Short HW 4

Ori Zohar - 205960750

Part A - Optimization

$$f(x) = \begin{cases} 2x & x \geq 0 \\ x^2 & x < 0 \end{cases}$$

1.1) אכן, $f(x)$ פונ׳ קמורה.

1.2) נבחר $g(u)$ ונראה שלכל $u \in \mathbb{R}$: לכל $v \in \mathbb{R}$ $f(v)-f(u) \geq g(u)(v-u)$

כלומר $g(u) \in \partial f(u)$ לכל $u \in \mathbb{R}$ $\Leftarrow$ g subgradient של f.

$$g(u) = \begin{cases} 2 & x \geq 0 \\ 2x & x < 0 \end{cases}$$

$\boxed{\text{עבור } u \geq 0 :}$

• $v \geq 0$:

$$f(v) - f(u) = 2v - 2u = 2(v-u) = g(u)(v-u)$$

• $v < 0$:

$$f(v) - f(u) = v^2 - 2u \geq 2v - 2u = 2(v-u) = g(u)(v-u)$$

$v^2 > 0 > 2v$

$\boxed{\text{עבור } u < 0 :}$

• $v \geq 0$:

$$f(v) - f(u) = 2v - u^2 \geq 2vu - u^2 > 2vu - 2u^2$$

$2v > 0 > 2vu$

$-u^2 < 0 \mid -u^2 \Rightarrow -2u^2 > -u^2$

$$= 2u(v-u) = g(u)(v-u)$$

• $v < 0$:

$$f(v) - f(u) = v^2 - u^2 \geq v^2 - u^2 - (v-u)^2 =$$

$$v^2 - u^2 - v^2 + 2uv - u^2 = 2uv - 2u^2 = 2u(v-u) = g(u)(v-u)$$

(1.3

$x_{i+1} = x_i - \eta \partial f(x_i)$

i	x_i	$f(x_i)$	$\partial f(x_i)$
0	$x_0 = -1$	1	-2
1	$x_1 = -1 - (0.25 \cdot (-2)) = -0.5$	0.25	-1
2	$x_2 = -0.5 - (0.25 \cdot (-1)) = -0.25$	0.125	-0.5
$\vdots$		$\vdots$	$\vdots$
n	$x_n = -\frac{1}{2^n}$	$f(x_n)$	$2 \cdot (x_n) =$

נשים לב שלא נגיע למינימום אלא רק נתקרב אליו כאשר $n \to \infty$

(1.4

i	x_i	$f(x_i)$	$\partial f(x_i)$
0	$x_0 = -1$	1	-2
1	$x_1 = -1 - (1 \cdot (-2)) = 1$	2	2
2	$x_2 = 1 - (1 \cdot (2)) = -1$	1	-2
			$\vdots$

נשים לב שלא נצליח להגיע למינימום.

Part 2 - Regression

נתון המודל הבא: $y = w^T x + \varepsilon \ , \ \varepsilon \sim Lap(0,b)$

ראשית נשים לב ש- $y|X=x \sim Lap(w^T x, b)$ מתכונות התפלגות ה- Lap

בהוספת סקלר, לכן.

$$f(y_i ; w^T x, b \,|\, X=x) = \frac{1}{2b} e^{-|y_i - w^T x|} = \frac{1}{2b} e^{-|w^T x - y_i|}$$

כעת נראה ש- $w_{MLE} = \underset{w}{\arg\min} \frac{1}{m} \sum_{i=1}^{m} |w^T x_i - y_i|$

הגדרה: $w_{MLE} =$

S מתפלג iid: $\arg\max P(S; w) =$

חוק הכפל וכלל השרשרת (לפי יונתן): $\arg\max \prod_i P(x_i, y_i ; w) =$

x בלתי תלוי ב- w: $\arg\max \prod_i P(y_i | x_i ; w) P(x_i ; w) =$

$\prod_i P(x_i)$ הוא סקלר ולכן לא משפיע argmax: $\arg\max \prod_i P(y_i | x_i ; w) P(x_i) =$

ln משמר argmax: $\arg\max \prod_i P(y_i | x_i ; w) =$

תכונות ln: $\arg\max \ln\left(\prod_i P(y_i | x_i ; w)\right) =$

הצבה: $\arg\max \sum_i \ln\left(P(y_i | x_i ; w)\right) =$

תכונות ln: $\arg\max \sum_i \ln\left(\frac{1}{2b} e^{-|w^T x_i - y_i|}\right) =$

$\arg\max \sum_i \ln\left(\frac{1}{2b}\right) + \ln\left(e^{-|w^T x_i - y_i|}\right) =$

האיבר $\ln(\frac{1}{2b})$ לא תלוי ב- w ולכן לא משפיע על argmax.

