מבוא ללמידה חישובית - תרגיל בית 4

II

ד) יהיו נקודות x,y ו-$\alpha\in[0,1]$. צריך להראות ש-

$g(\alpha x+(1-\alpha)y)\le\alpha g(x)+(1-\alpha)g(y)$. נשתמש באינדוקציה על m.

בסיס - $m=1 \Leftarrow g=f_1$ וסיימנו.

הנחה: נסמן ש-$h=\max_{1\le i\le m-1} f_i$ הנחת אינד' קמורה, ויהי f_m פונקציה קמורה כלשהי. נסמן $z=\alpha x+(1-\alpha)y$ ונקבל כדלקמן:

$$
\begin{aligned}
g(z) &= \max(f_m(z), h(z)) \le \\
&\le \max(\alpha f_m(x)+(1-\alpha)f_m(y),\ \alpha h(x)+(1-\alpha)h(y)) \le \\
&\le \max(\alpha[f_m(x)+h(x)],\ (1-\alpha)[f_m(y)+h(y)]) = \\
&= \alpha\max(f_n(x), h(x)) + (1-\alpha)\max(f_m(y)+h(y)) = \\
&= \alpha g(x)+(1-\alpha)g(y)
\end{aligned}
$$

ה) יהי $x\in\mathcal{X}^m$. מספיק להראות ש-$|F_x|\le\Pi_{F_1}(m)\cdot\Pi_{F_2}(m)$ מכיוון ש-X שרירותי.

$$
\begin{aligned}
|F_x| &= \left|\{f_2(f_1(x_1)),\dots f_2(f_1(x_m)) \mid f_1\in F_1, f_2\in F_2\}\right| = \\
&= \left|\bigcup_{y\in F_{1x}}\{f_2(y_1),\dots,f_2(y_m)) \mid f_2\in F_2\}\right|
\end{aligned}
$$

מחסם האיחוד (Union Bound) נקבל:

$$
\le \sum_{y\in F_{1x}}\left|\{f_2(y_1),\dots,f_2(y_m)\}_{f_2\in F_2}\right| \le \sum_{y\in F_{1x}}\Pi_{F_2}(m) \le \Pi_{F_1}(m)\cdot\Pi_{F_2}(m)
$$

כנדרש.

הערה: הסימון F_x או F_{1x} שקול לסימון H_C שראינו בהרצאה - כל החלוקות האפשריות של פונקציות מ-F על המדגם X.

5) נתון: $\ell(x_{t+1}) \le \ell(x_t) + \nabla \ell(x_t)^t (x_{t+1}-x_t) + \frac{\beta}{2}\|x_t - x_{t+1}\|^2$

ידוע ש- $x_{t+1}-x_t = -\eta \nabla \ell(x_t)$ ונקבל:

$$\ell(x_{t+1}) \le \ell(x_t) + \nabla\ell(x_t)^t(-\eta\nabla\ell(x_t)) + \frac{\beta}{2}\|\eta\nabla\ell(x_t)\|^2 =$$

$$= \ell(x_t) - \eta \nabla\ell(x_t)^t\nabla\ell(x_t) + \frac{\beta}{2}\eta^2\|\nabla\ell(x_t)\|^2$$

$$\circledast = \ell(x_t) + \left(\frac{\beta}{2}\eta^2-\eta\right)\|\nabla\ell(x_t)\|^2$$

כי $x^t x = \|x\|^2$

נתון $\eta < \frac{2}{\beta}$ ולכן: $\frac{\beta}{2}\eta^2-\eta < \frac{\beta}{2}\cdot\frac{4}{\beta^2}-\frac{2}{\beta} = \frac{2}{\beta}-\frac{2}{\beta}=0$

ולכן כשנחלק ב- $\frac{\beta}{2}\eta^2-\eta$ הסימן מתהפך ונקבל:

$$\|\nabla\ell(x_t)\|^2 \le \frac{\ell(x_{t+1})-\ell(x_t)}{\frac{\beta}{2}\eta^2-\eta}$$

ולכן:

$$\sum_{t=0}^{n}\|\nabla\ell(x_t)\|^2 \le \sum_{t=0}^{n}\frac{\ell(x_{t+1})-\ell(x_t)}{\frac{\beta}{2}\eta^2-\eta} = \frac{1}{\frac{\beta}{2}\eta^2-\eta}\sum_{t=0}^{n}\ell(x_{t+1})-\ell(x_t) =$$

זהו טור טלסקופי ולכן:

$$= \frac{-1}{\frac{\beta}{2}\eta^2-\eta}\left[\ell(x_0)-\ell(x_{n+1})\right] \le \frac{1}{\eta-\frac{\beta}{2}\eta^2}\ell(x_0) := K < \infty$$

כאשר המעבר האחרון נובע מכך ש- $\ell(x_{n+1}) \ge 0$

קיבלנו ש- $\sum_{t=0}^{n}\|\nabla\ell(x_t)\|^2 < K$ כאשר K לא תלוי ב-n

אלא רק ב- $\ell(x_0), \beta, \eta$ לכן זה נכון לכל n,

הטור חסום ע״י קבוע לכל n ומתכנס כנדרש.

