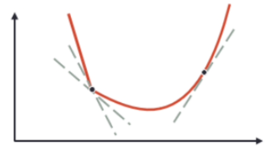
(*קיבלתי הארכה של שבוע להגשה*)

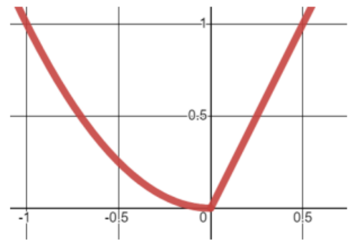
Part A – Optimization

Definition: the set of subgradients of $f: V \to \mathbb{R}$ at point $\boldsymbol{u} \in V$ is:



$$\partial f(\boldsymbol{u}) \triangleq \{\boldsymbol{q} \in V | \forall \boldsymbol{v} \in V : f(\boldsymbol{v}) \geq f(\boldsymbol{u}) + \boldsymbol{q}^{\top}(\boldsymbol{v} - \boldsymbol{u})\}.$$

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



1.1. Is f convex? No need to explain.

1.2. Propose a sub-derivative function g for f. That is, $g \in \partial f$.
Use the above definition to prove that $g(u) \in \partial f(u), \forall u \in \mathbb{R}$.

1.3. Set a learning rate of $\eta = 0.25$ and a starting point $x_0 = -1.5$.
Running subgradient descent, will the algorithm converge to a minimum?
Prove your answer by filling the following table like we did in Tutorial 07 using as many rows as needed.

i	x_i	$f(x_i)$	$\frac{\partial}{\partial x} f(x_i) = g(x_i)$
0	-1	1	
1			
⋮			

1.4. Repeat 1.3 with $\eta = 1,\ x_0 = -1.5$.

1. f קמורה.
2. נסתכל על הפונקציה הבאה :

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

נראה ש $g(u) \in \partial f(u)$, $\forall u \in \mathbb{R}$,

כאשר $\partial f(u) \triangleq \{q \in \mathbb{R} \,|\forall v \in \mathbb{R} : f(v) \geq f(u) + q^T(v-u)\}$.

ראשית, $\forall u \in \mathbb{R},\ g(u) \in \mathbb{R}$:

אם $u < 0 \Rightarrow g(u) = 2u \in \mathbb{R}$

אם $u \geq 0 \Rightarrow g(u) = 2 \in \mathbb{R}$

נחלק למקרים :

- כאשר $u, v < 0$:

$$\forall v \in \mathbb{R}, \qquad f(v) - f(u) - g(u)(v-u) = v^2 - u^2 - 2u(v-u) = v^2 - u^2 - 2uv + 2u^2$$
$$= v^2 - 2uv + u^2 = (v-u)^2 \geq 0$$
$$f(v) - f(u) - g(u)(v-u) \geq 0 \implies f(v) \geq f(u) + g(u)(v-u)$$

- כאשר $u, v \geq 0$:

$$\forall v \in \mathbb{R}, \qquad f(v) - f(u) - g(u)(v-u) = 2v - 2u - 2(v-u) = 2v - 2u - 2v + 2u$$
$$= 0 \geq 0$$
$$f(v) - f(u) - g(u)(v-u) \geq 0 \implies f(v) \geq f(u) + g(u)(v-u)$$

- כאשר $u \geq 0, v < 0$:

$$\forall v \in \mathbb{R}, \qquad f(v) - f(u) - g(u)(v-u) = v^2 - 2u - 2(v-u) = v^2 - 2u - 2v + 2u = v^2 - 2v \geq 0$$

$$f(v) - f(u) - g(u)(v-u) \geq 0 \implies f(v) \geq f(u) + g(u)(v-u)$$

- כאשר $v \geq 0, u < 0$:

$$\forall v \in \mathbb{R}, \qquad f(v) - f(u) - g(u)(v-u) = 2v - u^2 - 2u(v-u) = 2v - u^2 - 2uv + 2u^2 = u^2 - 2uv + 2v \geq 0$$

$$f(v) - f(u) - g(u)(v-u) \geq 0 \implies f(v) \geq f(u) + g(u)(v-u)$$

לכן,

$$\forall u \in \mathbb{R}: \;\; \forall v \in \mathbb{R}, f(v) \geq f(u) + g(u)(v-u), \qquad g(u) \in \partial f(u)$$

3. $x_0 = -1.5$, $\eta = 0.25$, $x_{i+1} = x_i - \eta \cdot \nabla f(x_i)$.

