(Q1)

$\mathcal{H} = \{x \to sign(w^T x) : w \in \mathbb{R}^d\}$ (Homogeneous linear classifiers) $(b=0)$

1.1 Prove that $VC(\mathcal{H}) \geq d$

• עמ״נ להוכיח ש-$VC(\mathcal{H}) \geq d$ נראה כי קיימת קבו׳ בת d דגימות כך ש-$\mathcal{H}$ מנתצת אותה.

• נבחר קבו׳ דגימות $A = \{e_1, \dots, e_d\} \subset X$. עפ״י הגדרה, $\mathcal{H}$ מנתצת את A

אמ״מ $\forall y_1, \dots, y_d \in Y : \exists h \in \mathcal{H} : \forall x_i \in A : h(x_i) = y_i$

• בה״כ נסמן: $y_i \in Y$ כלשהו של $x_i \in A$. נראה: $\forall i \in d : \exists h : h(x_i) = y_i$

• נגדיר: $w = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_d \end{pmatrix} \mid h = sign(w^T x)$

• $h \in \mathcal{H}$

• $\forall x_i \in A : h(x_i) = sign(w^T x_i) = sign\left[\begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} \cdot e_i\right] = sign(y_i) = y_i$

$\Leftarrow VC(\mathcal{H}) \geq d$

מש״ל

1.2 Prove that $VC\dim(\mathcal{H}) = d$ by proving $VC\dim(\mathcal{H}) < d+1$

• תהא קבו׳ כלשהי $B = \{x_1, \dots, x_{d+1}\} \subset X$.

• יהא מסווג כלשהו $h \in \mathcal{H}$.

• נראה כי קיים תיוג $y_i \in Y \mid \{y_1, \dots, y_{d+1}\}$ ש-h לא יכול להסכים איתו.

• $\dim(B) = d+1$, $\dim(x_i) = d$. מתכונות אלגברה לינארית הקבו׳ אינה בת״ל

ולכן קיים לפחות וקטור אחד שהוא קו״ל של שאר הוקטורים בקבוצה.

בה״כ נסמנו x_{d+1} ולכן מתקיים $x_{d+1} = \sum_{i=1}^{d} \alpha_i x_i$ כך ש-$\exists i : \alpha_i \neq 0$

* נתבונן בלייבלינג הבא:

$$\begin{cases} y_{d+1} = -1 \\ \forall i \in [1,d] \mid \alpha_i \neq 0 : y_i = \text{Sign}(\alpha_i) \end{cases}$$

* נניח בשלילה ש-h מסכים עם לייבלינג זה. כלומר: $\forall i \in [1,d+1] : \text{Sign}(w^T X_i) = \text{Sign}(\alpha_i)$

מהנחות אלו מתקיים:

$$\Downarrow$$

$$\forall i \in [1,d] \mid \alpha_i \neq 0 : \alpha_i w^T X_i > 0$$

$$\Rightarrow h(X_{d+1}) = \text{Sign}\left(w^T \cdot \left[\sum_{i=1}^{d} \alpha_i X_i\right]\right) = \text{Sign}\left(\underbrace{\sum_{i=1}^{d} \alpha_i w^T X_i}_{\text{סכום של ערכים חיוביים וערכי 0}}\right) = 1$$

קיבלנו ש $h(X_{d+1}) = 1$ בסתירה לכך ש- $y = -1$ לכן $\mathcal{H}$ לא מנתצת את B

ומתקיים $VCdim(\mathcal{H}) < d+1$

מש"ל

2 • $\phi : X \to \mathbb{R}^{n_1}$, $\phi' : X \to \mathbb{R}^{n_2}$, $n_1, n_2 \in \mathbb{N}$.

• $K, K' : (X, X) \to \mathbb{R}$, two valid kernels:

$$* \quad K(u,v) = \langle \phi(u), \phi(v) \rangle = \sum_{i=1}^{n_1} \phi_i(u)\phi_i(v)$$

$$* \quad K'(u,v) = \langle \phi'(u), \phi'(v) \rangle = \sum_{j=1}^{n_2} \phi'_j(u)\phi'_j(v)$$

Prove: $G(u,v) \triangleq K(u,v)\cdot K'(u,v)$ is a valid kernel (find $\varphi : X \to \mathbb{R}^{n_3}$ s.t: $G(u,v) = \langle \varphi(u), \varphi(v) \rangle$)

