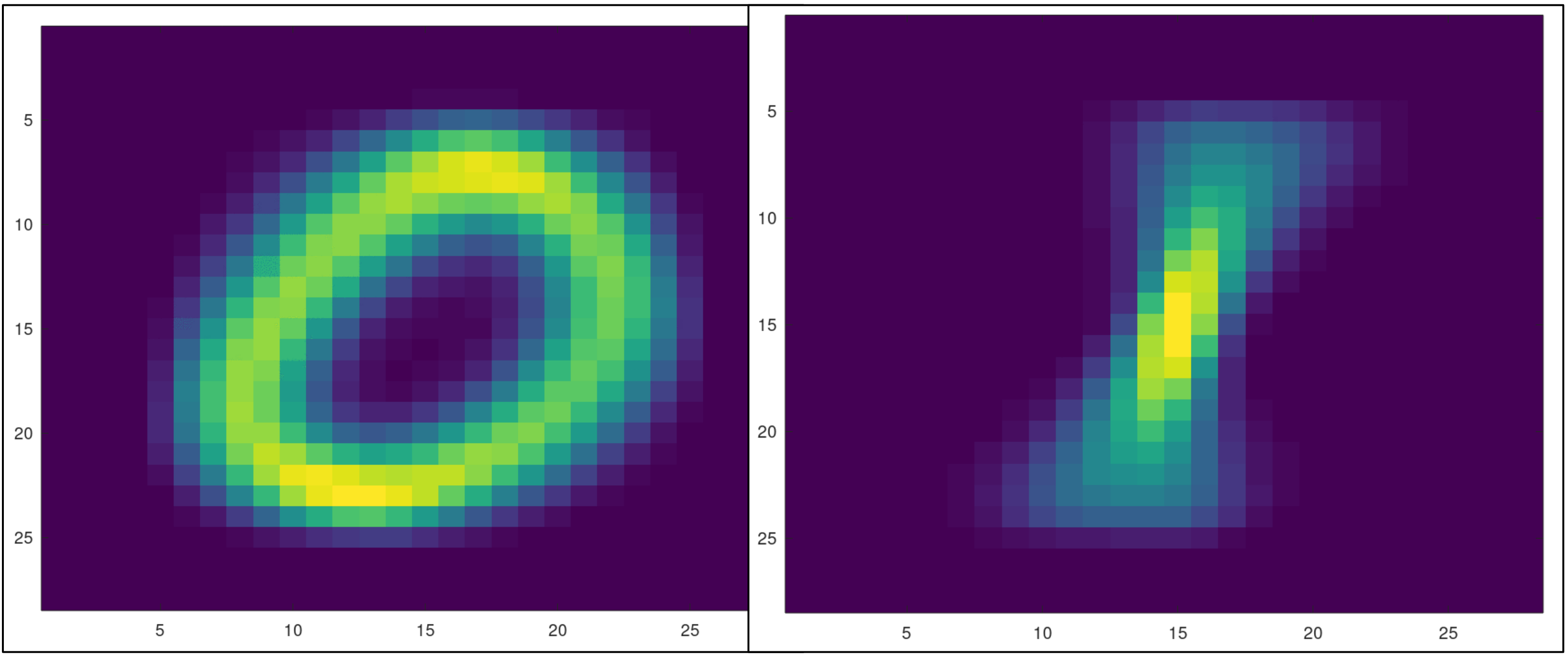
<u>שאלה 2</u>

א.

ב. ניתן להבחין כי הביצועים עבור הבחנה בין 0 ל 1 טובים יותר מהביצועים עבור הבחנה בין 3 ל 5. ניתן להסביר זאת ע"י כך שהספרות 0 ו 1 שונות לחלוטין, ואילו הספרות 3 ו 5 זהות בחלקן התחתון, ולכן יותר קשה להבחין ביניהן.
בנוסף, ניתן לראות כי ככל שגודל מדגם האימון גדל, כך השגיאה קטנה עבור 2 הבעיות. זאת משום שככל שמדגם האימון גדול יותר, כך הוא מיצג טוב יותר את ההתפלגות, ולכן האלגוריתם לומד טוב יותר את הבעיה.