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

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

i	x_i	$f(x_i)$	$\boldsymbol{\frac{\partial}{\partial x} f(x_i) = g(x_i)}$
0	$-\frac{3}{2}$	$(-\frac{3}{2})^2 = \frac{9}{4}$	$\boldsymbol{2 \cdot -\frac{3}{2} = -3}$
1	$-\frac{3}{2} - 0.25 \cdot (-3) = -\frac{3}{4}$	$\left(-\frac{3}{4}\right)^2 = \frac{9}{16}$	$\boldsymbol{2 \cdot -\frac{3}{4} = -\frac{3}{2}}$
2	$-\frac{3}{4} - 0.25 \cdot \left(-\frac{3}{2}\right) = -\frac{3}{8}$	$\left(-\frac{3}{8}\right)^2 = \frac{9}{64}$	$\boldsymbol{2 \cdot -\frac{3}{8} = -\frac{3}{4}}$
3	$-\frac{3}{8} - 0.25 \cdot \left(-\frac{3}{4}\right) = -\frac{3}{16}$	$\left(-\frac{3}{16}\right)^2 = \frac{9}{256}$	$\boldsymbol{2 \cdot -\frac{3}{16} = -\frac{3}{8}}$
4	$-\frac{3}{16} - 0.25 \cdot \left(-\frac{3}{8}\right) = -\frac{3}{32}$	$\left(-\frac{3}{32}\right)^2 = \frac{9}{1024}$	$\boldsymbol{2 \cdot -\frac{3}{32} = -\frac{3}{16}}$
5	$-\frac{3}{32} - 0.25 \cdot \left(-\frac{3}{16}\right) = -\frac{3}{64}$	$\left(-\frac{3}{64}\right)^2 = \frac{9}{4096}$	$\boldsymbol{2 \cdot -\frac{3}{64} = -\frac{3}{32}}$
⋮ ⋮			

המינימום הוא בנקודה $x = 0$, אנחנו רואים שבכל איטרציה x גדל (ושואף לאפס) הפונקציה $f(x)$ קטנה ומתקרבת למינימום שלה $f(0) = 0$.
ניתן לראות שx_i ישאר שלילי בכל איטרציה ולכן יתקיים

$$x_{i+1} = x_i - \eta \cdot 2 \cdot x_i \overset{\eta=0.25}{\cong} x_i(1-0.5) = 0.5x_i$$
$$x_{i+1} = (0.5)^i x_0 = (0.5)^{i+1}(-1.5) \underbrace{\to}_{i\to\infty} 0$$

לכן האלגוריתם יתכנס.

4. $x_0 = -1.5$, $\eta = 1$, $x_{i+1} = x_i - \eta \cdot \nabla f(x_i)$.

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

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

i	x_i	$f(x_i)$	$\boldsymbol{\frac{\partial}{\partial x} f(x_i) = g(x_i)}$
0	$-\frac{3}{2}$	$(-\frac{3}{2})^2 = \frac{9}{4}$	$\boldsymbol{2 \cdot -\frac{3}{2} = -3}$
1	$-\frac{3}{2} - 1 \cdot (-3) = \frac{3}{2}$	$2 \cdot \frac{3}{2} = 3$	$\boldsymbol{2}$
2	$\frac{3}{2} - 1 \cdot 2 = -\frac{1}{2}$	$\left(-\frac{1}{2}\right)^2 = \frac{1}{4}$	$\boldsymbol{2 \cdot -\frac{1}{2} = -1}$
3	$-\frac{1}{2} - 1 \cdot (-1) = \frac{1}{2}$	$2 \cdot \frac{1}{2} = 1$	$\boldsymbol{2}$
4	$\frac{1}{2} - 1 \cdot 2 = -\frac{3}{2}$	$\left(-\frac{3}{2}\right)^2 = \frac{9}{4}$	$\boldsymbol{2 \cdot -\frac{3}{2} = -3}$
5	$-\frac{3}{2} - 1 \cdot (-3) = \frac{3}{2}$	$2 \cdot \frac{3}{2} = 3$	$\boldsymbol{2}$
⋮ ⋮			

אפשר לראות שהאלגוריתם לא מתכנס(כי ה η גדול מידי) יש חזרה על ערכי הx, כבר באיטרצה הרביעית אנחנו רואים שהx חוזר להיות אותו ערך כמו הערך ההתחלתי ולכן קיבלנו חזרתיות ,כלומר הוא יתקיים אינסוף פעמים ויגיע לאותו ערך x (כל 4 איטרציות) ולכן לא נתכנס לנקודת המינימום, כלומר לא נגיע ל $x = 0$.
האלגוריתם לא יתכנס.

