Kernels & Kernelized Algorithms

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

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

polynomial $K(x,z) = (1 + x^T z)^d$

c

1. $K(x,z) = x^T z$
2. $K(x,z) = cK_1(x,z)$
3. $K(x,z) = K_1(x,z) + K_2(x,z)$
4. $K(x,z) = g(K(x,z))$
5. $K(x,z) = f(x)K_1(x,z)f(z)$
6. $K(x,z) = K_1(x,z)K_2(x,z)$
7. $K(x,z) = e^{K_1(x,z)}$
8. $K(x,z) = x^T A z$

Theorem RBF Kernel is well-defined

Proof

$K_1(x,z) = x^T z$ (rule 1)

$K_2(x,z) = \frac{2}{\sigma^2} K_1(x,z) = \frac{2}{\sigma^2} x^T z$ (rule 2)

$K_3(x,z) = e^{K_2(x,z)} = e^{\frac{2x^T z}{\sigma^2}}$ (rule 7)

$K_4(x,z) = e^{-\frac{x^T x}{\sigma^2}} K_3(x,z) e^{-\frac{z^T z}{\sigma^2}} = e^{-\frac{x^T x}{\sigma^2}} e^{\frac{2x^T z}{\sigma^2}} e^{-\frac{z^T z}{\sigma^2}}$ (rule 5)

$= e^{\frac{-x^T x + 2x^T z - z^T z}{\sigma^2}} = e^{\frac{-(x-z)^2}{\sigma^2}} = K_{RBF}(x,z)$ $f(x) = e^{\frac{-x^T x}{\sigma^2}}$

$\Rightarrow K$ is p.s.d.

1. All eigenvalues of K are ≥ 0
2. $\exists$ real matrix P s.t. $K = P^T P$
3. $\forall$ real vectors x, $x^T K x \geq 0$

Theorem

The following Kernel is well defined on any two sets $S_1, S_2 \subseteq \Omega$

$$K(S_1, S_2) = e^{|S_1 \cap S_2|}$$

elements of

Proof

List out all possible elements of Ω and arrange into a sorted list.

$X_s \in \{0,1\}^{|\Omega|}$

$K(S_1, S_2) = e^{x_{S_1}^T x_{S_2}}$ → # of elements shared between S_1 & S_2

(rules 1 & 7) □

$\Omega = \{1,2,3,4,5\}$

$S = \{2,5\}$ $x_s = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Kernel Machines

1. Prove that parameters lie in span of data (i.e. $w = \sum_{i=1}^n \alpha_i x_i$)
2. Rewrite learning algo. & classifier so all training & test x_i are only accessed via inner products $x_i^T x_j$
3. Define a Kernel function $K(x_i, x_j)$ & substitute for $x_i^T x_j$

Kernelized Linear Regression

$\min_w \sum_{i=1}^n (w^T x_i - y_i)^2 \Rightarrow w = (XX^T)^{-1} X^T y$

$[x_1 \cdots x_n]$ $[y_1 \cdots y_n]^T$

① $w = \sum_{i=1}^n \alpha_i x_i = X\alpha$ $[\alpha_1 \cdots \alpha_n]^T$, $K_{ij} = K(x_i, x_j)$

② Claim: Kernelized least squares has this solution: $\alpha = K^{-1} y$

proof

$X\alpha = w = (XX^T)^{-1} X y$ (multiply from left by $X^T XX^T$)

$(X^T X)(X^T X)\alpha = X^T(\cancel{XX^T (XX^T)^{-1}}) X y$ (sub. $K = X^T X$

$K^2 \alpha = K y$ (multiply from left by $(K^{-1})^2$)

$\alpha = K^{-1} y$ □

ridge regression

$\alpha = (K + \lambda I)^{-1} y$

$\lambda \|w\|_2^2$

$\lambda > 0$

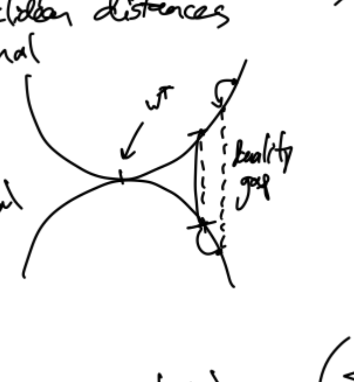
$h(z) = z^T w = z^T \underbrace{X\alpha}_{w} = \underbrace{K_*}_{z^T X} \underbrace{(K + \lambda I)^{-1} y}_{\alpha}$

Quiz: Let

$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$

How can you Kernelize KNN w/ Euclidean distances

$\boxed{\text{Kernel SVM}}$



primal

$\min_{w,b} w^T w + C\sum_{i=1}^n \xi_i$ (slack)

s.t. $\forall i$ $y_i(w^T x_i + b) \geq 1 - \xi_i$ (margin)

$\xi_i \geq 0$

dual

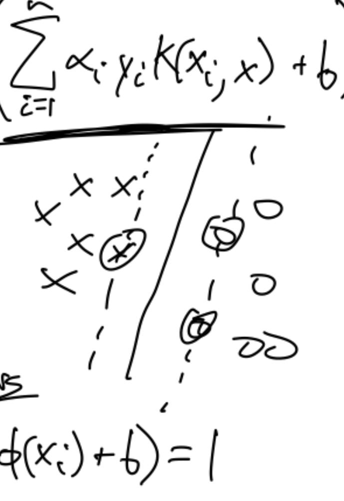
$\max_{\alpha_1, \ldots, \alpha_n} \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K_{ij} - \sum_{i=1}^n \alpha_i$ (K_{ij}: $x_i^T x_j$)

s.t. $0 \leq \alpha_i \leq C$

$\sum_{i=1}^n \alpha_i y_i = 0$

$h(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b\right)$

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



support vectors

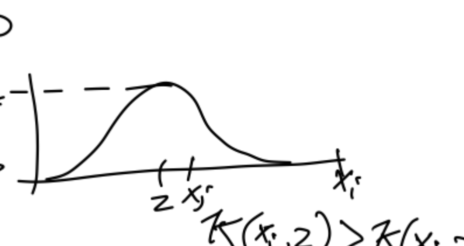
$y_i(w^T \phi(x_i) + b) = 1$

$y_i\left(\sum_{j=1}^n y_j \alpha_j K(x_j, x_i) + b\right) = 1$

$b = y_i - \sum_{j=1}^n y_j \alpha_j K(x_j, x_i)$

$\forall j \mid x_j$ is not a support vector

$\alpha_j = 0$



Kernel SVM — smart nearest neighbor algorithm

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

$K(x_j, z) > K(x_i, z)$

$K(z,z) = 1$

$y_i \in \{-1, 1\}$

$\in \{0,1\}$ → only if x_i is KNN of z

$\boxed{\text{NN}}$ $h(z) = \text{sign}\left(\sum_{i=1}^n y_i \delta^{nn}(x_i, z)\right)$

$\boxed{\text{SVM}}$ $h(z) = \text{sign}\left(\sum_{i=1}^n y_i \alpha_i K(x_i, z) + b\right)$