פתרון מטלה 2 — חישוביות וקוגניציה, 6119

22 בנובמבר 2025

שאלת הכנה

סעיף א׳

יהי פרספטרון לינארי N־מימדי הלומד פונקציה לא לינארית.

נבדוק מה ניתן לומר על שגיאת האימון ε_{tr} ועל שגיאת ההכללה ε_g לאחר שלמד $P < N$ דוגמות.

פתרון לצורך מענה על השאלה נניח שהפונקציה הנלמדת היא בלתי־לינארית לחלוטין, כלומר שאם $f : \mathbb{R}^N \to \mathbb{R}$ אז לכל $U \subseteq \mathbb{R}^N$ פתוחה לא קיימת העתקה לינארית כך ש־$f \restriction U$ לינארית.

במקרה זה נאמר ש־$\varepsilon_{\text{tr}}, \varepsilon_g > 0$ ושערכן המדויק תלוי בווקטור המשקולות המסוים שנבחר (תשובה ב׳), זאת שכן נוכל לבחור דוגמות שהתנהגותן יותר קרובה להיות לינארית.

סעיף ב׳

נבחן את המשוואה $C\bar{w} = \bar{u}$.

פתרון זוהי משוואה שנובעת מגזירה של תוחלת השגיאה, כלומר היא פתרון בעיית קיצון (תשובה א׳).

מטריצה היא רגולרית אם ורק אם $\det C \neq 0$, ולכן בהתאם נוכל לבודד בצורה יחידה את $\bar{w}$ אם ורק אם $\det(C) \neq 0$ (תשובה ד׳).

שאלה 1

יהי פרספטרון לינארי עם סף המנסה ללמוד את הפונקציה $Y_0 = X^3 - X^2$ כאשר $X \sim U([-1,1])$. נניח שלפרספטרון שני קלטים, $\bar{x} = (x, 1)$, ובהתאם מתקבל $y = w_1 x + w_2$.

סעיף א׳

נחשב את $\mathbb{E}(X^n)$.

פתרון נשתמש בטענה כי אם $Y = f(X)$ אז $\mathbb{E}(Y) = \int_{-\infty}^{\infty} f(x)\, f_X(x)\, dx$,

$$\mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n \cdot \frac{1}{2}\mathbb{1}_{[-1,1]}(x)\, dx = \frac{1}{2}\int_{-1}^{1} x^n\, dt = \frac{1}{2(n+1)} x^{n+1}\Big|_{x=-1}^{x=1} = \begin{cases} \frac{1}{n+1} & x \mod 2 = 0 \\ 0 & x \mod 2 = 1 \end{cases}$$

סעיף ב׳

נמצא וקטור משקולות אשר ממזער את שגיאת ההכללה $\varepsilon_g = \mathbb{E}(\frac{1}{2}(Y - Y_0)^2)$.

פתרון נבחין תחילה כי כנביעה מהסעיף הקודם נובע,

$$\mathbb{E}(Y_0^2) = \mathbb{E}(X^6 - 2X^5 + X^4) = \mathbb{E}(X^6) - 2\mathbb{E}(X^5) + \mathbb{E}(X^4) = \frac{1}{7} - 2 \cdot 0 + \frac{1}{5} = \frac{12}{35}$$

וכן מחישוב ישיר גם,

$$\mathbb{E}(Y^2) = \mathbb{E}(w_1^2 X^2 + 2w_1 w_2 X + w_2^2) = \frac{w_1^2}{3} + 0 + w_2^2 = \frac{w_1^2 + 3w_2^2}{3}$$

וגם,

$$\mathbb{E}(Y_0 Y) = \mathbb{E}((X^3 - X^2)(w_1 X + w_2)) = \mathbb{E}(w_1 X^4 - w_1 X^3 + w_2 X^3 - w_2 X^2) = w_1 \mathbb{E}(X^4) - w_2 \mathbb{E}(X^2) = \frac{w_1}{5} - \frac{w_2}{3}$$

נעבור לחישוב של שגיאת ההכללה,

$$\varepsilon_g = \mathbb{E}(\frac{1}{2}(Y - Y_0)^2) = \frac{1}{2}\mathbb{E}(Y^2 - 2Y_0 Y + Y_0^2) = \frac{1}{2}\mathbb{E}(Y^2) - \mathbb{E}(Y_0 Y) + \frac{1}{2}\mathbb{E}(Y_0^2) = \frac{w_1^2 + 3w_2^2}{6} - \frac{w_1}{5} + \frac{w_2}{3} + \frac{6}{35}$$

נבחן את השגיאה כפונקציה של w_1, w_2 ונגזור,

$$\frac{\partial \varepsilon_g}{\partial w_1} = \frac{1}{3}w_1 - \frac{1}{5}, \qquad \frac{\partial \varepsilon_g}{\partial w_2} = w_2 + \frac{1}{3}$$

ממשפט כופלי לגרנז׳ נובע ישירות ש-$\bar{w} = (\frac{3}{5}, -\frac{1}{3})^t$ הוא וקטור המשקולות בו $\varepsilon_g(\bar{w})$ מינימלי.

סעיף ג׳

נחשב את שגיאת ההכללה עבור הווקטור שמצאנו בסעיף הקודם.

פתרון נציב $w_1 = \frac{3}{5}, w_2 = \frac{1}{3}$,

$$\varepsilon_g = \frac{w_1^2 + 3w_2^2}{6} - \frac{w_1}{5} - \frac{w_2}{3} + \frac{6}{35} = \frac{3}{50} + \frac{1}{18} - \frac{3}{25} - \frac{1}{9} + \frac{6}{35} = \frac{88}{1575}$$

כלומר $\varepsilon_g \approx 0.05587$.

