HW3

1. In Tutorial 05, we showed that the VC-dimension of homogeneous linear classifiers is $\geq d$.

   Now we will show similarly that the VC-dimension of <u>non</u>homogeneous linear classifiers is $\geq d+1$.

   Define $\mathcal{H}^d = \{\boldsymbol{x} \mapsto \operatorname{sign}(\boldsymbol{w}^\top \boldsymbol{x} + b): \ \boldsymbol{w} \in \mathbb{R}^d, b \in \mathbb{R}\}$.

   Prove <u>rigorously</u> that $\mathrm{VCdim}(\mathcal{H}^d) \geq d+1$.

   <u>Hint</u>: As in the tutorial, you should show a <u>specific</u> set $x_1, \dots, x_{d+1} \in \mathbb{R}^d$ and prove that it holds:

   $$\forall y_1, \dots, y_{d+1} \in \{-1,+1\}: \exists \boldsymbol{w} \in \mathbb{R}^d, b \in \mathbb{R}: \forall i \in [d+1]: \operatorname{sign}(\boldsymbol{w}^\top \boldsymbol{x}_i + b) = y_i.$$

קיימת קבוצה C הבאה : $C = \{e_1, e_2, \dots \dots, e_d, 0\},\ e_i \in \mathbb{R}^d$

כאשר e_i הוא וקטור הבסיס הסטנדרטי, כלומר : $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots \dots, e_d = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$

ו-0 הוא וקטור האפס מימד d.

מתקיים ש- $|C| = d+1$.

יהי תיוג $y_1, y_2, \dots., y_{d+1} \in \{-1,1\}$.

נראה שקיים $b \in \mathbb{R}$ ו $w \in \mathbb{R}^d -$ כך ש$\forall i \in [d+1]$ מתקיים $sign(w^T x + b) = y_i$.

נגדיר את w להיות : $w = \begin{pmatrix} y_1 - b \\ y_2 - b \\ y_3 - b \\ \vdots \\ y_{d-1} - b \\ y_d - b \end{pmatrix}$ כאשר b : $b = y_{d+1}$ ונקבל:

$$\forall i \in [d], \qquad w^T x_i + b = w^T e_i + b = y_i - b + b = y_i$$

$$for\ i = d+1\,, w^T x_{d+1} + b = \ w^T 0_{d \times 1} + b = 0 + b = y_{d+1}$$

לכן אנחנו מקבלים : $\forall i \in [d+1],\ sign(w^T x + b) = sign(y_i) = y_i$

<u>תשובה אפשרית נוספת :</u>

לאותו C שהראנו, נגדיר את w : $w = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{d-1} \\ y_d \end{pmatrix}$, כלומר $w_i = y_i$.

אנחנו רוצים שיתקיים $\forall i \in [d]$: אם $y_i = 1$ אז $b > -1$ ואם $y_i = -1$ אז $b < 1$ $\Leftarrow$ $b \in (0,1)$,
ככה יתקיים :

$$\forall i \in [d], \qquad sign(w^T x_i + b) = sign(w^T e_i + b) = sign(y_i + b) = sign(y_i) = y_i$$

בנוסף, עבור $i = d+1$: אם $y_{d+1} = 1$ אז $b > 0$ ו-אם $y_{d+1} = -1$ אז $b < 0$,
ככה יתקיים $sign(w^T x_{d+1} + b) = sign(w^T 0_{d\times 1} + b) = sign(b) = sign(y_{d+1}) = y_{d+1}$

אם נבחר $b = \frac{y_{d+1}}{2}$, נעמוד בתנאים כי $b \in (0,1)$ וגם b שומר על הסימן של y_{d+1}.

ונקבל : $\forall i \in [d+1],\ sign(w^T x + b) = sign(y_i) = y_i$

בשני הדרכים נקבל ש : $VC_{\dim}(\mathcal{H}^d) \geq d+1$

2. Let $\phi: \mathcal{X} \to \mathbb{R}^{n_1}, \phi': \mathcal{X} \to \mathbb{R}^{n_2}$ be two feature mappings where $n_1, n_2 \in \mathbb{N}$.