ג.



נבחין כי מפות החום מפת החום של ppos מתאימה לספרה 1 ומפת החום של pneg מתאימה לספרה 0, זאת משום שהמפות מייצגות שבהינתן שהדוגמה מייצגת את אחת הספרות, מה ההסתברות שערך הקורדינטה הינו 1 (כהה). ואכן כפי שציפינו, בהינתן שהתמונה מייצגת את הספרה 0 או 1, כך ההסתברות של הקורדינטות שיוצרות את הספרה 0 או 1 בהתאמה להיות כהות, גבוהה יותר.

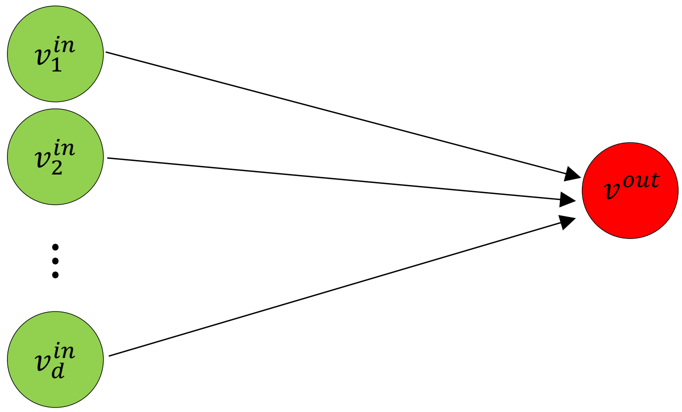
ד. אחוז השינוי עבור שני המקרים הינו 0. נבחין בהגדרת הפרדיקטור:

$$h(x) = 1 \Leftrightarrow \log(allpos) + \sum_{i=1}^{d} P[X(i) = x(i)|Y = 1] \geq \log(1 - allpos) + \sum_{i=1}^{d} P[X(i) = x(i)|Y = -1]$$

ניתן לראות כי ההשפעה של פרמטר ה allpos על אי השוויון זניחה, ולכן לא נמצא שינוי כלל.

שאלה 3

א. נגדיר גרף מכוון $G = \left(V = \{v_1^{in}, \dots, v_d^{in}, v^{out}\}, E = \{(v_i^{in}, v^{out}) : 1 \le i \le d\}\right)$.



טענה: $\mathcal{H}_{G,\sigma} \coloneqq \left\{f_w^{G,\sigma} \middle| w \in \mathbb{R}^{|E|}\right\} = \mathcal{H}_L^d \coloneqq \{h_w | w \in \mathbb{R}^d\}$.

הוכחה: $\supseteq$ תהי $h_w \in \mathcal{H}_L^d$.

ע"פ הגדרת h_w לכל $x \in \mathbb{R}^d$ מתקיים כי

$h_w(x) = sign(\langle w, x\rangle) = sign\left(\sum_{i=1}^d w(1)x(1)\right) = f_w^{G,\sigma}(x)$.

כלומר $h_w = f_w^{G,\sigma}$, ומכאן ש $h_w \in \mathcal{H}_{G,\sigma}$.

$\subseteq$ תהי $f_w^{G,\sigma} \in \mathcal{H}_{G,\sigma}$.

ע"פ הגדרת $f_w^{G,\sigma}$ לכל $x \in \mathbb{R}^d$ מתקיים כי

$f_w^{G,\sigma}(x) = sign\left(\sum_{i=1}^d w(1)x(1)\right) = sign(\langle w, x\rangle) = h_w(x)$.

