(20942) מבוא ללמידה חישובית | נקודות חשובות הרצאה 8

מנחה: ד"ר שי מימון
סמסטר: 2022א'
נכתב על ידי: מתן כהן

1 רגרסיה לינארית

ידוע לנו שבהינתן פתרון $\mathbf{w}^*$ לבעית הריבועים הפחותים וסט דוגמאות מ $\mathbb{R}$ ש:

(1) המטריצה X היא:

$$X_{N\times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$$

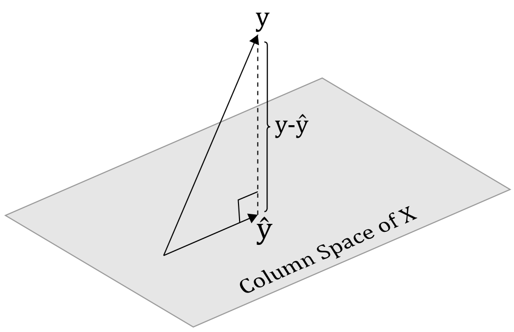
(2) $\hat{y}_i = w_0^* + w_1^* x_i$ וזהו ערך הדוגמה ה-i-ית על פי הכפולה בוקטור המשקלים $\mathbf{w}^*$

(א) מכך נסיק ש -

$$\hat{\mathbf{y}} = X\mathbf{w}^*$$

(ב) כמו-כן $\hat{\mathbf{y}}$ הוא במרחב העמודות של X ($\hat{\mathbf{y}} \in \mathcal{R}(X)$) ולכן הוא **אורתוגונלי** לשגיאה $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$

(3) הקשר בין $\hat{\mathbf{y}}$ ל- $\mathbf{y}$ הוא על פי:



(4) המשוואות הנורמליות מקיימות:

$$X^T \cdot \left(\underbrace{\mathbf{y} - X\mathbf{w}^*}_{\mathbf{y}-\hat{\mathbf{y}}=\mathbf{e}} \right) = \mathbf{0}$$

(5) מ- (4) נוכל להסיק שהשגיאה $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ היא **אורתוגונלית** לעמודות X

דוגמה. לשימוש בכל מה שציינו: נבדוק האם עבור פתרון $\mathbf{w}^*$ וסט סוגמאות מ- $\mathbb{R}$ נקבל $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - w_0^* - w_1^* x_i\right)\left(x_i - \tilde{x}\right) = 0$ (כאשר $\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$)

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - w_0^* - w_1^* x_i\right)\left(x_i - \tilde{x}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)x_i - \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)\tilde{x} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\underbrace{y_i - \hat{y}_i}_{e_i}\right)x_i - \tilde{x}\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)}_{=0} \\ &= \frac{1}{n}\sum_{i=1}^{n} e_i x_i = \frac{1}{n}\cdot \mathbf{x}^T\mathbf{e} \overset{(5)}{=} 0\end{aligned}$$

דרך אחרת היא להבין ש- $x_i - \tilde{x}$ נמצא במרחב העמודות של X ולכן הוא אורתוגונלי ל- **e**.

- הקשר בין הדוגמה שפתרנו לבין ביטוי של w_1^* הוא פשוט:
- תחילה נתבונן בשונות המדגם **x**:

$$C_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})(x_i - \tilde{x}) = \frac{1}{n}\sum_{i=1}^{n}x_i(x_i - \tilde{x}) - \tilde{x}\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x}) \tag{1.1}$$

- בצורה דומה עבור השונות המשותפת של **x** ו- **y** - $C_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})(y_i - \tilde{y}) \overset{\dots}{=} \frac{1}{n}\sum_{i=1}^{n} y_i\left(x_i - \tilde{x}\right)$ ואז:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - w_0^* - w_1^* x_i\right)\left(x_i - \tilde{x}\right) &= \frac{1}{n}\sum_{i=1}^{n} y_i\left(x_i - \tilde{x}\right) - w_0^*\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \tilde{x}\right) - w_1^*\frac{1}{n}\sum_{i=1}^{n}x_i\left(x_i - \tilde{x}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i\left(x_i - \tilde{x}\right) + w_0^*\left(\underbrace{\underbrace{\frac{1}{n}\sum_{i=1}^{n}x_i}_{\tilde{x}} - \underbrace{\tilde{x}\cdot\frac{1}{n}\cdot n}_{\tilde{x}}}_{\tilde{x}-\tilde{x}=0}\right) - w_1^*\frac{1}{n}\sum_{i=1}^{n}x_i\left(x_i - \tilde{x}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i\left(x_i - \tilde{x}\right) - w_1^*\left(\frac{1}{n}\sum_{i=1}^{n}x_i(x_i - \tilde{x}) - \tilde{x}\underbrace{\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})}_{=0}\right) \\ 1.1 &= C_{xy} - w_1^* C_{xx} \Rightarrow w_1^* = \frac{C_{xy}}{C_{xx}}\end{aligned}$$