Let $K, K': (\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$ be two valid kernels defined as:

$$K(u,v) = \langle \phi(u), \phi(v) \rangle = \sum_{i=1}^{n_1} \phi_i(u)\phi_i(v), \ \ K'(u,v) = \langle \phi'(u), \phi'(v) \rangle = \sum_{j=1}^{n_2} \phi'_j(u)\phi'_j(v).$$

Prove that $G(u,v) \triangleq K(u,v) \cdot K'(u,v)$ is a valid kernel. That is, propose a feature mapping $\psi: \mathcal{X} \to \mathbb{R}^{n_3}$ for some $n_3 \in \mathbb{N}$ and prove that it holds $G(u,v) = \langle \psi(u), \psi(v) \rangle$.

Hint: You should use $n_3 = n_1 \cdot n_2$.

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

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

$$\stackrel{(*)}{=} \sum_{k=1}^{n_3} \psi_k(u)\psi_k(v) = \langle\, \psi(u)\psi(v) \rangle$$

לכן קיבלנו ש$G(u,v)$ קרנל חוקי.

(*) נסמן את $\psi_k(u) = \phi_i(u)\phi'_j(u)$ ונתאים ערך של k לכל i ו-j על ידי: $k = (j-1)n_1 + i$,
מתקיים ש $1 \leq i \leq n_1,\ \ 1 \leq j \leq n_2$, ולכן $1 \leq k \leq n_1 n_2$.

נראה שהפונקציה $k: [n_1] \times [n_2] \to [n_1 n_2]$, $k = (j-1)n_1 + i$ חד חד ערכית ועל.
נראה חד חד ערכיות: יהי $(i_1, j_1), (i_2, j_2)$ אם $k(i_1, j_1) = k(i_2, j_2)$ נקבל :

$$(j_1 - 1)n_1 + i_1 = (j_2 - 1)n_1 + i_2$$
$$j_1 n_1 - n_1 + i_1 = j_2 n_1 - n_1 + i_2$$
$$n_1(j_1 - j_2) = i_2 - i_1$$

מתקיים ש$|i_2 - i_1| \leq n_1$ ולכן השוויון הזה יהיה נכון רק אם $|j_1 - j_2| \leq 1$ מכיוון ש$0 < j_1, j_2$ מתקיים ש $j_1 = j_2$
ולכן $i_2 - i_1 = 0$ מתקיים אמ"מ $i_1 = i_2$ כלומר קיבלנו שאם $k(i_1, j_1) = k(i_2, j_2)$ אז $(i_1, j_1) = (i_2, j_2)$.
נראה שהפונקציה היא על :
לכל k קיימים i, j כך ש-$k = (j-1)n_1 + i$.
נסתכל על $k \leq n_1$ קיימים $i = k$ ו-$j = 1$ ונקבל ש- $(j-1)n_1 + i = (1-1)n_1 + i = k$

אם $k > n_1$ קיימים $i = k(mod\, n_1)$ ו- $j = \left\lceil\frac{k}{n_1}\right\rceil$ ונקבל ש

$$(j-1)n_1 + i = \left\lceil\frac{k}{n_1}\right\rceil n_1 - n_1 + k(mod\, n1) \overset{\substack{k=nn_1+k(mod\, n_1)\\ n\in\mathbb{N}}}{\triangleq} \left\lceil\frac{nn_1 + k(mod\, n_1)}{n_1}\right\rceil n_1 - n_1 + k(mod\, n1)$$

$$= \ (n+1)n_1 - n_1 + k(mod\, n1) = nn_1 + k(mod\, n1)$$

$$\overset{k=nn_1+k(mod\, n_1)}{\triangleq} \quad k - k(mod\, n_1) + k(mod\, n1) = k$$

(דרך החישוב נבעה מההסתכלות הבאה על הבעיה :

נסתכל על הפיתוח של הסכומים:

$$= \left(\phi_1(u)\phi_1(v) + \cdots\cdots + \phi_{n_1}(u)\phi_{n1}\,(v)\right)\cdot\left(\phi'_1(u)\phi'_1(v) + \cdots\cdots + \phi'_{n_2}(u)\phi'_{n_2}(v)\right)$$

$$= \phi_1(u)\phi_1(v)\left(\phi'_1(u)\phi'_1(v) + \cdots\cdots + \phi'_{n_2}(u)\phi'_{n_2}(v)\right) + \cdots\cdots$$

$$+ \phi_{n_1}(u)\phi_{n1}\,(v)\left(\phi'_1(u)\phi'_1(v) + \cdots\cdots + \phi'_{n_2}(u)\phi'_{n_2}(v)\right)$$

