מבוא למערכות לומדות

רועי זכאי 208641757

<u>תרגיל 5 - חלק תיאורטי:</u>

1) <u>Regularization</u>: יהיו $X\in\mathbb{R}^{m\times d}$ ו-$y\in\mathbb{R}^{m}$
ונניח כי X^TX הפיכה.
נסמן ב-$\hat{w}$ את פתרון הרגרסיה הלינארית הפשוטה
ו-$\hat{w}_\lambda$ את פתרון Ridge (כאשר $\lambda=0$ $\hat{w}_\lambda=\hat{w}$)

(a) נראה כי $\hat{w}_\lambda=A_\lambda\hat{w}$ כאשר $A_\lambda=(X^TX+\lambda I)^{-1}(X^TX)$

<u>פתרון</u>: בתרגול ראינו כי

$$\hat{w}_\lambda=(X^TX+\lambda I)^{-1}\cdot X^Ty$$

כמו כן ניזכר כי

$$\hat{w}=X^{\dagger}y=V\Sigma^{\dagger}U^Ty \quad , \quad V,U$$

מט' אורתו'

<u>נסמן</u>: $X=U\Sigma V^T$

$$A_\lambda\hat{w}=(X^TX+\lambda I)^{-1}(X^TX)\cdot\hat{w}=$$

$$=(X^TX+\lambda I)\cdot(V\Sigma^TU^TU\Sigma V^TV\Sigma^{\dagger}U^Ty)$$

נרצה להביא ביטוי זה ל-X^Ty <u>וסיימנו</u>!

ונזכר כי

$$\Sigma^+_{i,i} = \begin{cases} 1/\sigma_i & \sigma_i \neq 0 \\ 0 & else \end{cases}$$

$$\Sigma_{i,i} = \sigma_i$$

$$\Leftarrow \quad \Sigma\Sigma^+_{i,i} = \begin{cases} 1 & \sigma_i \neq 0 \\ 0 & else \end{cases}$$

נתון כי X^TX הפיכה, לכן אנו במקרה ההפיך אשר ראינו כבר כי אין ערכים סינגולריים שווים ל-0!

$$\Leftarrow \quad \Sigma^+_{i,i} = 1/\sigma_i \quad \text{לכל } i \text{ ו-} \sigma_j$$

$$\Sigma\Sigma^+ = I$$

$$\Rightarrow (V\Sigma^T U^T \cancel{U\Sigma V^T V\Sigma^+ U^T} y) = X^T y$$

$$\Leftarrow \quad A_\lambda \hat{w} = (X^TX + \lambda I)^{-1} X^T y = \hat{w}_\lambda$$

כנדרש!

6) נראה כי רידג' הינו אומד מוטה, קרי $\mathbb{E}(\hat{w}_\lambda) \neq w$ לכל סכל.

פתרון:

$$\mathbb{E}[\hat{w}_\lambda] = \mathbb{E}(A_\lambda \hat{w}) = \mathbb{E}\left((X^TX + \lambda I)^{-1} X^TX \hat{w}\right)$$

$$(X^TX + \lambda I)^{-1} X^TX \cdot \mathbb{E}(\cancel{\hat{w}}) \neq w$$

w ראינו כי $\hat{w}$ אומד חסר מוטה

מכיוון ש- לכל λ אז $(X^TX+\lambda I)^{-1}$ אינה ההופכית של X^TX, לכן גם כופלת את w בגורם שונה מ- יחידה ולכן $w \neq \hat{w}_\lambda$.

ב.) נראה כי היא מוטה:

[illegible]

$$Var(\hat{w}_\lambda) = Var(A_\lambda \hat{w}) =$$

$$= A_\lambda Var(\hat{w}) \cdot A_\lambda^T = A_\lambda \cdot \sigma^2 (X^TX)^{-1} \cdot A_\lambda^T =$$

[illegible]

$$= \sigma^2 A_\lambda (X^TX)^{-1} A_\lambda^T$$

ד.) נפתור כיוון מבורר ע"פ פירוק Bias-variance כפונקציה של λ.