Part B – Regression

2. This exercise will investigate the regularization coefficient λ as it was presented in the ridge linear regression section of this course. Suppose we are trying to fit a polynomial to the following data:

x	y
0	0
1	3
2	12

Our hypothesis class for this problem will be

$$\mathcal{H} = \{w_0 + w_1x + w_2x^2 + w_3x^3 : (w_0, w_1, w_2, w_3) \in \mathbb{R}^4\}.$$

2.1. Show that we can fit the data with $w_0 = 0, w_1 = 2, w_2 = 0, w_3 = 1$.

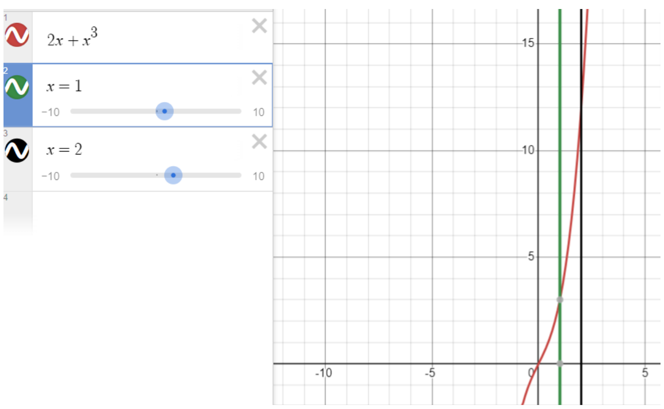
עבור $w = (0,2,0,1)$ נקבל פולינום $p(x) = 2x + x^3 \in \mathcal{H}$.

נראה שכל נקודה ב *data* נמצאת על הפולינום הנ"ל.

עבור $x_1 = 0, y_1 = 0$: $p(x_1) = p(0) = 0 = y_1$

עבור $x_2 = 1, y_2 = 3$: $p(x_2) = p(1) = 3 = y_2$

עבור $x_3 = 2, y_3 = 12$: $p(x_3) = p(2) = 12 = y_3$



2.2. Show that our hypothesis class is too expressive for the problem we're dealing with. In other words, find a simple quadratic polynomial that fits the data perfectly.

נמצא פונקציה ריבועית מהצורה: $p(x) = w_0 + w_1x + w_2x^2$

נציב $(2,12)$, $(1,3)$, $(0,0)$:

$$(0,0) \implies 0 = w_0$$
$$(1,3) \implies 3 = w_1 + w_2$$
$$(2,12) \Rightarrow 12 = 2w_1 + 4w_2 \Rightarrow 6 = w_1 + 2w_2$$
$$w_1 = 3 - w_2 \ \Rightarrow 6 = 3 - w_2 + 2w_2 = 3 + w_2 \ \Rightarrow w_2 = 3$$
$$w_1 = 0$$

הפונקציה הריבועית: $p(x) = 3x^2$

מתקיים ש $p(x) \in \mathcal{H}$

מצאנו שיש פונקציה ריבועית שמתאימה ל*data* ולכן מחלקת ההיפותזות שאנחנו משתמשים בה יותר מידי *expressive* וניתן להשתמש במחלקת היפותזו קטנה יותר, פולינומים ממעלה 2.

$$\mathcal{H} = \{w_0 + w_1x + w_2x^2 : (w_0, w_1, w_2) \in \mathbb{R}^3$$

2.3. Denote the mean squared error (MSE)

$$\mathcal{L}(w) = \frac{1}{m}\|\boldsymbol{X}w - y\|_2^2,$$

Where $\boldsymbol{X}$ is the appropriate Vandermonde matrix.

Calculate $\mathcal{L}(\mathrm{w})$ for the quadratic model in (2.2) and the cubic model in (2.1).

