# Nick 322315318

## 1.1

Let $C = \{x_1, x_2, \dots, x_d\}$ be the set of points such that $x_i = (0, \dots, 0, 1, 0, \dots, 0)$ (d elements, $i^{th}$ element is 1). Let $y_1, y_2, \dots, y_d \in \{-1, 1\}$ be the labels of each $x_i \in C$. Define $w = (y_1, y_2, \dots, y_d)$ and we'll show that for $y_1, \dots, y_d$ and $w$:

$$\forall i \in \{1, 2, \dots, d\} : \mathrm{Sign}(w^T x_i) = y_i$$

Let $i \in \{1, 2, \dots, d\}$:

$$w^T x_i = (y_1, y_2, \dots, y_d) \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = y_i \Rightarrow w^T x_i = y_i \Rightarrow \mathrm{Sign}(w^T x_i) = \mathrm{Sign}(y_i) = y_i$$

meaning that we found a set of $d$ points that is shattered $\mathcal{H}$.

## 1.2

We'll show that $\mathrm{VCdim}(\mathcal{H}) < d + 1$. Let $C = \{x_1, \dots, x_{d+1}\}$ be a set of $d + 1$ different points. Assume for the sake of contradiction: there exists $w \in \mathbb{R}^d$ such that

$$\forall i \in \{1, \dots, d + 1\} : h(x_i) = y_i$$

using the following labeling:

$$y_i = \begin{cases} 1 & z_i \geq 0 \\ -1 & z_i < 0 \, \mathrm{or} \, i = d + 1 \end{cases}$$

using the hint: $x_{d+1} = \sum_{i=1}^{d} z_i x_i \quad / * w^T$

$$h(x_{d+1}) = \mathrm{Sign}(w^T x_{d+1}) = \mathrm{Sign}\left(w^T \sum_{i=1}^{d} z_i x_i\right) = \mathrm{Sign}\left(\sum_{i=1}^{d} z_i w^T x_i\right) = \sum_{i=1}^{d} \mathrm{Sign}(z_i)\mathrm{Sign}(w^T x_i) =$$

$$= \sum_{i=1}^{d} \mathrm{Sign}(z_i) y_i \geq 0$$

meaning that $h(x_{d+1}) = 1$ which contradicts the fact that $h(x_{d+1}) = y_{d+1} = -1$. Therefore, our assumption that there exists $w \in \mathbb{R}^d$ such that $\forall i \in \{1, \dots, d{+}1\}$ : $h(x_i) = y_i$ is wrong. Therefore $\mathcal{H}$ does not shatter $C$, i.e. $\mathrm{VCdim}(\mathcal{H}) < d + 1$.

## 2

$G(u,v) = K(u,v) \cdot K'(u,v) = \sum_{i=1}^{n_1} \varphi_i(u)\varphi_i(v) \sum_{j=1}^{n_2} \varphi'_j(u)\varphi'_j(v) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} (\varphi_i(u)\varphi_i(v))(\varphi'_j(u)\varphi'_j(v)) =$

$= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \varphi_i(u)\varphi'_j(u)\varphi_i(v)\varphi'_j(v)$

Let's define $\psi : \mathcal{X} \to \mathbb{R}^{n_1 \cdot n_2}$ such that

$\psi(u) = (\varphi_1(u)\varphi'_1(u), \dots, \varphi_1(u)\varphi'_{n_2}(u), \dots, \varphi_{n_1}(u)\varphi'_1(u), \dots, \varphi_{n_1}(u)\varphi'_{n_2}(u))$

$(\mathrm{n}_1 \cdot n_2 components)$

Thus

$$\langle \psi(u), \psi(v) \rangle = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} (\varphi_i(u)\varphi'_j(u))(\varphi_i(v)\varphi'_j(v)) = G(u,v)$$

Therefore, $G(u,v)$ is a valid kernel.