תזכורת: בתרגול 4 ראינו כי

$$MSE = \mathbb{E}[\|\hat{y}-y\|^2] = Bias(\hat{y})^2 + Var(\hat{y})$$

$$\Downarrow$$

$$MSE = Bias(\hat{w}_\lambda)^2 + Var(\hat{w}_\lambda) =$$

$$= \mathbb{E}[\hat{w}_\lambda]^2 + \sigma^2 A_\lambda (X^TX)^{-1} A_\lambda^T =$$

$$\mathbb{E}[A_\lambda \hat{w}] = A_\lambda \mathbb{E}(\hat{w}) = A_\lambda w$$

$$\Rightarrow MSE = (A_\lambda w)^2 + \sigma^2 A_\lambda (X^T X)^{-1} A_\lambda^T$$

⇐ ניתן לראות כי המשוואה היא תלויה ב-λ ולכן הבענו לפי כנדרש!

ב)

$$\frac{d}{d\lambda} MSE(\hat{w}_\lambda)\Big|_{\lambda=0} :$$

ראשית נשים לב:

$$Bias^2(\hat{w}_\lambda) = \| \mathbb{E}[\hat{w}_\lambda] - \mathbb{E}[\hat{w}] \|^2 =$$

מן תכונות ש"מ!

$$= \| (A_\lambda - I) w \|^2 = w^T (A_\lambda - I)^T (A_\lambda - I) w$$

כמו כן, ראינו עוד כי בנוסף לפי השוואת תרג.2.11

$$Var(\hat{w}_\lambda) = \mathbb{E}\left[\| \hat{w}_\lambda - \hat{w} \|^2 \right] =$$

סכום Σ תוחלות הפרש הריבועים & אלכסון של השונות

$$= trace(Var(w_\lambda)) = \sigma^2 tr\left(A_\lambda (X^T X)^{-1} A_\lambda^T\right)$$

$$\Downarrow$$

$$\frac{d}{d\lambda} bias(\hat{w}_\lambda)^2 = \frac{d}{d\lambda} \sum_i \left(\sum_j (A_\lambda - I)_{ij} w_j \right)^2 \Big|_{\lambda=0}$$

$$= 2 \sum_i \left(\sum_j (A_\lambda - I)_{ij} w_j \frac{d}{d\lambda} \sum_j (A_\lambda - I)_{ij} w_j \right) = 0$$

⇐ לכן קיים רק λ חיובי ולכן הם מתאפס!

$$\frac{d}{d\lambda} Var(\hat{\alpha}_\lambda) = \sigma^2 \frac{d}{d\lambda} tr\left(A_\lambda X^T X A_\lambda^T\right)$$

מה

צו נג

הפונ' ב-A לפן נרצה

לגזור לפי ג

$$\frac{d}{d\lambda} Var(\hat{\alpha}_\lambda^T) = \sum_{i,j} \frac{d\, Var(\hat{\alpha}_\lambda^T)}{d(A_\lambda)_{i,j}} \cdot \frac{d(A_\lambda)_{i,j}}{d\lambda}\Big|_{\lambda=0} =$$

$$= Tr\left(\frac{d\, Var(\hat{\alpha}_\lambda^T)}{dA_\lambda} \cdot \frac{dA_\lambda}{d\lambda}\right)\Big|_{\lambda=0} =$$

$$= -2\underset{>0}{\sigma^2}\, Tr\big(\underset{>0}{(X^TX)^{-1}(X^TX)^{-T}}\big) < 0$$

⇐ קיבלנו כי נגזרת ה-MSE קטנה מ-0, כנדרש!

$\square$

f. נסיק מה ה' כי עבור $\hat{\alpha}_\lambda$ הנגזרת של MSE שלילית!
קרי מוסף ה-loss ומנקודה זה MSE יורדת, מה שאומר כי ההפסד קטן.

