מגיש: [illegible] [illegible]

ת"ז - 322357542

Theoretical Assignment 2

Intro to Machine Learning, TAU

November 29, 2024

Question 1

Show that the hyperbolic set $C_2 = \{x \in \mathbb{R}^2_+ | x_1 x_2 \geq 1\}$ is convex.
As a generalization, show that $C_n = \{x \in \mathbb{R}^n_+ | \prod_{i=1}^n x_i \geq 1\}$ is convex.
Hint: If $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$.

Question 2

The logistic loss function is frequently used in machine learning for binary classification. It is defined as:

$$f(z) = \log\left(1 + e^z\right),$$

where $z \in \mathbb{R}$.
Prove that $f(z)$ is convex.

Question 3

The following definition is a generalization of the definition of Jensen's inequality that we've seen in class for two points, to m points drawn from a distribution p:
Jensen's inequality states that for a convex function f and a probability distribution $\{p_i\}$ over points $\{\mathbf{x}_i\}$, we have:

$$f\left(\sum_{i=1}^m p_i \mathbf{x}_i\right) \leq \sum_{i=1}^m p_i f(\mathbf{x}_i).$$

Prove that $f(\mathbf{x}) = \|\mathbf{x}\|^2$ is convex using the above definition.

Question 4

Log in to Google Collab with your university email. Run this notebook, and follow the discussion points mentioned there. Add your answers to the pdf file with the solutions to the following questions.
Clarification: this is still considered a theoretical assignment as your goal is to analyze the results without implementing anything.

Question 1

Show that the hyperbolic set $C_2 = \{x \in \mathbb{R}^2_+ | x_1 x_2 \geq 1\}$ is convex.
As a generalization, show that $C_n = \{x \in \mathbb{R}^n_+ | \prod_{i=1}^n x_i \geq 1\}$ is convex.
Hint: If $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$.

מהגדרת קבוצה קמורה:

A set $C_2 \subseteq \mathbb{R}^2$ is convex if $\forall x, y \in C_2, \alpha \in [0,1]$

$$\Downarrow$$

$$(1-\alpha)x + \alpha y \in C_2$$

נגדיר

$$x = (x_1, x_2) \in C_2$$

$$\Downarrow$$

$$x_1 \cdot x_2 \geq 1$$

$$y = (y_1, y_2) \in C_2$$

$$\Downarrow$$

$$y_1 \cdot y_2 \geq 1$$

נגדיר

$$z = (1-\alpha)x + \alpha y = ((1-\alpha)x_1 + \alpha y_1, (1-\alpha)x_2 + \alpha y_2) = (z_1, z_2)$$

כיוון ש- $\boxed{\alpha \in [0,1]}$

ולכן $z \in \mathbb{R}^2_+$

(מורכב מסכום של מכפלות)

נוכיח ש- $z = (1-\alpha)x + \alpha y \in C_2$

כלומר, נראה

$$z_1 \cdot z_2 \geq 1$$

$$\left(\underbrace{(1-\alpha)x_1+\alpha y_1}\right)\cdot\left(\underbrace{(1-\alpha)x_2+\alpha y_2}\right)\ \geq$$

Hint: If $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^{\theta}b^{1-\theta} \leq \underbrace{\theta a+(1-\theta)b}$. מה נרמז -

$$x_1^{1-\alpha}\cdot y_1^{\alpha}\cdot x_2^{1-\alpha}\cdot y_2^{\alpha}=(x_1\cdot x_2)^{1-\alpha}\cdot(y_1\cdot y_2)^{\alpha}\geq 1^{1-\alpha}\cdot 1^{\alpha}$$

$$\boxed{=1}$$

$z_1\cdot z_2\geq 1$ לכן גם

$$z=(1-\alpha)x+\alpha y\in C_2$$

$C_n=\{x\in\mathbb{R}^n_+ \mid \prod_{i=1}^{n} x_i\geq 1\}$ עבור ה(n) נכליל

$x=(x_1,x_2,\ldots,x_n)\in C_n$ נבחר, נבחר -

$$\Downarrow$$

$$x_1\cdot x_2\cdot\ldots\cdot x_n\geq 1$$

$$y=(y_1,y_2,y_3\ldots y_n)\in C_n$$

$$\Downarrow$$

$$y_1\cdot y_2\cdots y_n\geq 1$$

נבחר

$$z=(1-\alpha)x+\alpha y=\left((1-\alpha)x_1+\alpha y_1,\ (1-\alpha)x_2+\alpha y_2,\ \ldots\ (1-\alpha)x_n+\alpha y_n\right)$$

