למידה חישובית - תרגיל 2 - תיאורטי

חגי ב.י., יונתן צ.

.1 -

(א) נגדיר

$$K_1 = \min_{w,b,\xi} \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 \quad s.t \ y_i(w^T x_i + b) \geq 1 - \xi_i \forall i = 1, ..., m$$

$$K_2 = \min_{w,b,\xi} \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 \quad s.t \ y_i(w^T x_i + b) \geq 1 - \xi_i \ \forall \wedge \xi_i \geq 0 \ \forall i = 1, ..., m$$

נוכיח כי $K_1 = K_2$.

i. $K_1 \leq K_2$: קיימים w, b, ξ כך ש: $K_2 = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2$ ומתקיימים עבורם האילוצים:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \wedge \xi_i \geq 0 \ \forall i = 1, ..., m$$

בפרט w, b, ξ מקיימים את כל האילוצים הנדרשים מפתרון של בעיית האופטימיזציה הראשונה. ופונק' המטרה בשני המקרים זהות, כלומר הפתרון הנתון מהווה פתרון לבעיה הראשונה, שערכו K_2ולכן המינימום של הבעיה הראשונה לא גדול מערך זה, כלומר $K_1 \leq K_2$.

ii. $K_2 \leq K_1$: קיימים w, b, ξ כך ש: $K_1 = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2$ ומתקיימים עבורם האילוצים:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \ \forall i = 1, ..., m$$

נגדיר פתרון חדש עבור בעיית האופטימיזציה השניה: $w, b, |\xi|$, (כלומר לוקחים ערך מוחלט של כל ξ_i מהפתרון המקורי ואת אותו w, b). אז הפתרון מקיים את האילוצים שכן לכל i: $y_i(w^T x_i + b) \geq 1 - \xi_i \geq 1 - |\xi_i|$ וכן ברור כי $0 \leq |\xi_i|$. הערך של הפתרון הוא: $\frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} |\xi_i|^2 = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 = K_1$, לכן $K_2 \leq K_1$.

(ב)

$$\mathcal{L}(w, b, \xi; \alpha) = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 + \sum_{i=1}^{m} \alpha_i (1 - \xi_i - y_i(w^T x_i + b))$$

(ג)

$$0 = \nabla_w L = w - \sum_{i=1}^{m} \alpha_i y_i x_i \Rightarrow w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$0 = \nabla_b L = -\sum_{i=1}^{m} \alpha_i y_i \Rightarrow \sum_{i=1}^{m} \alpha_i y_i = 0$$

$$0 = \nabla_{\xi_i} L = C\xi_i - \alpha_i \Rightarrow \xi_i = \frac{\alpha_i}{C}$$

נחשב את המינימום של הלגראנז'יאן בעזרת האילוצים שקבלנו:

$$
\begin{aligned}
g(\alpha) &= \min_{w,b,\xi} \mathcal{L}(w,b,\xi) \\
&= \min_{w,b,\xi} \frac{1}{2}w^Tw + \frac{C}{2}\sum_{i=1}^m \xi_i^2 + \sum_{i=1}^m \alpha_i(1-\xi_i - y_i(w^Tx_i+b)) \\
&= \min_{w,b,\xi} \frac{1}{2}w^Tw + \frac{C}{2}\sum_{i=1}^m \xi_i^2 + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i\xi_i - \sum_{i=1}^m \alpha_i y_i w^T x_i - b\sum_{i=1}^m \alpha_i y_i \\
&= \min_{w,b,\xi} \frac{1}{2}w^Tw + \frac{C}{2}\sum_{i=1}^m \xi_i^2 + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i\xi_i - w^T \overbrace{\sum_{i=1}^m \alpha_i y_i x_i}^{=w} - b\overbrace{\sum_{i=1}^m \alpha_i y_i}^{=0} \\
&= \min_{w,b,\xi} \frac{1}{2}w^Tw + \frac{1}{2C}\sum_{i=1}^m \alpha_i^2 + \sum_{i=1}^m \alpha_i - \frac{1}{C}\sum_{i=1}^m \alpha_i^2 - w^Tw \\
&= \min_{w,b,\xi} -\frac{1}{2}w^Tw + \frac{1}{2C}\sum_{i=1}^m \alpha_i^2 + \sum_{i=1}^m \alpha_i - \frac{1}{C}\sum_{i=1}^m \alpha_i^2 \\
&= -\frac{1}{2}(\sum_{i=1}^m \alpha_i y_i x_i)^T(\sum_{i=1}^m \alpha_i y_i x_i) + \sum_{i=1}^m \alpha_i + (\frac{1}{2C} - \frac{1}{C})\sum_{i=1}^m \alpha_i^2 \\
&= -\frac{1}{2}\sum_{i,j=1}^m \alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^m \alpha_i - \frac{1}{2C}\sum_{i=1}^m \alpha_i^2
\end{aligned}
$$