<u>נסתכל על הבעיה מ2.1 כאשר $p(x) = 2x + x^3$</u>

מטריצת ונדרמונד לפולינום ממעלה 3 ל*data* עם 3 נקודות: $X = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \end{pmatrix}$

נציב $(2,12)$, $(1,3)$,$(0,0)$ ונקבל: $X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix}$

$$w = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \qquad y = \begin{pmatrix} y_0 \\ y_1 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix}$$

$$\mathcal{L}(w) = \frac{1}{m} \| Xw - y \|_2^2 = \frac{1}{3} \| \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} \|_2^2 = \frac{1}{3} \| \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} \|_2^2 = 0$$

<u>נסתכל על הבעיה מ3.2 כאשר $p(x) = 3x^2$</u>

מטריצת ונדרמונד לפולינום ממעלה 2 ל*data* עם 3 נקודות: $X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}$

נציב $(2,12)$, $(1,3)$,$(0,0)$ ונקבל: $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}$

$$w = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}, \qquad y = \begin{pmatrix} y_0 \\ y_1 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix}$$

$$\mathcal{L}(w) = \frac{1}{m} \| Xw - y \|_2^2 = \frac{1}{3} \| \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} \|_2^2 = \frac{1}{3} \| \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} \|_2^2 = 0$$

- שני המודלים מתאימים ל*data* ולכן התוצאה שקיבלנו $\mathcal{L}(w) = 0$ בשני המקרים לא מפתיעה.

2.4. The best <u>line</u> for fitting this data is $y = 6x - 1$. Calculate $\mathcal{L}(w)$ for this line.

נסתכל על הבעיה כאשר $:p(x) = 6x - 1$

מטריצת ונדרמונד לפולינום ממעלה 3 לdata עם 3 נקודות: $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix}$

נציב $(2,12)$, $(1,3)$,$(0,0)$ ונקבל: $X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$

$$w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} -1 \\ 6 \end{pmatrix}, \qquad y = \begin{pmatrix} y_0 \\ y_1 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix}$$

$$\mathcal{L}(w) = \frac{1}{m} \| Xw - y \|_2^2 = \frac{1}{3} \| \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 6 \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} \|_2^2 = \frac{1}{3} \| \begin{pmatrix} -1 \\ 5 \\ 11 \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 12 \end{pmatrix} \|_2^2 =$$

$$\frac{1}{3} \| \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \|_2^2 = \frac{1}{3} \cdot 6 = 2 \quad \Rightarrow \mathcal{L}(w) = 2$$

2.5. Now denote the MSE with regularization as shown in class

$$\mathcal{L}_\lambda(w) = \frac{1}{m} \|\boldsymbol{X}w - y\|_2^2 + \lambda \|w\|_2^2.$$

Here $\lambda > 0$ is a hyperparameter, which is not given. As we learned in class, the regularization imposes a "cost" on models with large coefficients. Calculate $\mathcal{L}_\lambda(w)$ for each of the three models in (2.1), (2.2), (2.4).

עבור המודל $p(x) = 2x + x^3$

$$\mathcal{L}_\lambda(w) = \frac{1}{m} \| Xw - y \|_2^2 + \lambda \| w \|_2^2 \overset{q2.1}{\triangleq} 0 + \lambda \| \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} \|_2^2 = 5 \cdot \lambda$$

עבור המודל $p(x) = 3x^2$

$$\mathcal{L}_\lambda(w) = \frac{1}{m} \| Xw - y \|_2^2 + \lambda \| w \|_2^2 \overset{q2.2}{\triangleq} 0 + \lambda \| \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \|_2^2 = 9 \cdot \lambda$$

עבור המודל $:p(x) = 6x - 1$

$$\mathcal{L}_\lambda(w) = \frac{1}{m} \| Xw - y \|_2^2 + \lambda \| w \|_2^2 \overset{q2.4}{\triangleq} 2 + \lambda \| \begin{pmatrix} -1 \\ 6 \end{pmatrix} \|_2^2 = 2 + 37 \cdot \lambda$$

2.6. As it turns out, $\mathcal{L}_\lambda(w)$ would never prefer the simple quadratic polynomial over the cubic polynomial we found, no matter the value of $\lambda > 0$. Can you explain why?

ראינו בסעיף קודם ש $\mathcal{L}_\lambda(w)_{quadratic} = 9 \cdot \lambda > 5 \cdot \lambda = \mathcal{L}_\lambda(w)_{cubic}$ לכל $\lambda > 0$.

וזה כי בשני המקרים ה*data* מותאם למודל בצורה מושלמת ולכן ה"עונש" הוא רק על הנורמה של *w*

והנורמה של $w_{quadratic} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$ בעוד ש- $w_{cubic} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$ ומתקיים ש $\| w_{quadratic} \| < \| w_{cubic} \|$.