## 3.1

. Objective: $k(a,b) = e^{-\gamma(a-b)^2}$

1. $k_1(a,b) = a^T b$
2. $k_2(a,b) = 2\gamma k_1(a,b) = 2\gamma a^T b$
3. $k_3(a,b) = e^{k_2(a,b)} = e^{2\gamma a^T b}$
4. $k_u(a,b) = e^{-\gamma a^T a} k_3(a,b) \cdot e^{-\gamma b^T b} = e^{-\gamma a^T a} \cdot e^{2\gamma a^T b} \cdot e^{-\gamma b^T b} = e^{-\gamma(a^T a - 2a^T b + b^T b)} = e^{-\gamma(a-b)^2}$

Therefore,

$$e^{-\gamma(a-b)^2} \text{ is a valid kernel.}$$

$$K(a,b) = e^{-\gamma(a-b)^2} = e^{-\gamma a^2} \cdot e^{2\gamma ab} \cdot e^{-\gamma b^2} = e^{-\gamma a^2} \cdot e^{-\gamma b^2} \sum_{j=0}^{\infty} \frac{(2\gamma ab)^j}{j!} = e^{-\gamma a^2} \cdot e^{-\gamma b^2} \sum_{j=0}^{\infty} \frac{(\sqrt{2\gamma}a\sqrt{2\gamma}b)^j}{j!} =$$

$$= e^{-\gamma a^2} \cdot e^{-\gamma b^2} \sum_{j=0}^{\infty} \frac{(\sqrt{2\gamma}a)^j}{j!} \cdot \sum_{j=0}^{\infty} \frac{(\sqrt{2\gamma}b)^j}{j!} =$$

$$= e^{-\gamma a^2} \sum_{j=0}^{\infty} \frac{(\sqrt{2\gamma}a)^j}{j!} \cdot e^{-\gamma b^2} \sum_{j=0}^{\infty} \frac{(\sqrt{2\gamma}b)^j}{j!}$$

$$\varphi_j(a) = e^{-\gamma a^2} \frac{(\sqrt{2\gamma}a)^j}{j!}, \quad \varphi_j(b) = e^{-\gamma b^2} \frac{(\sqrt{2\gamma}b)^j}{j!} \quad \forall j \in \mathbb{N}$$

Therefore,

$$\langle \phi(a), \phi(b) \rangle = K(a,b)$$

## 3.2

Since the given dataset is very large, it would be possible and better to optimize the dual problem with the kernel we found since it saves us a lot of the computational time. The feature mapping we found is infinite meaning that unless we're able to use Euler's method to approximate the calculation is not possible in reasonable time.

## 4

$$f(x) = -x, \quad g(x) = |x|, \quad h(x) = f(g(x)) = f(|x|) = -|x|$$

$$f(x) \text{ is convex}: \quad f(\alpha u + (1-\alpha)v) = -\alpha u - (1-\alpha)v = \alpha f(u) + (1-\alpha)f(v)$$

$$g(x) \text{ is also convex}: \quad g(\alpha u + (1-\alpha)v) = |\alpha u + (1-\alpha)v| \leq \alpha|u| + (1-\alpha)|v| =$$
$$= \alpha g(u) + (1-\alpha)g(v) \quad \forall \alpha \in [0,1]$$

$$\text{Choose } u = 1, v = -1, \alpha = \frac{1}{2} \implies h(\alpha u + (1-\alpha)v) = h\left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (-1)\right) = h(0) = |0| = 0 > -1 =$$
$$= \alpha h(1) + (1-\alpha)h(-1)$$

therefore h(x) isn't convex

## 5.1)

$$f(z) \text{ is convex} \implies f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha)f(v) \quad / * \alpha$$

$$\alpha f(u + (1-\alpha)v) \leq \alpha(\alpha f(u) + (1-\alpha)f(v)) \implies \alpha f(z) \text{ is convex}$$

## 5.2)

I. 0 is convex w.r.t. $w : f(\alpha \cdot 0 + (1-\alpha)0) = 0 \leq \alpha \cdot 0 + (1-\alpha)0$

II. $1 - y_i w^T x_i$ is convex w.r.t. $w: \quad \nabla^2(1 - y_i w^T x_i) = \frac{\partial}{\partial w_i \partial w_j}\left(1 - y_i \mathbf{w}^T x_i\right) = 0 \succeq 0 \implies \text{PSD}$

$$\text{I} + \text{II} + \text{lemma 1} \implies \max\{0, 1 - y_i w^T x_i\} \text{ is convex w.r.t. } w$$

## 5.3)

$$\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\} \text{ is convex from } 5.2 + \text{lemma } 2$$

$$\frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\} \text{ is convex from } 5.1$$

$$\lambda\|w\|_2^2 \text{ is convex from Tirgul + 5.1}$$

$$\frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\} + \lambda\|w\|_2^2 \text{ is convex from lemma 2}$$

therefore the Soft-SVM problem is convex.