AI 534 WA-2

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

For a subgradient $f(x) \geq f(x_0) + g^T(x - x_0)$ $\quad x_0 = \bar{0}$ ($\bar{0}$ is 0 vector $= [0, 0 \ldots 0]^T$)

$$\Rightarrow f(x) \geq f(\bar{0}) + g^T(x - \bar{0})$$
$$\geq g^T(x).$$

$$\Rightarrow |x_1| + |x_2| + \cdots |x_d| \geq g^T(x)$$
$$\geq g_1 x_1 + g_2 x_2 + \cdots g_d x_d.$$

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

$$\Rightarrow \text{If } \quad |x_i| \geq g x_i \begin{cases} \text{if } x_i \geq 0 \Rightarrow x_i \geq g x_i \Rightarrow g \leq 1 \\ \\ \text{if } x_i < 0 \Rightarrow -x_i \geq g x_i \Rightarrow g \geq -1 \end{cases}$$

(multiplying by -1 switches inequality).

$$\Rightarrow |x_i| \geq g x_i \begin{cases} \text{if } x_i \geq 0 \text{ and } g \leq 1 \\ \text{or} \\ x_i < 0 \text{ and } g \geq -1 \end{cases}$$

$$\Rightarrow |x_i| \geq g x_i \quad \forall g \in [-1, 1]$$

$$\Rightarrow |x_1| + |x_2| + \cdots |x_d| \geq g_1 x_1 + g_2 x_2 + \cdots g_d x_d \quad \forall g_i \in [-1, 1]$$

$\Rightarrow g = [g_1, \ldots g_d]$ is a subgradient if $\forall i,\ g_i \in [-1, 1]$

2. This argument is incorrect and we show why:

Let $x_i^{\wedge}$ be the newly scaled data

$$\Rightarrow x_i^{\wedge} = \alpha x_i$$

As $|x_i| \leq D \Rightarrow \left|\frac{x_i^{\wedge}}{\alpha}\right| \leq D \Rightarrow |x_i^{\wedge}| \leq \alpha D$

$$\leq D^{\wedge}.$$

Also, $\exists u$, $y u^T x \geq \gamma$.

$$\Rightarrow y u^T \left(\frac{x_i^{\wedge}}{\alpha}\right) \geq \gamma \rightarrow y u^T x_i^{\wedge} \geq \alpha \gamma$$

$$\geq \gamma^{\wedge}.$$

Let us follow the proof again:

$\frac{u^T w_k^{\wedge}}{|u||w_k^{\wedge}|}$ converges to 1.

(i) $u^T w_k^{\wedge} = u^T(w_{k-1}^{\wedge} + y_k x_k^{\wedge}) \geq u^T w_{k-1}^{\wedge} + \gamma^{\wedge} \geq k\gamma^{\wedge}$

(ii) $w_k^{\wedge T} w_k^{\wedge} = (w_{k-1}^{\wedge} + y_k x_k^{\wedge})^T (w_{k-1}^{\wedge} + y_k x_k^{\wedge})$

$$= w_{k-1}^{\wedge T} w_{k-1}^{\wedge} + 2 w_{k-1}^{\wedge} y_k x_k^{\wedge} + x_k^{\wedge T} x_k^{\wedge}.$$

$$\leq w_{k-1}^{\wedge T} w_{k-1}^{\wedge} + x_k^{\wedge T} x_k^{\wedge} \leq w_{k-1}^{\wedge T} w_{k-1}^{\wedge} + D^{\wedge 2}$$

$$\leq k D^{\wedge 2}$$

$$\Rightarrow \frac{u^T w_k^{\wedge}}{|u||w_k^{\wedge}|} \geq \frac{k\gamma^{\wedge}}{\sqrt{k} D^{\wedge}}$$

$$\Rightarrow \sqrt{k}\left(\frac{\gamma^{\wedge}}{D^{\wedge}}\right) \leq 1 \Rightarrow k \leq \left(\frac{D^{\wedge}}{\gamma^{\wedge}}\right)^2 \leq \left(\frac{\alpha D}{\alpha \gamma}\right)^2$$

$$\leq \left(\frac{D}{\gamma}\right)^2.$$

3. Consider two vectors a and b.

$$K(a,b) = (a^Tb+1)^3 = (a^Tb)^3 + 3(a^Tb)^2 + 3(a^Tb) + 1$$

$$(a^Tb)^3 = \left(\sum_{i=1}^{d} a_i b_i\right)^3 = \sum_{i=1}^{d}(a_i b_i)^3 + 3\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}(a_i^2 a_j b_i^2 b_j) + 6\sum_{i=1}^{d}\sum_{j\neq i}^{d}\sum_{k\neq j\neq i}^{d} a_i a_j a_k b_i b_j b_k.$$

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

$$\Rightarrow K(a,b) = \sum_{i=1}^{d}(a_i b_i)^3 + 3\sum_{i=1}^{d}\sum_{\substack{j=1\\ j\neq i}}^{d}(a_i^2 a_j b_i^2 b_j) + 6\sum_{i=1}^{d}\sum_{j\neq i}^{d}\sum_{k\neq j\neq i}^{d} a_i a_j a_k b_i b_j b_k$$
$$+ 3\sum_{i=1}^{d} a_i^2 b_i^2 + 6\sum_{i=1}^{d}\sum_{j=i+1}^{d} a_i a_j b_i b_j + 3\sum_{i=1}^{d} a_i b_i + 1.$$