ב. לפי שאלה 1 מספיק להוכיח של-$\hat{y}\in L$ כלשהו מתקיים: $\ell(w_1,\dots,w_L,x,y)=w_{\hat{y}}\cdot x-w_y\cdot x+\Delta_{zo}(\hat{y},y)$ פונקציה קמורה וסיימנו כי מקסימום של פונקציות קמורות הוא קמור.

בנוסף ראינו שצירוף לינארי של פונקציות קמורות גם הוא קמור, לכן מספיק להראות שגם ל- $\Delta_{zo}(\hat{y},y)$, $w_y\cdot x$, $w_{\hat{y}}\cdot x$ קמורות.

- $\Delta_{zo}(\hat{y},y)$ - טריוויאלי (מדובר בפונקציה קבועה)
- $w_{\hat{y}}\cdot x, w_y\cdot x$ - פונקציות לינאריות ביחס ל- $w_y / w_{\hat{y}}$ לכן גם קמורות כנדרש.

ג. יהיו y,x,f. נסמן ב-y^* את התווית ש-f מחזיר ל-x, כלומר $f(x;w_1,\dots,w_L)=y^*$. אז:

$$\Delta_{zo}(f(x;w_1,\dots,w_L),y)=\Delta_{zo}(y^*,y)\le$$

$$\circledast\le\Delta_{zo}(y^*,y)+w_{y^*}x-w_y\cdot x\le$$

$$\le\max_{\hat{y}\in[L]}\left(\Delta_{zo}(\hat{y},y)+w_{\hat{y}}\cdot x-w_y\cdot x\right)=\ell(w_1,\dots,w_L,x,y)$$

כנדרש.

המעבר $\circledast$ נובע מכך ש- $w^*y\cdot x\ge w\cdot x$ מכך שבחרנו את y^*, $w_{y^*}x\ge w_yx$ לכל $y\in[L]$.

3) ג. נניח בשלילה שיש j כך: $\hat{y} := f(x_j; w_1^{opt},\dots,w_L^{opt}) \neq y_j$

כי הניבוי לא מדויק. אזי:

$$l(w_1^{opt},\dots,w_L^{opt},x_j,y_j) = \underbrace{\Delta_{zo}(\hat{y},y_j)}_{=1} + \underbrace{w_{\hat{y}}^{opt}\cdot x_j - w_{y_j}^{opt}\cdot x_j}_{\geq 0} \geq 1$$

לפי הגדרת f ונתון.

באופן דומה ניתן לראות שעבור w^* מדויק לכל $\hat{y}, i$:

$$l(w_1^*,\dots,w_L^*,x_i,y_i) \leq \Delta_{zo}(\hat{y},y_i) + w_{\hat{y}}^* x_i - w_{y_i}^* x_i$$

$$= \begin{cases} 0, & \hat{y}=y_i \\ 1+w_{\hat{y}}^* x_i - w_{y_i}^* x_i \end{cases} \leq 1$$

כי $w_{\hat{y}}^* x_i \leq w_{y_i}^* x_i$ לכל $\hat{y}, i$ לפי הגדרת w^*.

כלומר קיבלנו דוגמה (x_j,y_j) שבה w^* מקבל Loss קטן יותר מ-w^{opt}. אבל נתון:

$$\sum_i l(w_1^{opt},\dots,w_L^{opt},x_i,y_i) \leq \sum_i l(w_1^*,\dots,w_L^*,x_i,y_i)$$

לכן יש j כך ש- $l(w_1^*,\dots,w_L^*,x_j,y_j) > l(w_1^{opt},\dots,w_L^{opt},x_j,y_j)$

וגם בה"כ $\|x_i\|=1$ לכל i

נניח בה"כ ש- $\forall k: \|w_k^{opt}\| = \|w_k^*\| = 1$. מהשורה הקודמת

נובע:

$$\exists\hat{y}\,\forall\bar{y}.\ w_{\hat{y}}^* x_j - w_{y_j}^* x_j + \Delta_{zo}(\hat{y},y_j) > w_{\bar{y}}^{opt} x_j - w_{y_j}^{opt} x_j + \Delta_{zo}(\bar{y},y_j)$$

אם $\hat{y}=y_j$ סתירה להנחה שמצאנו את אלא מין א' אילי,

לכן $\hat{y}\neq y_j$, $\Delta_{zo}(\hat{y},y_j)=1$.