במודל ה*MSE* אנחנו מחפשים *w* שממזער את האובייקט $\mathcal{L}_\lambda(w) = \frac{1}{m} \| Xw - y \|_2^2 + \lambda \| w \|_2^2$ ובמקרה של ה*data* שלנו והמודלים המוצגים זה מתקיים עבור המודל $p(x) = 2x + x^3$. ולכן הוא ייבחר.

הערה : כאשר אנחנו מגדילים את λ אנחנו בעצם נותנים יותר משקל לנורמה של *w* שמייצגת את סיבוכיות המודל. כ λ גדלה למעשה אנחנו דורשים שהמודל שלנו יתאים פחות ל*data* ויהיה יותר פשוט.

במקרה הזה גם של λ גדול מאוד אנחנו עדיין נעדיף מודל יותר מורכב (מודל ממעלה 3 על פני מודל ממעלה 2) וזה כי אנחנו מקבלים *coefficient* דלילים שמתאימים ל*data* ולכן יש לדרוש מלכתחילה מחלקת היפותזות מתאימה ל*data* ולא מורכבת מידי כי יכול להיות שיבחר מודל מורכב יותר שלא לצרוך.

2.7. Suggest a way to fix the regularization method to prefer the model we consider to be simpler.

בשביל לתקן את שיטת הרגולריזציה שתבחר את המודל שניחשב פשוט יותר נשתמש ב"הענשה" של מודלים המציעים פולינום ממעלה גדולה יותר, כלומר במקרה שלנו, מודלים ממעלה 3.

נוסיף מטריצה $C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$, כך ש - $\mathcal{L}_\lambda(w) = \frac{1}{m} \| Xw - y \|_2^2 + \lambda \| Cw \|_2^2$.

נקבל עבור $w_{quadratic} = \begin{pmatrix} 0 \\ 0 \\ 3 \\ 0 \end{pmatrix}$ ש- $\| Cw_{quadratic} \|_2^2 = \| \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 3 \\ 0 \end{pmatrix} \|_2^2 = \| \begin{pmatrix} 0 \\ 0 \\ 6 \\ 0 \end{pmatrix} \|_2^2 = 36$

נקבל עבור $w_{cubic} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$ ש- $\| Cw_{cubic} \|_2^2 = \| \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} \|_2^2 = \| \begin{pmatrix} 0 \\ 2 \\ 0 \\ 8 \end{pmatrix} \|_2^2 = 66$

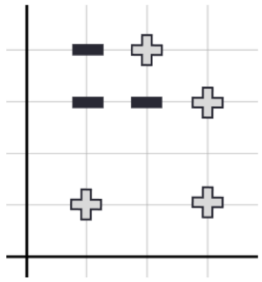
ולכן נקבל :

עבור המודל $p(x) = 2x + x^3$: $\mathcal{L}_\lambda(w_{quadratic}) = 36 \cdot \lambda$

עבור המודל $p(x) = 3x^2$: $\mathcal{L}_\lambda(w_{cubic}) = 66 \cdot \lambda$

לכן יבחר המודל הנחשב פשוט יותר *a simple quadratic polynomial*.

Part C – Boosting

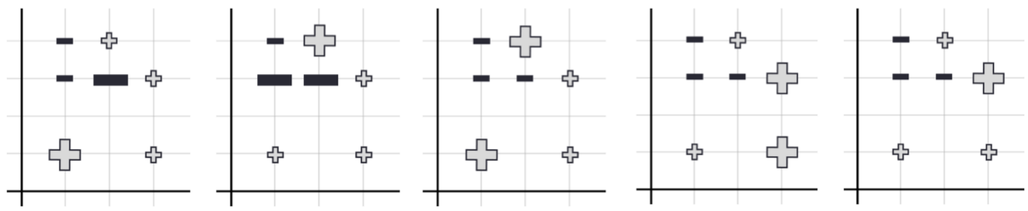


3. Given the following data with binary labels ("+", "-").

   We run AdaBoost with Decision stumps as weak classifiers.
   The sizes of the shapes in the figures indicate the probabilities that the algorithm assigns to each sample (high probability = large shape). Initially, the algorithm starts from a uniform distribution.

   Only some of the following figures depict possible distributions that can be obtained after one iteration of AdaBoost. **Which ones?** For each such distribution, propose a weak classifier that can lead to its figure (use a clear drawing or a short description of that classifier).