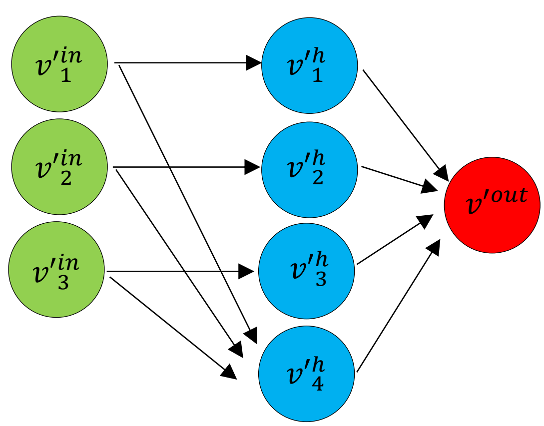
כלומר $f_w^{G,\sigma} = h_w$, ומכאן ש $f_w^{G,\sigma} \in \mathcal{H}_L^d$.

*נבחין כי ברשת לא קיימת hidden layer ומכאן ש σ אינה מופעלת כלל בחישוב $f_w^{G,\sigma}(x)$.

ב. נגדיר גרף מכוון $G' = (V', E')$, כאשר:

$V' = \left\{v'^{in}_1, v'^{in}_2, v'^{in}_3, v'^{h}_1, v'^{h}_2, v'^{h}_3, v'^{d}_4, v'^{out}\right\}$

$E' = \{(v'^{in}_i, v'^{h}_i), (v'^{in}_i, v'^{h}_4), (v'^{h}_j, v'^{out}) : 1 \le i \le 3, 1 \le j \le 4\}$



טענה: $\mathcal{H}_{G,\sigma} \subseteq \mathcal{H}_{G\prime,\sigma}$ עבור $d = 3$.

הוכחה: תהי $f_w^{G,\sigma} \in \mathcal{H}_{G,\sigma}$ כך ש $w = (w(1), w(2), w(3))$.

נסתכל על $f_{w\prime}^{G\prime,\sigma} \in \mathcal{H}_{G\prime,\sigma}$ כך ש $w' = (0, w(1), 0, w(2), 0, w(3), 0,0,0,1)$.

ע״פ הגדרת $f_{w\prime}^{G\prime,\sigma}$ לכל $x \in \mathbb{R}^3$ מתקיים כי

$$\begin{aligned} f_{w\prime}^{G\prime,\sigma}(x) &= sign(0 + 0 + 0 + 1 \cdot sign(w(1)x(1) + w(2)x(2) + w(3)x(3)) \\ &= sign(w(1)x(1) + w(2)x(2) + w(3)x(3)) = sign(\langle w, x\rangle) = h_w(x) \\ &= f_w^{G,\sigma}(x) \end{aligned}$$

כלומר $f_{w\prime}^{G\prime,\sigma} = f_w^{G,\sigma}$, ומכאן ש $f_w^{G,\sigma} \in \mathcal{H}_{G\prime,\sigma}$. בסה״כ הראינו כי $\mathcal{H}_{G,\sigma} \subseteq \mathcal{H}_{G\prime,\sigma}$, כנדרש.

טענה: $\mathcal{H}_{G,\sigma} \not\supseteq \mathcal{H}_{G\prime,\sigma}$ עבור $d = 3$.

הוכחה: נסתכל על $f_w^{G\prime,\sigma} \in \mathcal{H}_{G\prime,\sigma}$ כך ש $w = (1,0,1,0,1,0,1,1,1,0)$.