(ד) בעזרת סעיף קודם, הבעיה הדואלית:

$$
\begin{aligned}
\max_\alpha \min_{w,b,\xi} \mathcal{L}(w,b,\xi) = \max_\alpha & -\frac{1}{2}\sum_{i,j=1}^m \alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^m \alpha_i - \frac{1}{2C}\sum_{i=1}^m \alpha_i^2 \\
& s.t \alpha_i \geq 0 \\
& \sum_{i=1}^m \alpha_i y_i = 0
\end{aligned}
$$

.2 -

(א) כאמור, יש שיוויון בין ערך פונק' המטרה של הצורה הפרימאלית והדואלית של בעיית ה־SVM (בצורה ה־"רכה").
כלומר לפי הסימון שהוגדר בשאלה עבור הפתרונות האופטימליים והפיתוח מהשיעור מתקיים:

$$
\frac{1}{2}w^{*T}w^* + C\sum_{i=1}^m \xi_i^* = \sum_{i=1}^m \alpha_i^* - \frac{1}{2}\sum_{i,j=1}^m \alpha_i^*\alpha_j^* y_i y_j x_i^T x_j
$$

(ב) בשיעור ראינו כי מתקיים: $w^* = \sum_{i=1}^m \alpha_i^* y_i x_i$ (עבור פתרון המינימום), לכן:

$$
\begin{aligned}
& \sum_{i=1}^m \alpha_i^* - \frac{1}{2}\sum_{i,j=1}^m \alpha_i^*\alpha_j^* y_i y_j x_i^T x_j \\
= & \sum_{i=1}^m \alpha_i^* - \frac{1}{2}\sum_{i=1}^m \alpha_i^* y_i x_i^T \sum_{j=1}^m \alpha_j^* y_j x_j \\
= & \sum_{i=1}^m \alpha_i^* - \frac{1}{2}w^{*T}w^*
\end{aligned}
$$

$$
\frac{1}{2}w^{*T}w^* + C\sum_{i=1}^m \xi_i^* = \sum_{i=1}^m \alpha_i - \frac{1}{2}w^{*T}w^*
$$

$$w^{*T}w^* = \sum_{i=1}^{m} \alpha_i - C\sum_{i=1}^{m} \xi_i^*$$

לכן נובע ש: $\alpha_i < C$ וכן $C, \xi_i \geq 0$

$$|w^*|^2 = w^{*T}w^* \leq \sum_{i=1}^{m} \alpha_i \leq mC$$
$$|w^*| \leq \sqrt{mC}$$

.3 -

(א) *

(ב) הבעיה הדואלית היא: $\max_\alpha \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j x_i \cdot x_j,\ s.t\, \forall i.0 \leq \alpha_i \leq C$. אם נניח כי כל הקואורדינטות מלבד הראשונה קבועות נקבל את הבעיה הבאה:

$$\begin{aligned}
&\max_{\alpha_1} g(\alpha_1) \\
=&\max_{\alpha_1} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j x_i \cdot x_j \\
=&\max_{\alpha_1} \alpha_1 + \sum_{i\neq 1}\alpha_i - \frac{1}{2}\alpha_1^2 y_1^2 |x_1|^2 - (\alpha_1 y_1 x_1)\cdot\sum_{i\neq 1}(\alpha_i y_i x_i) - \frac{1}{2}\sum_{i\neq 1, j\neq 1}(\alpha_i y_i x_i)(\alpha_j y_j x_j) \\
=&\max_{\alpha_1} \left[-\frac{1}{2}\overbrace{y_1^2}^{=1}|x_1|^2\right]\alpha_1^2 + \left[1 - (y_1 x_1)\cdot\sum_{i\neq 1}(\alpha_i y_i x_i)\right]a_1 + \left[\sum_{i\neq 1}\alpha_i - \frac{1}{2}\sum_{i\neq 1, j\neq 1}(\alpha_i y_i x_i)\cdot(\alpha_j y_j x_j)\right] \\
=&\max_{\alpha_1} \left[-\frac{1}{2}|x_1|^2\right]\alpha_1^2 + \left[1 - (y_1 x_1)\cdot\sum_{i\neq 1}(\alpha_i y_i x_i)\right]a_1 + \left[\sum_{i\neq 1}\alpha_i - \frac{1}{2}\sum_{i\neq 1, j\neq 1}(\alpha_i y_i x_i)\cdot(\alpha_j y_j x_j)\right] \\
&s.t\ 0 \leq \alpha_1 \leq C
\end{aligned}$$