If vector $a = (x_1, x_2, x_3)^T$

$\Rightarrow k(a,b) = \langle \phi(a) . \phi(b) \rangle$, where

$\phi(a) = \cancel{(a_i^3}, \sqrt{3}$

$$\phi(a) = \big(x_1^3, x_2^3, x_3^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1^2x_3, \sqrt{3}x_2^2x_1, \sqrt{3}x_2^2x_3,$$
$$\sqrt{3}x_3^2x_1, \sqrt{3}x_3^2x_2, \sqrt{6}x_1x_2x_3, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{3}x_3^2,$$
$$\sqrt{6}x_1x_2, \sqrt{6}x_2x_3, \sqrt{6}x_1x_3, \sqrt{3}x_1, \sqrt{3}x_2, \sqrt{3}x_3, 1\big)$$

4. a. $k'(x,z) = ck(x,z) \quad c>0$

$k(x,z) = \phi^1(x)\phi^1(z) + \phi^2(x)\phi^2(z) + \dots \phi^N(x)\phi^N(z)$ where ϕ^i is the i^{th} feature in $\phi(x)$.

$ck(x,z) = c\phi^1(x)\phi^1(z) + c\phi^2(x)\phi^2(z) + \dots$

$= \sqrt{c}\,\phi^1(x)\sqrt{c}\,\phi^1(z) + \sqrt{c}\,\phi^2(x)\sqrt{c}\,\phi^2(z) + \dots \sqrt{c}\,\phi^N(x)\sqrt{c}\,\phi^N(z)$

$\langle\sqrt{c}\,\phi(x) . \sqrt{c}\,\phi(z)\rangle \Rightarrow \phi' = \sqrt{c}\,\phi$

b. From the above we see that $\phi' = \sqrt{c}\,\phi$. However if $c<0$ $\sqrt{c}$ does not exist $\Rightarrow k'(x,z) = k(x,z)\ \forall\ c<0$ is not a valid kernel. Also.

$$k'(x,z) = \begin{bmatrix} \phi^1(x) \\ \phi^2(x) \\ \vdots \\ \phi^N(x) \end{bmatrix}^T \underbrace{\begin{bmatrix} c & 0 & \cdots & 0 & 0 \\ 0 & c & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & & c & 0 \\ 0 & \cdots & & 0 & c \end{bmatrix}}_{M} \begin{bmatrix} \phi^1(z) \\ \phi^2(z) \\ \vdots \\ \phi^N(z) \end{bmatrix}$$

As $c<0$,

matrix M is negative definite. Thus this is an invalid kernel.

c. $k'(x,z) = c_1 k_1(x,z) + c_2 k_2(x,z)$. Let ϕ_1 have N_1 dimensions and ϕ_2; N_2 dims.

$= c_1\left(\phi_1^1(x)\phi_1^1(z) + \phi_1^2(x)\phi_1^2(z) + \dots \phi_1^{N_1}(x)\phi_1^{N_1}(z)\right)$

$+ c_2\left(\phi_2^1(x)\phi_2^1(z) + \phi_2^2(x)\phi_2^2(z) + \dots \phi_2^{N_2}(x)\phi_2^{N_2}(z)\right)$.

$= \sqrt{c_1}\,\phi_1^1(x)\sqrt{c_1}\,\phi_1^1(z) + \sqrt{c_1}\,\phi_1^2(x)\sqrt{c_1}\,\phi_1^2(z) + \dots \sqrt{c_1}\,\phi_1^{N_1}(x)\sqrt{c_1}\,\phi_1^{N_1}(z)$

$+ \sqrt{c_2}\,\phi_2^1(x)\sqrt{c_2}\,\phi_2^1(z) + \dots \sqrt{c_2}\,\phi_2^{N_2}(x)\sqrt{c_2}\,\phi_2^{N_2}(z)$

$\Rightarrow \phi = [\sqrt{c_1}\,\phi_1 : \sqrt{c_2}\,\phi_2]$, where $[:]$ is concatenation of the two feature sets.

d. $K'(x,z) = K_1(x,z)\, K_2(x,z)$

$$= \left[\phi_1^1(x)\phi_1^1(z) + \phi_1^2(x)\phi_1^2(z) + \dots \phi_1^{N_1}(x)\phi_1^{N_1}(z)\right].$$

$$\left[\phi_2^1(x)\phi_2^1(z) + \phi_2^2(x)\phi_2^2(z) + \dots \phi_2^{N_2}(x)\phi_2^{N_2}(z)\right]$$

$$= \phi_2^1(x)\phi_2^1(z) . K_1(x,z) + \phi_2^2(x)\phi_2^2(z) K_1(x,z) + \dots \phi_2^{N_2}(x)\phi_2^{N_2}(z) K_1(x,z)$$

$$= \phi_1^1(x)\phi_1^1(z)\phi_2^1(x)\phi_2^1(z) + \phi_1^1(x)\phi_1^1(z)\phi_2^2(x)\phi_2^2(z) + \dots$$

$$+ \phi_1^{N_1}(x)\phi_1^{N_1}(z)\phi_2^1(x)\phi_2^1(z) + \dots \phi_1^{N_1}(x)\phi_1^{N_1}(z)\phi_2^{N_2}(x)\phi_2^{N_2}(z)$$

$\Rightarrow$ ϕ is the feature set formed from element wise multiplication of the features in ϕ_1 and ϕ_2.