ע״פ הגדרת $f_w^{G\prime,\sigma}$ לכל $x \in \mathbb{R}^3$ מתקיים כי

$$f_w^{G\prime,\sigma}(x) = sign\big(1 \cdot sign\big(1 \cdot x(1)\big) + 1 \cdot sign\big(1 \cdot x(2)\big) + 1 \cdot sign\big(1 \cdot x(3)\big) + 0 \cdot sign(0 \cdot x(1) + 0 \cdot x(2) + 0 \cdot x(3))\big)\big) = sign\Big(sign\big(x(1)\big) + sign\big(x(2)\big) + sign\big(x(3)\big)\Big)$$

נניח בשלילה כי $f_w^{G\prime,\sigma} \in \mathcal{H}_{G,\sigma}$, כלומר קיים $w' \in \mathbb{R}^3$ כך שכל $x \in \mathbb{R}^3$ מתקיים כי

$$f_{w\prime}^{G,\sigma}(x) = sign\big(w'(1)x(1) + w'(2)x(2) + w'(3)x(3)\big)$$

$$\overset{(*)}{=} sign\Big(sign\big(x(1)\big) + sign\big(x(2)\big) + sign\big(x(3)\big)\Big)$$

נסתכל על $x_1 = (1,1,1),\ x_2 = (-3,1,1),\ x_3 = (1,-3,1),\ x_4 = (1,1,-3)$.

ע״פ הגדרת $f_{w\prime}^{G,\sigma}$ נקבל כי

$$f_{w\prime}^{G,\sigma}(x_1) = sign\big(w'(1) + w'(2) + w'(3)\big)$$

$$f_{w\prime}^{G,\sigma}(x_2) = sign\big(-3w'(1) + w'(2) + w'(3)\big)$$

$$f_{w\prime}^{G,\sigma}(x_3) = sign\big(w'(1) - 3w'(2) + w'(3)\big)$$

$$f_{w\prime}^{G,\sigma}(x_4) = sign\big(w'(1) + w'(2) - 3w'(3)\big)$$

ע״פ $(*)$ נקבל כי

$$w'(1) + w'(2) + w'(3) > 0\ , \qquad -3w'(1) + w'(2) + w'(3) > 0$$

$$w'(1) - 3w'(2) + w'(3) > 0\ , \qquad w'(1) + w'(2) - 3w'(3) > 0$$

נחבר את אי השוויונים ונקבל כי $0 > 0$. סתירה! מכאן ש $f_w^{G\prime,\sigma} \notin \mathcal{H}_{G,\sigma}$, ולכן $\mathcal{H}_{G,\sigma} \not\supseteq \mathcal{H}_{G\prime,\sigma}$ כנדרש.

<u>שאלה 4</u>

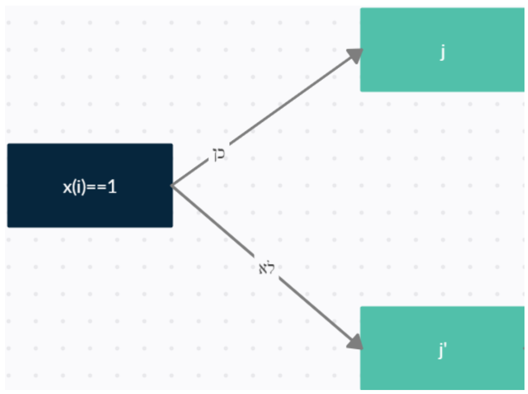
א. נניח בשלילה שקיים עץ החלטה T בעל עומק $d<2$ אשר מסווג באופן מושלם את המדגם הנ״ל.

<u>$d=0$:</u> אז T הינו מהצורה

עבור $j\in\{-1,1\}$.

- $j=-1$: אז הדוגמא $(0,0,0)$ מקבלת סיווג -1 , בסתירה להגדרת המדגם.
- $j=1$: אז הדוגמא $(0,1,0)$ מקבלת סיווג 1, בסתירה להגדרת המדגם.