$$=(z_1,z_2,\ldots,z_n)$$

$\boxed{\alpha\in[0,1]}$ כי

$z\in\mathbb{R}^n_+$ כאשר

(מוגדר עבור אי שליליים)

(נניח ש- $z \in C_n$

צריך, נראה $z_1 \cdot z_2 \cdot \cdots z_n \geq 1$

$$\Big(\overbrace{(1-\alpha)x_1+\alpha y_1}\Big)\cdot\Big(\overbrace{(1-\alpha)x_2+\alpha y_2}\Big)\cdot\cdots\big((1-\alpha)x_n+\alpha y_n\big) \geq$$

Hint: If $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \overbrace{\theta a + (1-\theta)b}$. מהרמז-

$$x_1^{1-\alpha}\cdot y_1^{\alpha}\cdot x_2^{1-\alpha}\cdot y_2^{\alpha}\cdots x_n^{1-\alpha}\cdot y_n^{\alpha} = (x_1\cdot x_2\cdots x_n)^{1-\alpha}\cdot(y_1\cdot y_2\cdots y_n)^{\alpha} \geq 1^{1-\alpha}\cdot 1^{\alpha}$$

$$\boxed{=1}$$

צריך $z_1 \cdot z_2 \cdots z_n \geq 1$

לכן קמורה

Question 2

The logistic loss function is frequently used in machine learning for binary classification. It is defined as:

$$f(z) = \log\left(1 + e^{z}\right),$$

where $z \in \mathbb{R}$.
Prove that $f(z)$ is convex.

$$f(z) = \log(1+e^{z})$$

הוכחה - פונ׳ קונבקסית - מספיק להראות שהנגזרת השנייה
(פונ׳ של משתנה אחד) של $f''(z)$ חיובי לכל $z \in \mathbb{R}$

$$f'(z) = \frac{1}{1+e^{z}} \cdot e^{z} \geq 0$$

$$1+e^{z} > 0$$

$$f''(z) = \frac{e^{z}(1+e^{z}) - e^{z}(e^{z})}{(1+e^{z})^2} = \frac{e^{z} + \cancel{e^{2z}} - \cancel{e^{2z}}}{(1+e^{z})^2}$$

$$= \frac{e^{z}}{(1+e^{z})^2} \quad e^{z} \geq 0, \quad (1+e^{z})^2 > 0$$

הנגזרת השנייה חיובית - עבור כל $z \in \mathbb{R}$

ראינו שההסיאן של פונ׳ קמורה היא מטריצה PSD
ושמטריצה שלנו מסדר 1x1 מטריצה זאת ע״ע של אחד
חיובי ממש

ווידאנו שאכן פונקציה קמורה.

Question 3

The following definition is a generalization of the definition of Jensen's inequality that we've seen in class for two points, to m points drawn from a distribution p:
Jensen's inequality states that for a convex function f and a probability distribution $\{p_i\}$ over points $\{\mathbf{x}_i\}$, we have:

$$f\left(\sum_{i=1}^{m} p_i \mathbf{x}_i\right) \leq \sum_{i=1}^{m} p_i f(\mathbf{x}_i).$$

Prove that $f(\mathbf{x}) = \|\mathbf{x}\|^2$ is convex using the above definition.