כלומר קבלנו בעית מקסימיזציה של פרבולה תחת אילוץ על המשתנה.

(ג) נגזור ונשווה ל־0 כדי לקבל את a_1 בו מתקבל המקסימום (ללא התחשבות באילוץ):

$$f' = \left[-|x_1|^2\right]\alpha_1 + \left[1 - (y_1 x_1)\cdot\sum_{i\neq 1}(\alpha_i y_i x_i)\right] = 0$$
$$a_1^* = \frac{1 - (y_1 x_1)\cdot\sum_{i\neq 1}(\alpha_i y_i x_i)}{|x_1|^2} = \frac{1 - \sum_{i\neq 1}(\alpha_i y_1 y_i x_1 \cdot x_i)}{|x_1|^2}$$

אז הפתרון הוא: $\begin{cases} g(a_1^*) & 0 \leq \alpha_1^* \leq C \\ g(0) & \alpha_1^* < 0 \\ g(C) & \alpha_1^* > C \end{cases}$ כי g היא פרבולה, הנגזרת השניה שלילית, ולכן הפתרון רק קטן ככל שמתרחקים בציר ה־x מ־α_1^*.

(ד) -

- <u>סיבוכיות זמן:</u> נסמן ב־m את מספר הדגימות וב־d את מימד מרחב הדגימות. בכל איטרציה מחשבים את α_i^*. לשם כך צריך לחשב $m-1$ פעמים: מכפלה פנימית - $O(d)$ ומספר קבוע של פעולות כפל, ולחבר את $m-1$ התוצאות, סכ"ה $O(md)$ (צריך לבצע עוד מספר קבוע של פעולות אריתמטיות שלא משנות את הסיבוכיות).
- <u>סיבוכיות מקום:</u> ניתן לבצע את חישוב a_i^* ב־$O(logd+logm)$ מקום. בכל איטרציה בחישוב הסכום צריך לשמור בזכרון שני אינדקסים - לוקטור ה־i ולאינדכס הרץ של הסכום - $O(logm)$. בנוסף בשביל חישוב המפכלה הפנימית צריך אינדקס בגודל $logd$. שאר החישוב מורכב ממספר קבוע של פעולות אריתמטיות הדורשות מקום קבוע (כפל/חיבור).

.4 -

(א) **לא בהכרח**, למשל: נגדיר $K_1(x,z)=0,\ K_2(x,z)=(1+xz)^2$. אז K_1 הוא הקרנל הטריוויאלי ז.א $K_1(x,z)=\phi_1(x)\phi_1(z)=\bar{0}\cdot\bar{0}=0$ כשהגדרנו $\phi_1(a)=\bar{0}$. ו־K_2 קרנל שראינו בשיעור. קבלנו:

$$K(x,z)=K_1(x,z)-K_2(x,z)=-(1+xz)^2$$

נניח בשלילה כי זה קרנל, אז קיימת פונקציה וקטורית ממשית ϕ כך ש־ $K=\phi\cdot\phi$. אז מצד אחד חייב להתקיים $K(x,x)=\phi(x)\cdot\phi(x)\geq 0$ מצד שני מקבלים עבור למשל $x=[0,0]$ (במימד שתיים),

$$K([0,0],[0,0])=-(1+[0,0]\cdot[0,0])^2=-1<0$$

בסתירה.

(ב) **לא בהכרח**, למשל עבור $K_1=K_2=0$ נקבל $K=\frac{K_1}{K_2}=\frac{0}{0}$ שהיא בכלל אינה מוגדרת (עבור אף קלט). ($\phi(x)=0\Rightarrow K(x,y)=\phi(x)\phi(y)=0\cdot 0=0$ לכן 0 הוא אכן קרנל).