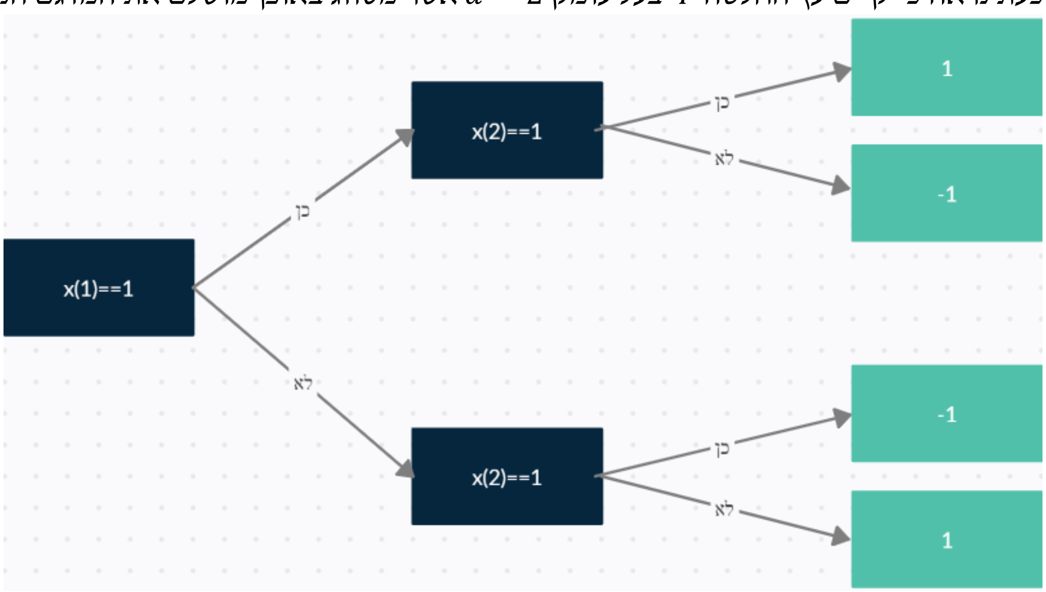
<u>$d=1$:</u> אז T הינו מהצורה



עבור $j,j'\in\{-1,1\},\ i\in\{1,2,3\}$.

- $i=1$: אז הדוגמאות $(1,0,0),(1,1,0)$ מקבלות את אותו סיווג j, בסתירה להגדרת המדגם.
- $i=2$: אז הדוגמאות $(0,1,0),(1,1,0)$ מקבלות את אותו סיווג j, בסתירה להגדרת המדגם.
- $i=3$: אז הדוגמאות $(0,0,0),(1,0,0)$ מקבלות את אותו סיווג j', בסתירה להגדרת המדגם.

כעת נראה כי קיים עץ החלטה T בעל עומק $d=2$ אשר מסווג באופן מושלם את המדגם הנ״ל:



$\Leftarrow$ תהי דוגמא x מהמדגם כך ש $x(1) = x(2) = j$ עבור $j \in \{0,1\}$.

- $j = 0$: אז הסיווג מתקדם במסלול של לא->לא->1 כנדרש.
- $j = 0$: אז הסיווג מתקדם במסלול של כן->כן->1 כנדרש.

$\Rightarrow$ תהי דוגמא x מהמדגם כך ש $j' = x(2) \neq x(1) = j$ כך ש $j = 0, j' = 1$ או $j = 1, j' = 0$.

- $j = 0, j' = 1$: אז הסיווג מתקדם במסלול של לא->כן->1- כנדרש.
- $j = 1, j' = 0$: אז הסיווג מתקדם במסלול של כן->לא->1- כנדרש.

ב. נסמן:

n = מספר הפעמים שכל אחת מהדוגמאות מופיעה במדגם.

S = המדגם הראשוני.

$A = \{1,2,3\}$.

נבחן את $err_{after}(S,i)$ לכל $i \in A$:

<u>$i = 1$:</u>

$$p_1^0 = \frac{num\ of\ examples\ from\ the\ form\ \big(0, x(2), x(3)\big)}{num\ of\ examples} = \frac{4n}{8n} = \frac{1}{2}$$
$$q_1^0 = \frac{num\ of\ examples\ from\ the\ form\ (0,0,x(3))}{num\ of\ examples\ from\ the\ form\ \big(0, x(2), x(3)\big)} = \frac{2n}{4n} = \frac{1}{2}$$
$$q_1^1 = \frac{num\ of\ examples\ from\ the\ form\ (1,1,x(3))}{num\ of\ examples\ from\ the\ form\ \big(1, x(2), x(3)\big)} = \frac{2n}{4n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,1) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$.