$$\arg\max \sum_i -|w^T x_i - y_i|$$

גורם משותף מסכום .

$$\arg\max -\sum_i |w^T x_i - y_i|$$

$$\arg\min \sum_i |w^T x_i - y_i|$$

$$\arg\min \frac{1}{m} \sum_i |w^T x_i - y_i|$$

תכונת argmax/argmin - הכפלה במינוס משמעותה

שינוי מין argmin ל- argmax ולהפך וסקלר חיובי

לא משנה

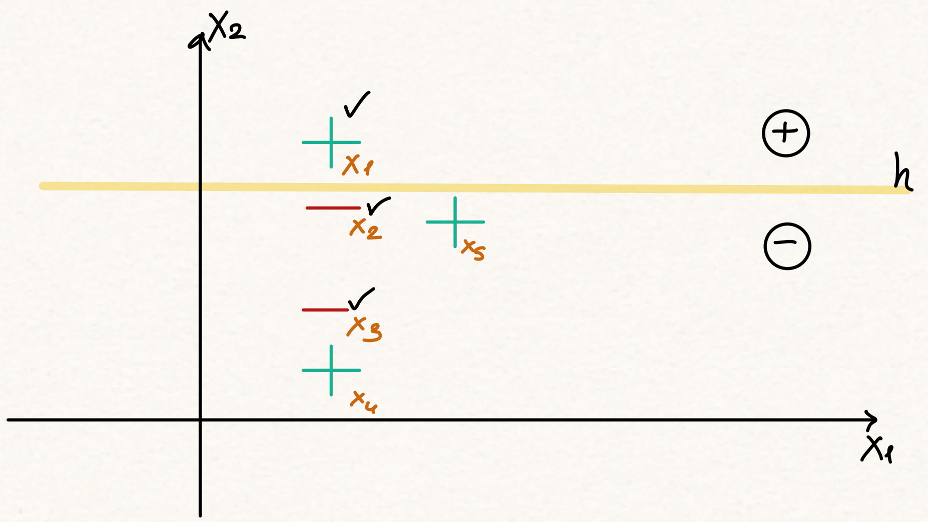
Part C - Boosting

נרצה למצוא h חלש כך שאחרי איטרציה 1 של AdaBoost נקבל את אחד מהאיורים a,b,c,d.

מהגדרת h חלש עליו להיות טוב יותר מסיווג רנדומלי כלומר ההסתברות לכך שיצדוק בתשובותיו צריכה להיות גדולה מ-$\frac{1}{2}$.

מאופן הפעולה של Adaboost ניתן משקל גדול יותר לנקודות שטעינו בהם לכן נקבל שבאיורים אנו מקבלים תמונת מצב לאחר איטרציה אחת נצפה לראות יותר מ-2 נקודות גדולות יותר, איור d נפסל.

נסתכל על המסווג הבא:



נשים לב שהוא טועה ב-2 דגימות ולכן הוא מסווג חלש.

נחשב את $D^{(2)}$:

- $D^{(1)} = (X_1, X_2, X_3, X_4, X_5) = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$
- $t = 1$:

1. $h_1(\bar{x}_i) = + \quad \forall x_i$

2. $\varepsilon_1 = \sum_{i=1}^{5} \frac{1}{5} \mathbb{1}\{y_i \neq h_1(\bar{x}_i)\} = \frac{2}{5}$

3. $\alpha_1 = \frac{1}{2} \log\left(\frac{1}{\frac{2}{5}} - 1\right) = \log\left(\sqrt{\frac{3}{2}}\right)$

אם: $h_1(\bar{x}_i) = y_i$ אז $h_1(\bar{x}_i) \cdot y_i = 1$ ולכן $D_i^{(2)} \propto \frac{1}{5} \cdot e^{-\sqrt{\frac{3}{2}}} = \frac{1}{5}\sqrt{\frac{2}{3}}$

אם: $h_1(\bar{x}_i) \neq y_i$ אז $h_1(\bar{x}_i) \cdot y_i = -1$ ולכן $D_i^{(2)} \propto \frac{1}{5} \cdot e^{\sqrt{\frac{3}{2}}} = \frac{1}{5}\sqrt{\frac{3}{2}}$

4. $D^{(2)} = \left(\frac{1}{5}\sqrt{\frac{2}{3}}, \frac{1}{5}\sqrt{\frac{2}{3}}, \frac{1}{5}\sqrt{\frac{2}{3}}, \frac{1}{5}\sqrt{\frac{3}{2}}, \frac{1}{5}\sqrt{\frac{3}{2}}\right)$

ואכן ניתן לראות כי עבור h שבחרנו $x_1, x_2, x_3 < x_4, x_5$ כפי שהנחנו

ולכן האיור שנבחר הוא (a).