לפי ג'נסן, לפונק' קונבקסית f עם התפלגות הסתברות $\{p_i\}$ מעל נקודות $\{x_i\}$

מתקיים:

$$f\left(\sum_{i=1}^{m} p_i x_i\right) \leq \sum_{i=1}^{m} p_i f(x_i)$$

נוכיח: $f(x) = \|x\|^2$ קונבקסית –

נניח ש-$f(x)$ קונבקסית ונראה שהטענה אכן מתקיימת

נוכיח באינדוקציה –

בסיס: עבור $m=1$

$$f(p_1 x_1) = \|p_1 x_1\|^2 = p_1^2 \cdot \|x_1\|^2 \leq p_1 \cdot \underbrace{\|x_1\|^2}_{= f(x)}$$

$$0 \leq p \leq 1$$
$$\Downarrow$$
$$p_1^2 < p$$

צעד:

נניח שנכון עבור m -

$$\left\|\sum_{i=1}^{m} p_i x_i\right\|^2 \leq \sum_{i=1}^{m} p_i \|x_i\|^2$$

נוכיח עבור m+1 -

$$f\left(\sum_{i=1}^{m} p_i x_i\right) = \left\|\sum_{i=1}^{m+1} p_i x_i\right\|^2 = \left\|\left(\sum_{i=1}^{m} p_i x_i\right) + p_{m+1} x_{m+1}\right\|^2 \underset{\text{אי שוויון}}{\leq} \left\|\left(\sum_{i=1}^{m} p_i x_i\right)\right\| + \|p_{m+1} x_{m+1}\|^2 \leq$$

[illegible]

$$\sum_{i=1}^{m} p_i \|x_i\| + \|p_{m+1} x_{m+1}\|^2 \underset{(p_{m+1}^2 < p_{m+1})}{\leq}$$

$$\sum_{i=1}^{m} p_i \|x_i\| + p_{m+1} \|x_{m+1}\|^2 = \sum_{i=1}^{m+1} p_i \|x_i\|^2$$

$$= \sum_{i=1}^{m+1} p_i f(x_i)$$

לכן $f(x) = \|x_i\|^2$

קונבקסי -

Question 4

Log in to Google Collab with your university email. Run this notebook, and follow the discussion points mentioned there. Add your answers to the pdf file with the solutions to the following questions.
Clarification: this is still considered a theoretical assignment as your goal is to analyze the results without implementing anything.

1) training error - ניתן לראות שככל שהדרגה
הפולינום שלנו מקרב יותר טוב, ככה השגיאה
קטנה. במקרה זה יש קשר ישר בין שיעור העלייה
בדרגה לבין הירידה השגיאה.

2) testing error - ככל שעולה דרגת הפולינום
שלאחר מכן מקרבים השגיאה גדלה, כי יש קשר
ישיר בין שיעור העלייה דרגת הפולינום לגודל
השגיאה.

3) ניתן לראות שבדרגה degree 10 אנחנו מקבלים פער
גדול בין ה- training ל- testing, כלומר overfitting
ניתן לכן לראות שזה מספר אחרי מעלה 5.
וגם ראינו מודל ב- degree 1
היה עם פער מינימלי בין המודל לשגיאה
אבל לא היה מספיק להסביר הפולינום היעיל.

שמגיעים ל- best tradeoff.

אין נקודה של underfitting כי אין לנו אזור
שבו השגיאות של ה- testing, training דומות
וקרובות לאפס כו לא.

4. מהגרפים של קירוב מעלה 1 שמתאים לו ל. bias גבוה
כי הקירוב למטבע יהיה כ-קו ישר לגמרי
יחסית, כשהמטפל המשפט קצת שמאלה
למטה מהערך המקורי

ואפשר להוריד שלמקרה כבר השונות יחסית
בגלל המרחק של הנקודות
הרחוקות ביותר מהערך למעלה ולמטה.

מצד שני, במעלה מספר 3, יש שיפור קטן ב- bias
וב- variance אבל לא שיפור שווה של שונות והטיה
כמו מעלה 1.

במעלה מספר 7, השיפור שמקבלים יותר ב- bias.
הגרף נוטה למה קרוב למה שרוצים בדוגמה של
וקצב יותר מפורט, ומתרחק מהנקודות הרחוקות
יותר כלומר יש פער הרבה לכל של השונות.

בנוסף המקרה מעלה 10 אפשר לראות שגם trade off
הוא הכי טוב כי הפיצול הגדול יחסי — דפוסי הנתונ-
מגוון וגם השונות הוקטנה, וגם. bias עדיין קיים
אבל לא יותר מדי קטן כמו לקודם, כי יש יותר
פיצול לאיזור כבר מקודם.