$$= \phi_1(u)\phi'_1(u)\phi_1(v)\phi'_1(v) + \phi_1(u)\phi'_2(u)\phi_1(v)\phi'_3(u) + \cdots + \phi_1(u)\phi'_{n_2}(u)\phi_1(v)\phi'_{n_2}(v)$$

$$+\phi_2(u)\phi'_1(u)\phi_2(v)\phi'_1(v) + \phi_2(u)\phi'_2(u)\phi_2(v)\phi'_3(u) + \cdots + \phi_2(u)\phi'_{n_2}(u)\phi_2(v)\phi'_{n_2}(v)$$

$$+\cdots\cdots+$$

$$+\phi_{n_1}(u)\phi'_1(u)\phi_{n_1}(v)\phi'_1(v) + \phi_{n_1}(u)\phi'_2(u)\phi_{n_1}(v)\phi'_3(u) + \cdots + \phi_{n_1}(u)\phi'_{n_2}(u)\phi_{n_1}(v)\phi'_{n_2}(v)$$

נסמן את $\psi_k(u) = \phi_i(u)\phi'_j(u)$,

נרצה למצוא את $i = f(k), j = g(k)$, אנחנו רואים פה שבכל טור i גדל ב-1, כלומר רץ מ1 ל-n_1 ובכל שורה i נשאר קבוע וj רץ מ1 ל-n_2.

לכן נבחר את הפונקציות הבאות שמתאימות לזה : $i = f(k) = k(mod\, n1) + 1 \ , j = g(k) = \left\lceil\frac{k}{n_1}\right\rceil$

לפי פתיחת הסכומים ניתן לראות שk רץ מ1 ל $n_1 \cdot n_2$, נסמן $n_3 = n_1 \cdot n_2$.

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

$$= \sum_{k=1}^{n_3}\left(\phi_{k(mod\, n1)}(u)\phi'_{\left\lceil\frac{k}{n_1}\right\rceil}(u)\right)\left(\phi_{k(mod\, n1)}(v)\phi''_{\left\lceil\frac{k}{n_1}\right\rceil}(v)\right)$$

$$\sum_{k=1}^{n_3}\psi_k(u)\psi_k(v) = \langle\,\psi(u)\psi(v)\rangle$$

3. **Refute** (with a simple example)**:** Let $f, g: \mathbb{R} \to \mathbb{R}$ be two convex functions.

The composition $h \triangleq f \circ g$ (that is, $h(x) = f\big(g(x)\big)$) is also a convex function.

נגדיר את שתי הפונקציות הבאות :

$$f(x) = -x$$

$$g(x) = x^2$$

נראה ששתי הפונקציות הבאות הן קמורות בדרכים שונות:

<u>$f(x) = -x$ קמורה :</u>

*פונקציה $f: C \to \mathbb{R}$ היא קמורה אם :

$$\forall x_1, x_2 \in C, \quad \forall t \in [0,1]: tf(x_1) + (1-t)f(x_2) \geq f(tx_1 + (1-t)x_2)$$

לכן, $f: \mathbb{R} \to \mathbb{R}$

$$\forall x_1, x_2 \in \mathbb{R}, \forall t \in [0,1]: \quad tf(x_1) + (1-t)f(x_2) = -tx_1 - x_2(1-t)$$
$$f(tx_1 + (1-t)x_2) = -tx_1 - x_2(1-t)$$
$$tf(x_1) + (1-t)f(x_2) = f(tx_1 + (1-t)x_2)$$

- אנחנו גם יודעים מהתרגול שכל פונקציה לינארית היא קמורה.
- $\mathbb{R}$ זאת קבוצה קמורה.

<u>$g(x) = x^2$ קמורה</u> –

נראה לפי המשפט מהתרגול : כל פונקציה גזירה פעמים $f: C \to \mathbb{R}$ היא קמורה אמ"מ $\nabla^2 f \succcurlyeq 0$
במקרה שלנו g היא $g: \mathbb{R} \to \mathbb{R}$ ומתקיים $g'(x) = 2x$ ו - $g''(x) = 2$,
לכן $g''(x) \succcurlyeq 0$ ולכן g קמורה.

נסתכל על $h(x) = f \circ g = f\big(g(x)\big) = -x^2$,

<u>נראה ש h לא קמורה:</u>

מתקיים ש- $h'(x) = -2x, \quad h''(x) = -2$
לכן לפי המשפט $h''(x) \prec 0$ ולכן h לא קמורה.