שאלה 2

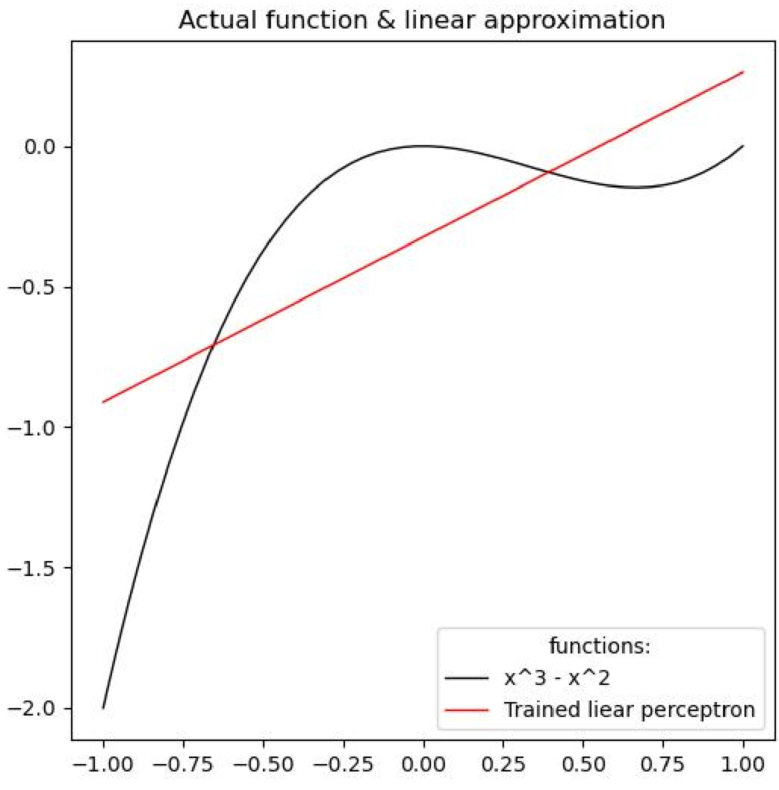
בשאלה הבאה נריץ מבחן חישובי לבדיקת פרספטרון לינארי על הפונקציה שהוצגה בשאלה הקודמת. נבחין כי הסעיפים השונים לאו דווקא מתייחסים לאותה ההרצה של המבחן, ובהתאם עלולים להיות הבדלים זעירים בתוצאות.

סעיף ב׳

נגדיר את המבחן אשר יוצר $P = 500$ דוגמות לווקטורים מהצורה $(x, 1)$ עבור $x \sim U([0, 1])$ וסיווגים בהתאם לפונקציה הלא לינארית המופיעה בשאלה 1. נפעיל את אלגוריתם האימון על הדוגמות והסיווגים. בהרצה שרירותית של קוד המבחן התקבל הווקטור $\bar{w} = (0.568, -0.33)^t$, וקטור זה מאוד קרוב לווקטור שחושב בשאלה הקודמת, הוא $(0.6, -0.33)^t$ בקירוב עשרוני.

סעיף ג׳

נציג את פעולת הפרספטרון המאומן, קרי את ההעתקה הלינארית אותה הוא מבצע, לצד ההעתקה הלא לינארית אותה הוא לומד מדוגמות. עבור הרצה שרירותית של קוד האימון נציג גרף של ההעתקה הלינארית אותה מבצע הפרספטרון לעומת הפונקציה המקורית הנתונה $x^3 - x^2$ בתחום.



סעיף ד׳

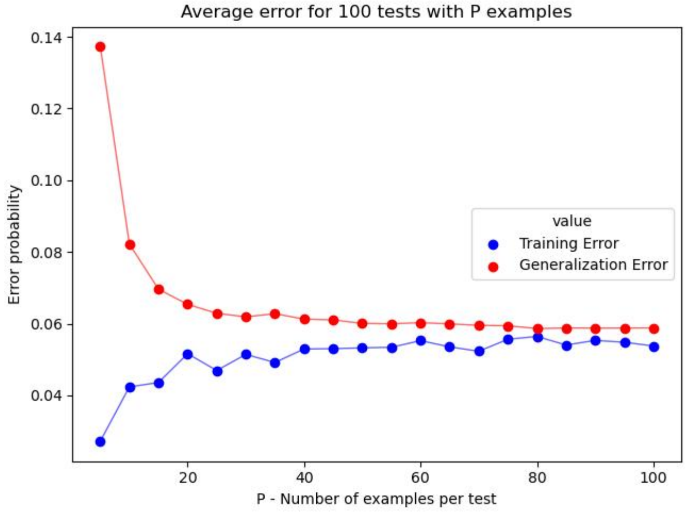
חישוב של שגיאת האימון ושגיאת הכללה של ריצה שרירותית מניב את התוצאות,

$$\varepsilon_{\text{tr}} = 0.055, \quad \varepsilon_g = -0.005$$

ערכים אלה לא דומים במיוחד, אך שניהם קטנים במידה שמאלצת אותנו להטיל ספק בדיוק החישוב.

סעיף ה׳

נריץ את הניסוי הממוחשב 100פעמים לכל מספר משתנים בין 5 ל־100 בקפיצות של 5. בגרף המצורף מופיע חישוב של שגיאת האימון וההכללה כפונקציה של מספר הדוגמות.



ניתן לראות כי שגיאת ההכללה שואפת מטה ומתכנסת ככל שהמערכת מקבלת עוד נתונים, מהצד השני ניתן לראות את שגיאת האימון גדלה ככל שיש יותר דוגמות, זאת שכן המערכת נמצאת במצב של איזון בין הדוגמות השונות.