⇐ מכך קיבלנו כי יש נקודה בה $\lambda>0$, קרי זהו מודל LS סטנדרטי, אנו רואים כי הנגזרת שלילית ⇐ הפונק' יורדת.

מכיוון שהפונק' רציפה (מן הציורים) כאשר λ גדל

אך עדיין היה מאוד נקה כי ערך ה-MSE יהיה

$\Leftarrow$ הוספת ערך רגולריזציה קלות Ridge מעלין

את ה-MSE כנדרש!

PCA:

(ב) פתרון: נראה כי הערך המקסימלי של השונות מתקבל ע"י הו"ע של M מעל הגדול ביותר => השונות של v^Tx תסומן על ידי.

ראשית: יהי $v \in \mathbb{R}^d$, $\|v\|_2 = 1$

$$\mathbb{E}(v^TX) = \frac{1}{m}\sum_{i=1}^{m} v^Tx_i$$ כש שונות המדגם של x

$$= v^T\bar{x}$$ כמו כפי שרצינו בכתה

$$var(v^TX) = \mathbb{E}_x\left(\left(v^Tx_i - \mathbb{E}_x[v^Tx]\right)^2\right) =$$

$$= \frac{1}{m}\sum_i \left(v^Tx_i - v^T\bar{x}\right)^2 =$$

$$= \frac{1}{m}\sum_i \left[v^T(x_i - \bar{x})\right]^2 =$$

$$= \frac{1}{m}\sum_i \left(v^T(x_i - \bar{x})\right)\left(v^T(x_i - \bar{x})\right)^T =$$

$$= \frac{1}{m}\sum_i v^T(x_i - \bar{x})(x_i - \bar{x})^Tv =$$

$$= v^T \underbrace{\sum_i (x_i - \bar{x})(x_i - \bar{x})^T}_{\text{cov sample mat}} v =$$

$$= v^TSv$$

⇐ נרצה לכן

$$\hat{v} = \underset{\|v\| = 1}{\arg\max}\ v^TSv$$

דרך לינקסל נוספת!

נעזר ב- Lagrange multipliers

נגדיר $g(v) = 1 - v^T v$

$$\Downarrow$$

$$L = v^T v + \lambda g(v)$$

$$\frac{d}{dv} L = 2Sv - 2\lambda v = 0$$

=> אנו מקבלים כי המקסימום Q $v^T v$ הינו ו"ע

ניתן להסיק כי המקסימום הוא ו"ע S עם ע"ע גדול ביותר!

⇐ קיבלנו כי ההטלה שלנו תהיה הוקטור ע"ע ו"ע
של הע"ע הגדול ביותר מקסימום את שונות
$X^T v$ => ומכאן זו אכן ביטוי למקסימום של הבעיה
מההנחה, כנדרש!

∎

בונוס:

(3) יהי k הקרנל החוקי הנתון שלנו, כפי שראינו
כי עם פונק' f

$$\tilde{k}(x,x') = f(x)\cdot k(x,x')\cdot f(x')$$

הינו קרנל PSD חוקי!

ונבחר

$$f(x) = \begin{cases} 1/\sqrt{k(x,x)} & k(x,x) \geq 0 \\ -1/\sqrt{-k(x,x)} & \text{else} \end{cases}$$

⇐ מכאן נקבל:

$k(x,x) \geq 0$

$$\tilde{k}(x,x) = f(x)\,k(x,x)\cdot f(x) = \frac{1}{k(x,x)}\,k(x,x) = 1 \,//$$

אחרת $k(x,x) < 0$

⇐

$$k(x,x) = \frac{+1}{\sqrt{+k(x,x)}}\,k(x,x)\cdot\frac{+1}{\sqrt{+k(x,x)}} = 1 \,//$$

קרי באופן זה קיבלנו את הנדרש!