(ג) **כן**, יהיו K_1,a>0, אז קיים ϕ_1 כך ש־ $K_1=\phi_1\cdot\phi_1$ אז אפשר לכתוב:

$$K(x,z)=aK_1(x,z)=a\phi_1(x)\cdot\phi_2(z)=(\sqrt{a}\phi_1(x))\cdot(\sqrt{a}\phi_1(z))$$

(השתמשנו בכך ש־a>0 כדי להוציא שורש ובלינאריות). נגדיר $\phi_2(y)=\sqrt{a}\phi_1(y)$ אז ϕ_2 פונק' וקטורית ממשית (מכפלה של פונק' כזו בקבוע ממשי), ומתקיים $K(x,z)=\phi_2(\mathrm{x})\cdot\phi_2(z)$ אז לפי הגדרה K הוא קרנל.

הוכחה נוספת: נגדיר $\overline{K}=aK$. אז לכל קבוצה $S=\{x_1,\ldots,x_n\}$ מתקיים

$$\left[\overline{K}_S\right]_{i,j}=a\left[K_S\right]_{i,j}$$

לכן לכל וקטור v מתקיים

$$\begin{aligned}v^T\overline{K}_Sv&=\sum_{i,j}v_ia\left[K_S\right]_{i,j}v_j\\&=a\sum_{i,j}v_i\left[K_S\right]_{i,j}v_j\\&=\overbrace{a}^{>0}\overbrace{v^TK_Sv}^{\geq 0}\geq 0\end{aligned}$$

לכן $\overline{K}$ הוא PSD, ולכן לפי משפט מרסר $\overline{K}$ הוא קרנל (FBK).

(ד) **לא בהכרח**, נגדיר $a=-1,\ K_1(x,z)=(1+xz)^2$, אז נקבל $K(x,z)=-(1+xz)^2$ וזו בדיוק הדוגמא מסעיף (a) שאינה קרנל.

(ה) **כן**, יהיו K,f כנ"ל, נגדיר $\phi(x)=[f(x)]$ (וקטור 1×1). אז $K(x,z)=f(x)f(z)=[f(x)]\cdot[f(x)]=\phi(x)\cdot\phi(z)$ כלומר K ניתנת לביטוי כמכפלה סקלרית של פונק' בעצמה, ולכן היא קרנל.

(ו) **כן**, נוכיח כי קב' הקרנלים סגורה לחיבור, כפל, ומכפלה בסקלר חיובי. מכך תנבע הטענה שכן כל פולינום הוא חזקה של משתנה (מכפלה של המשתנה בעצמו n פעמים), מכפלה של החזקה בסקלר חיובי, וחיבור של מספר סופי של ביטויים כאלו (מונומים).

- סגירות לחיבור: עבור כל שני קרנלים $K_1=\phi\cdot\phi, K_2=\psi\cdot\psi$ מתקיים $K=K_1+K_2$ הוא קרנל.

$$K(x,z)=\phi(x)\cdot\phi(z)+\psi(x)\cdot\psi(z)=[\phi(x);\psi(x)]\cdot[\phi(z);\psi(z)]=\gamma(x)\cdot\gamma(z)$$

כשהגדרנו $\gamma(x)=[\phi(x);\psi(x)]$ להיות השרשור של הפיצ'רים של K_1,K_2 המוגדרים על ידי ϕ,ψ.

- סגירות לכפל: עבור כל שני קרנלים $K_1=\phi\cdot\phi, K_2=\psi\cdot\psi$ מתקיים $K=K_1*K_2$ הוא קרנל:

$$\begin{aligned}K(x,z)&=(\phi(x)^T\phi(z))(\psi(x)^T\psi(z))\\&=(\phi(x)^T\phi(z))(\psi(z)^T\psi(x))\\&=tr(\phi(x)^T\phi(z)\psi(z)^T\psi(x))\\&=tr(\psi(x)\phi(x)^T\phi(z)\psi(z)^T)\\&=\left\langle VEC(\psi(x)\phi(x)^T),VEC(\phi(z)\psi(z)^T)\right\rangle\\&=\gamma(x)\cdot\gamma(z)\qquad(\gamma(a):=VEC(\psi(a)\phi(a)^T))\end{aligned}$$

כשהשתמשנו בתכונת סמטריות של מכפלה סקלרית, בתכונת הציקליות של tr, ובפונק VEC שמסמנת את העובדה שאנחנו מתייחסים למטריצה כאל וקטור (ראה scribes של שיעור 6: 6.1 Kernels for SVM Classifiers (page 7).

