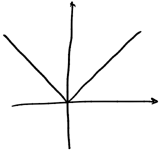
WA2 Yongsung Cho

1. (Subgradient) (5 pts) Consider the L_1 form function for $\mathbf{x} \in R^d$: $f(\mathbf{x}) = |\mathbf{x}|_1 = \sum_{i=1}^{d} |x_i|$. Show that $\mathbf{g} = [g_1, g_2, ..., g_d]^T$ is a subgradient of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$ if every $g_i \in [-1, 1]$. Hint: go back to the definition of subgradient: g is a subgradient of $f(x)$ at x_0 if $\forall x$, $f(x) \geq f(x_0) + g^T(x - x_0)$

$f(x) = |x|_1 = \sum_{i=1}^{d} |x_i|$

$g = [g_1, g_2 \cdots g_d]^T \Rightarrow$ subgradient of $f(x)$ at $x=0$



$f(x) = |x|$

g is a subgradient of $f(x)$ at x_0 if $\forall x$, $f(x) \geq f(x_0) + g^T(x - x_0)$

when $x_0 = 0$, $f(x) \geq f(0) + g^T(x - x_0)$

$f(x) \geq g^T x$

Because the slope of $f(x)$ at $x > 0$ is 1,

the slope of $f(x)$ at $x < 0$ is -1,

in $g \in [-1, 1]$ situation, $f(x) \geq g^T x$ is always True.

As a result, $g = [g_1, g_2 \cdots g_d]^T$ is a subgradient of $f(x)$ at $x = 0$

2. (Perceptron) (5 pts) Consider the following argument. We know that the number of steps for the perceptron algorithm to converge for linearly separable data is bounded by $(\frac{D}{\gamma})^2$. If we multiple the input $\mathbf{x}$ by a small constant α, which effectively reduces the bound on $|\mathbf{x}|$ to $D' = \alpha D$, we can reduce the upper bound to $(\alpha\frac{D}{\gamma})^2$. Is this argument correct? Why?

$\forall_i \; y_i u^T x_i \geq \alpha r \qquad w_k = w_{k-1} + y_k x_k$

① — $u^T w_k = u^T(w_{k-1} + y_k x_k) = u^T w_{k-1} + y_k u^T x_k \geq u^T w_{k-1} + \alpha r \geq k r$

② — $w_k^T w_k = (w_{k-1} + y_k x_k)^T (w_{k-1} + y_k x_k)$

$= w_{k-1}^T w_{k-1} + 2 y_k w_{k-1}^T x_k + x_k^T x_k \leq w_{k-1}^T w_{k-1} + \alpha^2 D^2$

$\leq \alpha^2 k D^2$

According to ①, ②, $\dfrac{u^T w_k}{|u||w_k|} \geq \dfrac{k r}{\alpha\sqrt{k}D}$

But, this cannot exceed 1

$\therefore \dfrac{kr}{\alpha\sqrt{k}D} \leq 1$

$kr \leq \alpha\sqrt{k}D$

$\cancel{k}^{\,k}\, r^2 \leq \alpha^2 \cancel{k} D^2$

$k \leq \left(\alpha \frac{D}{r}\right)^2$

3. (Cubic Kernels.) (10 pts) In class, we showed that the quadratic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^2$ was equivalent to mapping each $\mathbf{x} = (x_1, x_2) \in R^2$ into a higher dimensional space where

$$\Phi(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1).$$

Now consider the cubic kernel $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^3$. What is the corresponding Φ function?

$$(a^Tb+1)(a^Tb+1)(a^Tb+1)$$

$$\left((a^Tb)^2 + 2(a^Tb) + 1\right)(a^Tb+1)$$

$$(a^Tb)^3 + 2(a^Tb)^2 + a^Tb + (a^Tb)^2 + 2(a^Tb) + 1$$

$$= (a^Tb)^3 + 3(a^Tb)^2 + 3(a^Tb) + 1$$

$$\left[3\left(\sum_{i=1}^{d} a_i^2 b_i^2\right) + 6\sum_{i=1}^{d}\sum_{j=i+1}^{d} a_i a_j b_i b_j\right] + 3\sum_{i=1}^{d} a_i b_i$$

$$+1$$

$$\left(\sum_{i=1}^{d} a_i b_i\right)^3 = \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} a_i a_j a_k b_i b_j b_k$$

$$= a^3b^3 + 3a^2b + 3ab^2$$

$$(x_1^3 x_2^3, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{3}x_1, \sqrt{3}x_2), \quad \sqrt{6}x_1x_2$$

$$\sqrt{3}x_1^2, \sqrt{3}x_2^2, \quad \sqrt{3}x_1, \sqrt{3}x_2, 1$$

$$\therefore \Phi(x) = (x_1^3, x_2^3, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{6}x_1x_2,$$

$$\sqrt{3}x_1, \sqrt{3}x_2, 1)$$

4. (Kernel or not). In the following problems, suppose that K, K_1 and K_2 are kernels with feature maps ϕ, ϕ_1 and ϕ_2. For the following functions $K'(x, z)$, state if they are kernels or not. If they are kernels, write down the corresponding feature map, in terms of ϕ, ϕ_1 and ϕ_2 and c, c_1, c_2. If they are not kernels, prove that they are not.

- (5 pts) $K'(\mathbf{x}, \mathbf{z}) = cK(\mathbf{x}, \mathbf{z})$ for $c > 0$.
- (5 pts) $K'(\mathbf{x}, \mathbf{z}) = cK(\mathbf{x}, \mathbf{z})$ for $c < 0$.
- (5 pts) $K'(\mathbf{x}, \mathbf{z}) = c_1K_1(\mathbf{x}, \mathbf{z}) + c_2K_2(\mathbf{x}, \mathbf{z})$ for $c_1, c_2 > 0$.
- (5 pts) $K'(\mathbf{x}, \mathbf{z}) = K_1(\mathbf{x}, \mathbf{z})K_2(\mathbf{x}, \mathbf{z})$.

(1) $\phi' = \sqrt{c}\,\phi$ $K(X.z)$ has positive values.

if $c > 0$, $K'(X.z)$ also has positive values.

(2) On the other hand, if $c < 0$, $K'(X, z)$ has negative values. Because kernal matrix is positive semi-definite, every kernal matrix should be positive values.

As a result,

$K'(X.z)$ is not kernels in equation

$K'(X.z) = cK(X.z)$ for $c < 0$

(3) $K_1(X.Z)$ and $K_2(X.Z)$ are kernels. So,

$c_1K_1(XZ)$, $c_2K_2(X.Z)$ are also kernels

(Because $c_1, c_2 > 0$ according to Q1)

As a result, $c_1K_1(X, Z) + c_2K_2(X, Z)$ is also kernels.

feature map $\Rightarrow$ $\phi' = \sqrt{c_1}\,\phi_1 + \sqrt{c_2}\,\phi_2$

(4) if $K_1(X, z)$ has N_1 features

$K_2(X, z)$ has N_2 features,

ϕ will have $N_1 \times N_2$ features,

$$\phi_{ij} = \phi_{1i} \cdot \phi_{2j}$$