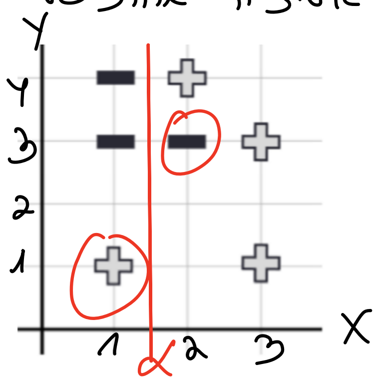
(a) (b) (c) (d) (e)



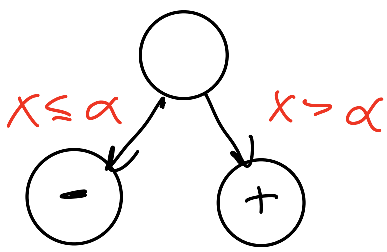
• כל התשובות מתבססות על כך שהאלגוריתם מגדיל את ההסתברות של נק' שסווגו לא נכון ומקטין לנק' שסווגו נכון, בנוסף לא ניתן למצוא מסווג על הdata הזה שיטעה רק על נק' אחת.

(a) מתאר התפלגות אפשרית אחרי איטרציה אחת של adaboost

ע"י



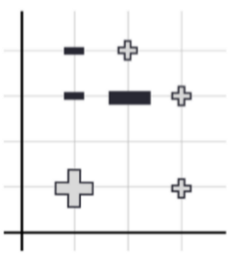
כאשר



• α יכול להיות שווה ל 1.5.

• כל הנקודות מסווגות נכון חוץ מהנק' (2,3) שמסווגת כ "-" והנק' (1,1) שמסווגת כ "+" ולכן נקבל אחרי איטרציה אחת ←

(a)



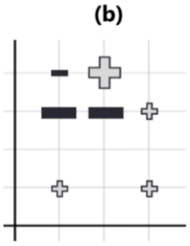
(b) לא מתאר התפלגות אפשרית.

adaboost בוחר בכל שלב מסווג באמצעות ERM עם השגיאה המינימלית

ראינו בסעיף (a) שקיים מסווג שטועה על 2 נקודות ולכן המסווג שטועה על 3 נק' כמו בדוגמא לא "יבחר" ע"י מסווג ERM.

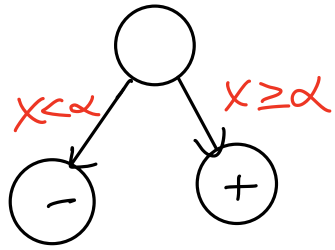
(נשים לב שכן ניתן למצוא מסווג שיגרום להתפלגות הנ"ל אבל adaboost לא יבחר בו)

(b)

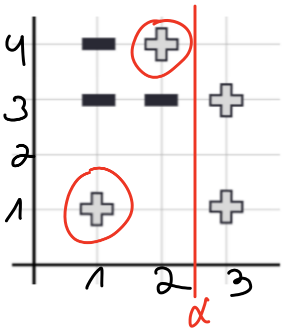


(c) מתאר התפלגות אפשרית אחרי איטרציה אחת של adaboost

ע"י

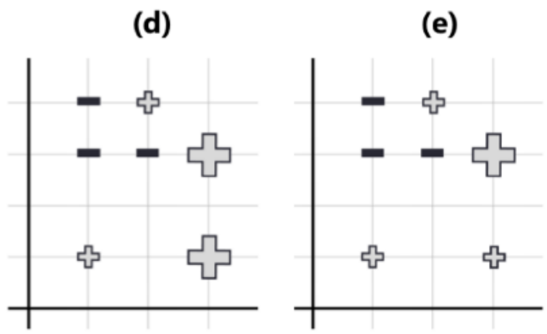


כאשר



- α יכול להיות שווה ל 2.5
- כל הנק' מסווגות נכון חוץ מהנקודות (2,4) ו (1,1) שמסווגות כ "-". לכן נקבל.

(d) + (e)



ניתן לראות שאי אפשר להפריד ע"י מסווג פשוט את הdata שיטעה רק על הנק' המוצגות פה. זאת כי אם יש מפריד שטועה עליהם הוא יטעה גם על עוד נקודות ונקודות אלה לא מוגדלות ולכן זאת לא התפלגות אפשרית בשני המקרים.

ב-(e) מוצג אפשרות לשימוש במסווג שטועה רק טעות אחת וכמו שכבר ראינו זה לא אפשרי על הdata שלנו
(כי אפשרות להפריד את הנק' ע"י מפריד תגרור טעויות על נק' נוספות עם ציר x וגם ציר y, פיצ'ר x ופיצ'ר y בהתאמה).