- סגירות לכפל בסקלר חיובי, ע"פ סעיף (c).

.5 -

(א) נוכיח באינדוקציה ש־ w_{T+1}הוא צירוף לינארי שלוקטורים מ־ $\{x_i\}$ $S=$, לכל $T\geq 0$.

- <u>בסיס:</u> $T=0$. צריך להראשות ש־ w_1 הוא צירוף לינארלי של איברים מ־S. לפי הגדרת האלגוריתם $w_1=(0,..0)$, לכן הוא ניתן לביטוי כצירוף לינארי טריוויאלי של איברי S.
- <u>צעד האינדוקציה:</u> נניח כי w_T הוא צירוף לינארי של וקטורים מ־S ונראה שגם w_{T+1} הוא צירוף לינארי. האלגוריתם יש שני מקרים אפשריים בכל צעד:
  - התחזית של האלגוריתם על הדוגמא x_T הייתה נכונה. (כלומר $\langle w_T,x_T\rangle\geq 0$ וגם $c*(x)=1$, או $\langle w_T,x_T\rangle<0$ וגם $c*(x)=-1$). אז $w_{T+1}=w_T$, לכן ע"פ ה.א w_{T+1} צ"ל של איברי S.
  - האלגוריתם שגה בתחזית שלו על הדוגמא X_T, אז מתקיים $w_{T+1}=w_T+x_T$ או $w_{T+1}=w_T-x_T$ (לפי אם שגינו על דוגמא חיובית או שלילית). ובכל מקרה, w_{T+1} הוא צ"ל של איברי S שכן ע"פ הנחת האינדוקציה $w_T=\sum a_ix_i$, אז

$$w_{T+1}=w_T+x_T=\sum a_ix_i+x_T=(a_T+1)x_T+\sum_{i\neq T}a_ix_i$$

וזה אכן צ"ל של איברי S (באותו אופן עבור המקרה בו מחסרים את x_T).

(ב) -

- לא ניתן לשמור את w^* באופן ישיר כי הוא עשוי להיות ממימד גדול מאוד (או אינסופי) וחישוב ישיר עליו עשוי להיות יקר מאוד (או בלתי אפשרי).
- נסמן ב־ϕ מיפוי של וקטורים למימד אחר ונניח כי קיים K כך ש: $K(x,y)=\langle\phi(x),\phi(y)\rangle$
- כדי לבצע את אלגוריתם ה־perceptron במימד אליו אנו ממפים יש לחשב את המפלכה הפנימית: $\langle w_{\mathrm{t}},\phi(x_t)\rangle$ לפי סעיף קודם מתקיים $w_t=\sum c^*(x_i)\phi(x_i)$ כש־x_i היא הדוגמא ה־i שסווגה לא נכון עד שלב t.

$$\begin{aligned}&\langle w_{\mathrm{t}},\phi(x_t)\rangle\\ =&\left\langle\sum c^*(x_i)\phi(x_i),\phi(x_t)\right\rangle\\ =&\sum c^*(x_i)\left\langle\phi(x_i)_i,\phi(x_t)\right\rangle\\ =&\sum c^*(x_i)K(x_i,x_t)\end{aligned}$$

כלומר אם נשמור את כל הדגימות עליהן טעינו ואת הסיווג הנכון עבורן, נוכל לבצע את החישוב ואת אלגוריתם ה־kernel perceptron. להלן פסודו־קוד של המימוש:

```
def predict(xt, samples, ops):
        dot_prod = 0
        for (s, op) in zip(samples, ops):
                dot_prod += op * K(s, xt)
        return dot_prod >= 0 ? 1 : -1

samples = []
ops = []
//training - output is (samples, ops)
while xt = get_next_sample():
        prediction = predict(xt, samples, ops)
        if prediction == c*(xt):
                continue

        samples.append(xt)
        ops.append( c*(xt) )

// example x is new sample
predict(x, samples, ops)
```