מש״ל

ג.) נכונה. נניח: $x \in \mathbb{R}^2$ כאשר $x_i = (x_i, y_i)$
נקודה ב-$\mathbb{R}^2$ ונקבע כי ה-label y_i נקבע
באופן הבא:
אם $\sqrt{x_i^2+y_i^2}$ זוגי אז נגדיר 1 אחרת
-1.
⇐ ונשים לב כי לא ניתן להפריד את S
הנתון ב-$\mathbb{R}^2$ באופן לינארי.

נגדיר $\psi(x) = (x_i, y_i, \sqrt{x_i^2+y_i^2} \,\%\, 2)$
⇐ נוסיף מימד שמייצג את זוגיות המרחק
באופן זה נוכל לקבל הפרדה בקלות כאשר
נוכל להפריד ע״י משטח את S שמפריד
אחרי!

5) (a) לא נכון! נראה דוגמה נגדית:

נראה כי הוא גרעין אינו ספק.

תהי

דגימות מה $S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

כאשר:

$$k = e^{(\|x-y\|^2)}$$

$$\Downarrow$$

$$K = \begin{pmatrix} 1 & e^2 \\ e^2 & 1 \end{pmatrix}$$

נראה ש K לא ספק, יהי $v = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

$$v^T K v = (-2+e^2 \quad -2e^2+1) \begin{pmatrix} -2 \\ 1 \end{pmatrix} =$$

$$-2(-2+e^2) - 2e^2 + 1 = 4 - 4e^2 + 1 = 5 - 4e^2$$

$\Leftarrow e^2 > 2 \Leftarrow 5-4e^2 < 0 \Leftarrow$ K לא ספק, כנדרש!

$\boxtimes$

ב) לא קרנל! נראה כי $k(x,x)<0$

נבחר שני קרנלים פולינומיאלים כפי שראינו בתרגול

$$k_1(x,z) = (\langle x,z\rangle + -5)^3$$

$$k_2(x,z) = (\langle x,z\rangle + 0)^2$$

$$\Downarrow$$

$$k(0,0) = k_1(0,0) - k_2(0,0) = -5^3 =$$

$$= -125 < 0$$

קיבלנו כי הנ"ל לא PSD ולכן הנ"ל לא קרנל חוקי!

$\blacksquare$

ג) הטענה נכונה!

הוכחה: יהיו k_a, k_b קרנלים ולידים כך ש-

$k = k_a + k_b$ נראה שהוא ולידי גם

ו- $x \in \mathbb{R}^d$ $\quad x = \begin{bmatrix} x_a \\ \hline x_b \end{bmatrix}$

נשים לב כי $k_{i,j} = k_{a_{i,j}} + k_{b_{i,j}}$

<= ולכן סימטרית!

נראה כי היא PSD.

$$k(x,x) = x k x^T = x(k_a + k_b)x^T$$

נראה כי גם קרנל זה מוגדר mxm כאשר כמות הדגימות של תצפית בגודל.

$$X = x_a^* + x_b^*$$

זה השורות עבור a ו-b ובכניסות הלא מוגדרות יש 0.

$$X K X^T = (x_a^* + x_b^*) K (x_a^* + x_b^*)^T =$$

$$= x_a^* K x_a^T + x_b^* K x_b^{*T} =$$

$$= x_a^* (K_a + K_b) x_a^{*T} + x_b^* (K_a + K_b) x_b^{*T}$$

⇐ מהגדרה הקרנלים החדשים מוגדרים רק על הקורד' בגלל הרלוונטיות אחרת נותן ערך 0

$$\Rightarrow x_a^* K_a x_a^{*T} + 0 + 0 + x_b^* K_b x_b^{*T} =$$

$$= \overset{\geq 0}{x_a K_a x_a^T} + \overset{\geq 0}{x_b K_b x_b^{*T}} \geq 0$$

כי הכפל של אפסים את וקטור הנ"ל.

⇐ הוא גם psd וזהו קרנל חוקי!