- מטריצת האסיאן לפונקציה במימד אחד היא מהצורה הבאה :

$$\nabla^2 f = \left[\frac{\partial^2 f}{\partial x^2}\right] \succcurlyeq 0 \iff \frac{\partial^2 f}{\partial x^2} \geq 0$$

הראנו שהרכבה של פונקציות קמורות לא בהכרח קמורה.

4. We will now prove that the following Soft-SVM problem is convex:

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

Let $f,g: C\to\mathbb{R}$ be two convex functions defined over a convex set C.

Lemma (no need to prove)**:** $q(z)\triangleq\max\{f(z),\ g(z)\}$ is convex w.r.t z.

Lemma (no need to prove)**:** the sum of any number of convex functions is convex.

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

4.2. Using a rule from Tutorial 07, conclude that $\max\{0,1-y_i\boldsymbol{w}^\top\boldsymbol{x}_i\}$ is convex w.r.t $\boldsymbol{w}$.

4.3. Using the above (and properties from Tutorial 07), conclude that the Soft-SVM optimization problem is convex w.r.t $\boldsymbol{w}$.

:4.1

בהינתן $\alpha\in\mathbb{R}_{\geq 0}$ ו- $f: C\to\mathbb{R}$ נראה ש-$h(z)=\alpha f(z)$ היא קמורה לפי הגדרה:

כלומר נראה ש -

$$\forall z_1,z_2\in C,\ \ \forall t\in[0,1]: th(z_1)+(1-t)h(z_2)\geq h(tz_1+(1-t)z_2)$$

הוכחה :

$$\forall x_1,x_2\in C,\ \ \forall t\in[0,1]:\ th(z_1)+(1-t)h(z_2)=\ t\alpha f(x_1)+(1-t)\alpha f(x_2)$$

$$=\alpha\big(tf(x_1)+(1-t)f(x_2)\big)\underbrace{\geq}_{\substack{f\ is\ convex\\ \alpha\geq 0}}\alpha\big(f(tz_1+(1-t)z_2)\big)=h(tz_1+(1-t)z_2)$$

: 4.2

- נסתכל על הפונקציה הבאה : $s(w)=-yw^Tx$, $s:\mathbb{R}^d\to\mathbb{R}$.
  $\mathbb{R}^d$ היא קבוצה קמורה ו-s זאת פונקציה לינארית לכל y וx,
  בתרגול 7 , לפי למה 1, כל פונקציה לינארית היא קמורה ולכן s היא פונקציה קמורה.
- הפונקציה הקבועה 1 היא גם לינארית ולכן גם היא קמורה.
- אנחנו יודעים מהנתון בשאלה שסכום פונקציות קמורות גם היא קמורה ולכן הפונקציה $1+s(w)$ היא קמורה.
- הפונקציה הקבועה 0 היא גם לינארית ולכן גם היא קמורה.
- מהנתון בשאלה אנחנו יודעים ש$q(z)\triangleq\max\{f(z),g(z)\}$ קמורה אם $f(z),g(z)$ קמורות .
  נסמן $f(w)=0$ ו -$g(w)=1+s(w)$, הראנו ששתיהן קמורות ולכן $q(w)=\max\{0,1-yw^Tx\}$ קמורה.

קיבלנו ש$\max\{0,1-yw^Tx\}$ קמורה על w .

:4.3

- נתון בשאלה שכל סכום של פונקציות קמורות היא קמורה וראינו בסעיף הקודם ש$\max\{0,1-yw^Tx\}$ קמורה ולכן $\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ קמורה.
- מתקיים ש $\frac{1}{m}$ חיובית ולכן נוכל להשתמש בסעיף 4.1 ונקבל ש: $\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_iw^Tx_i\}$ קמורה על w.

- ראינו בתרגול בשקופית 12 ש- $||w||^2$ היא קמורה ולפי סעיף 4.1 גם $\lambda\, ||w||^2$ קמורה.
- סכום של פונקציות קמורות היא פונקציה קמורה ולכן : $\frac{1}{m}\sum_{i=1}^{m}\max\{0,1-y_i w^T x_i\}+\lambda\, ||w||^2$ קמורה.

קיבלנו שבעיית האופטימזציה *soft-SVM* היא קמורה.