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

• We'll define $\varphi : X \to \mathbb{R}^{n_3}$, $n_3 = n_1 \cdot n_2$ s.t:

$$\varphi(u) = \begin{pmatrix} \phi(u)_1 \cdot \begin{pmatrix} \phi'(u)_1 \\ \vdots \\ \phi'(u)_{n_2} \end{pmatrix} \\ \vdots \\ \phi(u)_{n_1} \begin{pmatrix} \phi'(u)_1 \\ \vdots \\ \phi'(u)_{n_2} \end{pmatrix} \end{pmatrix} = \left.\begin{pmatrix} \left.\begin{matrix} \phi(u)_1 \cdot \phi'(u)_1 \\ \vdots \\ \phi(u)_1 \cdot \phi'(u)_{n_2} \end{matrix}\right\} n_2 \\ \vdots \\ \left.\begin{matrix} \phi(u)_{n_1} \cdot \phi'(u)_1 \\ \vdots \\ \phi(u)_{n_1} \cdot \phi'(u)_{n_2} \end{matrix}\right\} n_2 \end{pmatrix}\right\} n_3 = n_2 \cdot n_1$$

מימד הוקטור

$$\Rightarrow \langle \varphi(u), \varphi(v) \rangle = \sum_{m=1}^{n_3} \varphi_m(u)\cdot\varphi_m(v) = \begin{pmatrix} \phi(u)_1 \cdot \phi'(u)_1 \\ \vdots \\ \phi(u)_1 \cdot \phi'(u)_{n_2} \\ \vdots \\ \phi(u)_{n_1} \cdot \phi'(u)_1 \\ \vdots \\ \phi(u)_{n_1} \cdot \phi'(u)_{n_2} \end{pmatrix}^T \cdot \begin{pmatrix} \phi(v)_1 \cdot \phi'(v)_1 \\ \vdots \\ \phi(v)_1 \cdot \phi'(v)_{n_2} \\ \vdots \\ \phi(v)_{n_1} \cdot \phi'(v)_1 \\ \vdots \\ \phi(v)_{n_1} \cdot \phi'(v)_{n_2} \end{pmatrix}$$

$$= \phi(u)_1 \phi'(u)_1 \cdot \phi(v)_1 \cdot \phi'(v)_1 + \dots + \phi(u)_1 \phi'(u)_{n_2} \cdot \phi(v)_1 \phi'(v)_{n_2}$$
$$+ \dots + \phi(u)_{n_1} \phi'(u)_1 \cdot \phi(v)_{n_1} \cdot \phi'(v)_1 + \dots + \phi(u)_{n_1} \phi'(u)_{n_2} \cdot \phi(v)_{n_1} \phi'(v)_{n_2}$$

$$= \phi(u)_1 \phi(v)_1 \cdot \sum_{j=1}^{n_2} \phi'(u)_j\, \phi'(v)_j + \dots + \phi(u)_{n_1} \phi(v)_{n_1} \sum_{j=1}^{n_2} \phi'(u)_j\, \phi'(v)_j$$

$$= \sum_{i=1}^{n_1} \phi(u)_i \, \phi(v)_i \cdot \sum_{j=1}^{n_2} \phi'(u)_j \, \phi'(v)_j$$

$\Rightarrow G(u,v) = \langle \psi(u), \psi(v) \rangle \Rightarrow$ $G(u,v)$ is a valid kernel

dèn

3) For $\gamma > 0$: $K: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $K(a,b) = \exp(-\gamma(a-b)^2)$

3.1) Provide $\phi: \mathbb{R} \to \mathbb{R}^P$, $P \in \mathbb{N} \cup \{\infty\}$ and prove that k is a valid kernel.

$$K(a,b) = \exp(-\gamma(a-b)^2) = \exp(-\gamma(a^2 - 2ab + b^2)) = \exp(-\gamma a^2 + 2ab\gamma - \gamma b^2)$$

$$= \exp(2ab\gamma) \cdot \exp(-\gamma a^2) \cdot \exp(\gamma b^2) \qquad \bullet\, e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$= \sum_{n=0}^{\infty} \frac{(2\gamma ab)^n}{n!} \cdot \exp(-\gamma a^2) \cdot \exp(\gamma b^2)$$

$$= \sum_{n=0}^{\infty} \frac{2^n \gamma^n}{n!} a^n \cdot \exp(-\gamma a^2) \cdot b^n \exp(\gamma b^2)$$

$$\Rightarrow \phi: \mathbb{R} \to \mathbb{R}^\infty: \quad \underbrace{\phi(X)_n}_{n\text{'th entry}} = \sqrt{\frac{2^n \gamma^n}{n!}}\, X^n \exp(-\gamma X^2)$$

3.2) Would it be better to optimize the primal problem with feature mapping or the dual problem with the kernel?

