi(x')\rangle$, proof Gram matrix is positive-definite, and $k(x,x')$ is symmetry

Symmetry : $k(x,x') = \langle\phi(x), \phi(x')\rangle$ $\quad k(x',x) = \langle\phi(x'), \phi(x)\rangle$

$\because$ Inner product has symmetry $\therefore \langle\phi(x), \phi(x')\rangle = \langle\phi(x'), \phi(x)\rangle$

$\therefore k(x,x') = k(x',x)$ $\therefore$ Satisfy symmetry

$A_{n\times n}$ is positive-definite $\begin{cases} ① \text{ Eigenvalue} \geq 0 \\ ② \forall \alpha \in \mathbb{R}^n \quad \alpha^T A \alpha \geq 0 \ ✓ \end{cases}$

Proof. positive-definite : $\forall \alpha \in \mathbb{R}^N$ $\quad \alpha^T k \alpha = \alpha^T \begin{bmatrix} k_{11} & \cdots & k_{1N} \\ \vdots & & \vdots \\ k_{N1} & \cdots & k_{NN} \end{bmatrix} \alpha$

原式 $= \alpha^T [\phi(x_1) \cdots \phi(x_N)]^T [\phi(x_1) \cdots \phi(x_N)] \alpha$
$= \|\alpha[\phi(x_1) \cdots \phi(x_N)]\|^2 \geq 0$
$= \|\sum_{i=1}^{N} \alpha_i \phi(x_i)\|^2$

P38.5 Widely Used Kernal Function

Common kernel Functions ① Linear kernal $k(x,x') = \langle x, x'\rangle$

② Polynomial (多项式) kernal $k(x,x') = (\langle x, x'\rangle + 1)^r$

③ Gaussian Kernal $k(x,x') = \exp\{\frac{-\|x-x'\|^2}{2\sigma^2}\}$

Distance in the Feature Space : $\|\phi(x) - \phi(x)\|^2 = \langle\phi(x), \phi(x)\rangle - 2\langle\phi(x), \phi(x')\rangle + \langle\phi(x'), \phi(x')\rangle$
$= k(x,x) - 2k(x,x') + k(x',x')$

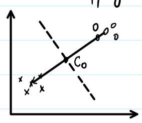
Angle in the Feature Space : $\cos\theta = \frac{\langle\phi(x), \phi(x')\rangle}{\|\phi(x)\|\,\|\phi(x')\|} = \frac{k(x,x')}{\sqrt{k(x,x)\,k(x',x')}}$

kernel function binary classification

Feature mapping Use each dimension of y as a coordinate

$\uparrow$ ↘ ○ ○ $\quad x: C_+ = \frac{1}{N_+}\sum_{i|y_i \in C_+} \phi(x_i)$ (c is the mean value of the vectors after dimension upgrading) kernel function binary classification

Feature mapping Use each dimension of y as a coordinate



$x: C_+ = \frac{1}{N_1} \sum_{\{i|y_i \in C_1\}} \phi(x_i)$ (c is the mean value of the vectors after dimension upgrading)

$o: C_- = \frac{1}{N_2} \sum_{\{j|y_j \in C_2\}} \phi(x_j)$

$\swarrow: w = C_+ - C_-$

$C_0: \frac{1}{2}(C_+ + C_-)$

$\because$ Take the vertical line at the midpoint C_0 as the division line $\because$ b is difficult to find, so only the angle between the sample and w is used for classification

$$\cos\theta = \frac{\langle \phi(x_0) - C_0, w \rangle}{\|\phi(x_0) - C_0\| \|w\|} \propto \langle \underset{1\times p}{\phi(x_0) - C_0}, \underset{p\times 1}{w^T} \rangle \quad \therefore \hat{y}_0 = \mathrm{sgn}(\langle \phi(x_0) - C_0, w^T \rangle)$$

① If $\phi(x)$ is known then the solution can be solved directly

② Only know kernal function: $\langle \phi(x_0) - C_0, w \rangle = [\phi(x_0) - C_0]^T w = \phi(x_0)^T \left[\frac{1}{N_1} \sum_{\{i|y_i \in C_1\}} \phi(x_i) - \frac{1}{N_2} \sum_{\{j|y_j \in C_2\}} \phi(x_j)\right] - \frac{1}{2}(C_+ + C_-)(C_+ - C_-)^T$

$$= \frac{1}{N_1} \sum_{\{i|y_i \in C_1\}} k(x_0, x_i) - \frac{1}{N_2} \sum_{\{j|y_j \in C_2\}} k(x, x_j) \underbrace{- \frac{1}{2N_1^2} \sum_{\{i|y_i \in C_1\}} \sum_{\{j|y_j \in C_1\}} k(x_i, x_j) + \frac{1}{N_1 N_2} \sum_{\{i|y_i \in C_1\}} \sum_{\{j|y_j \in C_2\}} k(x_i, x_j) - \frac{1}{2N_2^2} \sum_{\{i|y_i \in C_2\}} \sum_{\{j|y_j \in C_2\}} k(x_i, x_j)}_{b}$$