2 שאלות על SVM

(1) כאשר הדוגמאות **לא** ניתנות להפרדה לינארית, ה- Hard-Margin SVM (ללא slack variable) יחזיר לנו $\mathbf{w}=0$

(א) **לא נכון** - מדובר ב- Hard-Margin SVM ולכן בהכרח לא נקבל תשובה בכלל כיוון שאין נקודה פיסיבילית שתקיים את האילוצים - משמע המערכת שתפתור את ה quadratic programming תקרוס.

(2) נניח שמשתמשים בבעיה הפרימלית בגרסאת ה-SVM עם Soft-Margin. כיצד נוכל להבטיח שסט הדוגמאות שלנו ניתן להפרדה בצורה לינארית?

(א) ניזכר בבעיה:

$$\begin{aligned} &\underset{\mathbf{w},b,\boldsymbol{\xi}}{\textbf{minimize}}\ \frac{1}{2}\mathbf{w}^T\mathbf{w}+{\color{red}C\cdot\sum_{n=1}^{N}\xi_n} \\ &\textbf{s.t}\quad y_n\cdot\left(\mathbf{w}^T\mathbf{x}_n+b\right)\geq 1{\color{red}-\xi_n},\quad n=1,..,N \\ &\qquad\ {\color{red}\xi_n\geq 0,\quad n=1,..,N} \end{aligned}$$

וכעת על מנת להבטיח את הרצוי יש צורך ב - $C\longrightarrow\infty$ כדי לא להפר את ה־Margin, ולאחר מכן לבדוק כמה דברים על מנת "לסגור את ההוכחה":

(i) $\xi_i=0$ לכל i

(ii) כלל הנקודות מסווגות נכון

(iii) כלל האילוצים מתקיימים

(iv) $\frac{1}{2}\mathbf{w}^T\mathbf{w}+{\color{red}C\cdot\sum_{n=1}^{N}\xi_n}\in\mathbb{R}$

(3) לאחר אימון SVM נוכל להסיר את כל הדוגמאות שהן לא support vectors ועדיין לסווג נכון נקודות חדשות

(א) **נכון** - כמו שאמרנו ניתן "לזרוק" את כל ה $\alpha_i=0$ כי רק אלו ששונות מ-0 הן אלו שמגדירות את השוליים של המפריד.

(4) עבור כל אחד מהמקרים הבאים נבדוק במה יותר כדאי להשתמש - בבעיה הפרימלית או הדואלית

(א) מיפינו את ממד הפיצ'רים שלנו לממד אינסופי

(i) כאן ברור כי הרבה יותר נוח להשתמש בבעיה הדואלית כיוון שהיא איננה מסתמכת על ממד הפי'צרים כמו הבעיה הפרימלית! הרי בבעיה הפרימלית אנחנו מנסים למזער את הפונקציה על פי $\mathbf{w}$ **מממד הפיצ'רים** ואילו בבעיה הדואלית רוצים למזער את α **מממד הדוגמאות**.

(ב) הכפלנו את ממד הפיצ'רים שלנו. מספר הדוגמרות הוא כמיליארד והדוגמאות ניתנות להפרדה לינארית

(i) זה בדיוק הפוך מהמקרה הקודם - כאן אנחנו נעדיף להתמודד עם בעיה שלא מסתמכת על ממד מגובה מיליארד, אלא על ממד הפי'צרים שככל הנראה פשוט יותר ורק הוכפל - משמע נרצה להשתמש בבעיה הפרימלית.

(ii) כמו-כן נבחין בעובדה **שהדוגמאות ניתנות להפרדה לינארית**, פרט זה חשוב כיוון שבגללו נבחר להשתמש ב-Hard-Margin כיוון שאם היינו בוחרים ב- Soft-Margin היינו צריכים עדיין להסתמך על מספר הדוגמאות כיוון שלכל דוגמה נצמיד ξ_i משלה.

(5) במידה והשתמשנו בבעיה הפרימלית לפתרון ה-SVM, כיצד היינו בודקים מי הם candidate support vectors?

(א) בודקים אילו נקודות מקיימות את האילוץ $y_n\cdot\left(\mathbf{w}^T\mathbf{x}_n+b\right)\geq 1$ בשוויון.