Comparing the time/Actions <u>complexity</u>:

- Dual problem with the kernels:

Calculating $\exp(-\gamma(a-b)^2)$ is $O(1)$ for $\dim(a) = \dim(b) = 1$

- Primal problem with feature mappings:

$\dim(\phi(a)) = \dim(\phi(b)) = P = \infty \Rightarrow$ Impractical to directly calculate/optimize. (Depending on implementation, also memory consuming).

$\Rightarrow$ It is better to optimize with the RBF kernel dual problem where calculations are made on the inputs low dimension.

4) Refute: let $f,g:\mathbb{R}\to\mathbb{R}$ <u>convex</u> func's $\Rightarrow h \triangleq f\circ g$ is convex.

$f(x) = -x$

- $f'(x) = -1 \Rightarrow f''(x) = 0 \Rightarrow f''(x) \geq 0 \Rightarrow$ $f(x)$ is convex

$g(x) = x^2$

- $g'(x) = 2x \Rightarrow g''(x) = 2 \geq 0 \Rightarrow g(x)$ is convex

$h(x) = f(g(x)) = f(x^2) = -x^2$

- $h'(x) = -2x \Rightarrow h''(x) = -2 < 0 \Rightarrow h(x)$ is <u>not</u> convex.

[Convexity checked by property seen it the Tirgul convex $\Leftrightarrow D^2 f \geq 0$]

5) Prove that the following Soft-SVM problem is convex:

$$\underset{w \in \mathbb{R}^d}{\arg\min} \ \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1-y_i w^T x_i\} + \lambda \|w\|_2^2$$

- Let $f, g: C \to \mathbb{R}$ be two convex func's over a convex set C.
- Lemma 1: $q(z) \triangleq \max\{f(z), g(z)\}$ is convex w.r.t z.
- Lemma 2: The sum of any number of convex funcs is convex.

5.1) Prove (by definition): Given a const $\alpha \in \mathbb{R}_{\geq 0}$, the function $\alpha f(z)$ is convex w.r.t z.

- Definition: A function $f: C \to \mathbb{R}$ is a convex function if $\forall x_1, x_2 \in C, \forall t \in [0,1]: \ t f(x_1) + (1-t) f(x_2) \geq f(t x_1 + (1-t) x_2)$

$$\forall z_1, z_2 \in C, \forall t \in [0,1]: \ \overbrace{t f(z_1) + (1-t) f(z_2) \geq f(t z_1 + (1-t) z_2)}^{\text{Given that } f(z) \text{ is convex}} \ / \cdot \alpha \in \mathbb{R}_{\geq 0}$$

$$t \alpha f(z_1) + (1-t) \alpha f(z_2) \geq \alpha f(t z_1 + (1-t) z_2)$$

$\Rightarrow \alpha f(z)$ is convex

5.2) Conclude that $\max\{0, 1-y_i w^T x_i\}$ is convex w.r.t w.

- $g(w) = 1 - y_i w^T x_i$ is linear $\rightarrow g(w)$ is convex W.R.T w.
- $f(w) = 0$ is linear $\Rightarrow f(w)$ is convex W.R.T w.
- by lemma 1: for convex functions $f(w), g(w): C \to \mathbb{R}$, the function $g(w) = \max(f(w), g(w))$ is also convex W.R.T w.

5.3) conclude that the Soft SVM Problem is convex W.R.T w.

$$\underset{w \in \mathbb{R}^d}{\arg\min} \ \frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\} + \lambda \|w\|_2^2$$

- $h(w) = \|w\|_2^2$ is convex $\left(\frac{\partial^2 h(w)}{\partial w_i w_j} = \begin{cases} 2 & i=j \\ 0 & i \neq j \end{cases} \geq 0 \right)$
- $\lambda h(w)$ is convex (based on 5.1)
- $\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\}$ is convex based on (5.2) and (lemma 2).
- $\frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\}$ is convex based on (5.1)
- $\frac{1}{m}\sum_{i=1}^{m} \max\{0, 1 - y_i w^T x_i\} + \lambda \|w\|_2^2$ is convex based on (Lemma 2).

$\Rightarrow$ the Soft SVM Problem is convex W.R.T w.