בפרט הנ"ל נכון לכל $\bar{y}$ בפרט $\bar{y}=\hat{y}$:

$$w_{\hat{y}}^* x_j - w_{y_j}^* x_j + 1 > w_{\hat{y}}^{opt} x_j - w_{y_j}^{opt} x_j + 1$$

$$\Downarrow$$

$$(w_{\hat{y}}^* - w_{y_j}^*) x_j > (w_{\hat{y}}^{opt} - w_{y_j}^{opt}) x_j$$

אם נמצא סתירה כי $w_{\hat{y}}^* - w_{y_j}^*$ יכול להיות קטן כרצוננו

שכן אם כל נכפול את w_k ב-$0<\alpha$ משהו לכל $k\in[L]$

הגדרת w^* לא תשתנה - הוא עדיין לא יעלה, כלומר

יש w^* חוקי עבורו הא"ש הנ"ל מדויק לכל

$x_j, y_j, \hat{y}\neq y_j$.

א. נניח בשלילה שיש פתרון אופטימלי לבעיה כאשר
מוגדרים כאשר יש j עבור $\varepsilon_j<0$. יהיו $\varepsilon_1,\dots,\varepsilon_m$
ונגדיר $\varepsilon'_1,\dots,\varepsilon'_m$:

$$\forall i\neq j:\ \varepsilon'_i=\varepsilon_i,\quad \varepsilon'_j=0$$

מבחינת אילוצים - כל אילוץ שלא כולל את ε_j
נשאר אותו דבר, כלו' האילוץ האחר:

$$y_j(w^Tx_j+b)\geq 1-\varepsilon_j \implies y_j(w^Tx_j+b)\geq 1-\varepsilon'_j$$

כי $1-\varepsilon'_j=0$ אבל $1-\varepsilon_j>1$ כי $\varepsilon_j<0$.
ומבחינת הערך האופטימלי ברור ש- $\frac{C}{2}\sum_{i=1}^m\varepsilon_i'^2<\frac{C}{2}\sum_{i=1}^m\varepsilon_i^2$
כי לכל $i\neq j$ $\varepsilon_i^2=\varepsilon_i'^2$ ו- $\varepsilon_j'^2=0<\varepsilon_j^2$.
כלומר מצאנו פתרון יותר טוב מ- $w,b,\varepsilon_1,\dots,\varepsilon_m$
בסתירה.

ב. הלגרנז'יאן הוא:

$$L(w,b,\varepsilon,\lambda)=\frac{1}{2}w^Tw+\frac{C}{2}\sum_{i=1}^m\varepsilon_i^2+\sum_{i=1}^m\lambda_i(1-\varepsilon_i-y_i(w^Tx_i+b))$$

ג. נגזור ונשווה ל-0:

$$0=\nabla_wL(w,b,\varepsilon,\lambda)=w+\sum_{i=1}^m-\lambda_iy_ix_i\implies w=\sum_{i=1}^m\lambda_iy_ix_i$$

$$0=\nabla_bL(w,b,\varepsilon,\lambda)=\sum_{i=1}^m-\lambda_iy_i\implies\sum_{i=1}^m\lambda_iy_i=0$$

$$0=\nabla_{\varepsilon_i}L(w,b,\varepsilon,\lambda)=C\varepsilon_i-\lambda_i\implies\varepsilon_i=\frac{\lambda_i}{C}$$

נציב חזרה ונקבל:

$$L(w,b,\varepsilon,\lambda)=\frac{1}{2}\Big(\sum_{i=1}^m\lambda_ix_iy_i\Big)^2+\frac{C}{2}\sum_{i=1}^m\frac{\lambda_i^2}{C^2}+\sum_{i=1}^m\lambda_i-\frac{\lambda_i^2}{C}=$$

$$=\frac{1}{2}\Big(\sum_{i=1}^m\lambda_ix_iy_i\Big)^2-\frac{1}{2C}\sum_{i=1}^m\lambda_i^2+\sum_{i=1}^m\lambda_i$$

ד. הבעיה הדואלית:

$$\max_{\lambda}\frac{1}{2}\Big(\sum_{i=1}^m\lambda_ix_iy_i\Big)^2-\frac{1}{2C}\sum_{i=1}^m\lambda_i^2+\sum_{i=1}^m\lambda_i$$

תחת האילוץ:

$$\sum_{i=1}^m\lambda_iy_i=0$$