Dual of SVM problem

$$\min_{\{w,b\}} \frac{1}{2}\|w\|^2$$

$$y_i(w^T x_i + b) \geq 1$$

$\mathcal{L}(w, b, \lambda)$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i \left\{ y^{(i)} \left(w^T x^{(i)} + b \right) - 1 \right\}$$

$$\nabla_w \mathcal{L} = 0 \Rightarrow w - \sum_{i=1}^{N} \lambda_i y^{(i)} x^{(i)} = 0$$

$$\nabla_b \mathcal{L} = 0 \Rightarrow -\sum_{i=1}^{N} \lambda_i y^{(i)} = 0$$

$$\lambda_i \left\{ y_i \left(w^T x^{(i)} + b \right) - 1 \right\} = $$

$$\lambda_i \geq 0$$

Wolfe Dual

$$\max_{w,b,\lambda} \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i \left\{ y^{(i)} \left(w^T x^{(i)} + b \right) - 1 \right\}$$

$$w = \sum_{i=1}^{N} \lambda_i y^{(i)} x^{(i)}$$

$$\sum_{i=1}^{N} \lambda_i y^{(i)} = 0$$

Eliminate $\underline{\underline{w}}$, b

$$\max_{\lambda \geq 0} \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \left\| \sum_{i=1}^{N} \lambda_i y^{(i)} x^{(i)} \right\|^2$$

$$\sum_{i=1}^{N} \lambda_i y^{(i)} = 0$$

$$\max_{\lambda \geq 0} \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y^{(i)} x^{(i)T} x^{(j)}$$

$$\sum_{i=1}^{N} \lambda_i y^{(i)} = 0$$

$$h(x) = \text{sign}\left(\sum_{i=1}^{N} \lambda_i y^{(i)} \left(x^{(i)}\right)^T x + b\right)$$

Solve SVM in feature space

Instead of x use $\Phi(x)$

SVM Dual

$$\max_{\lambda \geq 0} \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y^{(i)} y^{(j)} \Phi(x^{(i)})^T \Phi(x^{(j)})$$

$$\sum_{i=1}^{N} \lambda_i y^{(i)} = 0$$

Classifier

$$h(x) = \text{sign}\left(\sum_{i=1}^{N} \lambda_i y^{(i)} \Phi(x^{(i)})^T \Phi(x) + b \right)$$

Example

$$K(x, \tilde{x}) = \Phi(x)^T \Phi(\tilde{x}) = (1 + x^T \tilde{x})^2$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\Phi(x) = \begin{bmatrix} 1 \\ \sqrt{2} x_1 \\ \sqrt{2} x_2 \\ \sqrt{2} x_1 x_2 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$

$$\Phi(x)^T \Phi(\tilde{x})$$

$$= (1 + x^T \tilde{x})^2$$

Kernel methods

for Pattern Analysis

Kernel functions

$K : \mathcal{X} \times \mathcal{X} \Rightarrow \mathbb{R}$ is called a Kernel function if

1. K is symmetric
2. K is positive semi-definite.

K is positive semi-definite:
For every $n \in \mathbb{N}$ (Set of natural numbers)
and for every $\mathcal{D}_n = \{x^{(i)} \in \mathcal{X} \mid i \in [n]\}$

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

$$K \succeq 0.$$

Example

$$K(x,z) = x^T z = K(z,x)$$

For any $x^{(1)} \dots x^{(n)} \in \mathbb{R}^d$

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

$$u^T K u = \sum_i \sum_j u_i u_j x^{(i)T} x^{(j)}$$

$$= \left(\sum_i u_i x^{(i)}\right)^T \left(\sum_j u_j x^{(j)}\right)$$

$$= \left\|\sum_i u_i x^{(i)}\right\|^2 \geq 0$$

Mercer's Theorem (Thm 3.13)

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a compact set. Let $K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a continuous symmetric function.

Then for every $f \in L_2(\mathcal{X})$

$$\int_{\mathcal{X} \times \mathcal{X}} K(x, z) f(x) f(z) \, dx \, dz \geq 0$$

is true iff there exists

$\phi_i \in L_2(\mathcal{X})$, $\quad \int \phi_i(x) \phi_j(x) \, dx = 0$

such that

$$K(x, z) = \sum_{i=1}^{\infty} \phi_i(x) \phi_i(z).$$

$$K(x, z) = \Phi(x)^T \Phi(z)$$

$$\Phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_N(x) \\ \vdots \end{bmatrix}$$

$K(x, z)$ be a valid Kernel

then $\tilde{K}(x, z) = \dfrac{K(x, z)}{\sqrt{K(x, x)}\sqrt{K(z, z)}}$

$$K(x,x) = \|\Phi(x)\|$$

$$\tilde{K}(x,z) = \frac{\Phi(x)^T \Phi(z)}{\|\Phi(x)\| \, \|\Phi(z)\|}$$

$$= \tilde{K}(z,x) \qquad i \in [N]$$

$$\tilde{K}_{ij} = \tilde{K}(x^{(i)}, x^{(j)})$$

$$u^T \tilde{K} u = \sum_i \sum_j u_i u_j \frac{\Phi(x^{(i)})^T \Phi(x^{(j)})}{\|\Phi(x^{(i)})\| \, \|\Phi(x^{(j)})\|}$$

$$= \left\| \sum_i u_i \frac{\Phi(x^{(i)})}{\|\Phi(x^{(i)})\|} \right\|^2 \geq 0$$

for $\forall u \quad u \in \mathbb{R}^N$

$\tilde{K}$ is called normalized Kernel

$$\tilde{K}(x,x) = 1$$

Let K_1, K_2 be two Kernel functions.

For any $\alpha, \beta \geq 0$

$K = \alpha K_1 + \beta K_2$ is a valid Kernel function

$K(x,z) = K_1(x,z)K_2(x,z)$

is a Kernel function

For any N.

$(K_1)_{ij} = K(x^{(i)}, x^{(j)})$

is a covariance matrix.

$A \in \mathbb{R}^N$

$A \sim N(0, K_1)$

$B \in \mathbb{R}^N$

$B \sim N(0, K_2)$

$C_i = A_i B_i$ $\quad E(C_i) = E(A_i)E(B_i)$

$= 0$

$$E(C_i C_j) = E(A_i B_i A_j B_j)$$

$$= E(A_i A_j) E(B_i B_j)$$

$$K_{ij} = (K_1)_{ij} (K_2)_{ij}$$

$K \in \mathbb{R}^{N \times N}$ is positive definite

$$K(x, z) = \sum_{i=0}^{m} a_i (x^T z)^i$$

is Kernel function if

$a_i \geq 0$ for all i

$K(x,z) = (1+x^T z)^m$ is a Kernel function

$$K(x,z) = \lim_{m\to\infty} \sum_{i=0}^{m} \frac{1}{i!} (x^T z)^i = e^{x^T z}$$

is a Kernel function

$$\tilde{K}(x,z) = \frac{e^{x^T z}}{e^{\frac{1}{2}x^T x}\, e^{\frac{1}{2}z^T z}}$$

$$= e^{-\frac{1}{2}\|x-z\|^2}$$

Gaussian (RBF) Kernel