<u>$i = 2$:</u>

$$p_2^0 = \frac{num\ of\ examples\ from\ the\ form\ \big(x(1), 0, x(3)\big)}{num\ of\ examples} = \frac{4n}{8n} = \frac{1}{2}$$
$$q_2^0 = \frac{num\ of\ examples\ from\ the\ form\ (0,0,x(3))}{num\ of\ examples\ from\ the\ form\ \big(x(1), 0, x(3)\big)} = \frac{2n}{4n} = \frac{1}{2}$$
$$q_2^1 = \frac{num\ of\ examples\ from\ the\ form\ (1,1,x(3))}{num\ of\ examples\ from\ the\ form\ \big(x(1), 1, x(3)\big)} = \frac{2n}{4n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,1) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$.

<u>$i = 3$:</u>

$$p_3^0 = \frac{num\ of\ examples\ from\ the\ form\ (x(1), x(2), 0)}{num\ of\ examples} = \frac{4n}{8n} = \frac{1}{2}$$
$$q_3^0 = \frac{num\ of\ examples\ from\ \{(0,0,0), (1,1,0)\}}{num\ of\ examples\ from\ the\ form\ (x(1), x(2), 0)} = \frac{2n}{4n} = \frac{1}{2}$$
$$q_3^1 = \frac{num\ of\ examples\ from\ \{(0,0,1), (1,1,1)\}}{num\ of\ examples\ from\ the\ form\ (x(1), x(2), 1)} = \frac{2n}{4n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,1) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$.

הראינו כי $err_{after}(S,i) = \frac{1}{2}$ לכל $i \in A$, ומכאן ש $argmax_{i \in A} Gain(S,i) = \{1,2,3\}$.
נסתכל על המקרה בו האלגוריתם בוחר $j = 3$ עבור המבחן הראשון, ונביט בקריאות הרקורסיביות:

<u>ענף שמאל:</u>
$S_0 = \{(x,y) \in S | x(3) = 0\}$ = המדגם בקריאה הרקורסיבית הנ"ל.
$A = \{1,2\}$.

נבחן את $err_{after}(S_0, i)$ לכל $i \in A$:
<u>$i = 1$:</u>

$$p_1^0 = \frac{num\ of\ examples\ from\ the\ form\ (0, x(2), 0)}{num\ of\ examples\ from\ the\ form\ (x(1), x(2), 0)} = \frac{2n}{4n} = \frac{1}{2}$$

$$q_1^0 = \frac{num\ of\ examples\ from\ the\ form\ (0, x(2), 0)\ and\ x(1) = x(2)}{num\ of\ examples\ from\ the\ form\ (0, x(2), 0)} = \frac{n}{2n} = \frac{1}{2}$$

$$q_1^1 = \frac{num\ of\ examples\ from\ the\ form\ (1, x(2), 0)\ and\ x(1) = x(2)}{num\ of\ examples\ from\ the\ form\ (1, x(2), 0)} = \frac{n}{2n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$.
<u>$i = 2$:</u>

$$p_3^0 = \frac{num\ of\ examples\ from\ the\ form\ (x(1), 0, 0)}{num\ of\ examples\ from\ the\ form\ (x(1), x(2), 0)} = \frac{2n}{4n} = \frac{1}{2}$$

$$q_3^0 = \frac{num\ of\ examples\ from\ the\ form\ (x(1), 0, 0)\ and\ x(1) = x(2)}{num\ of\ examples\ from\ the\ form\ (x(1), 0, 0)} = \frac{n}{2n} = \frac{1}{2}$$

$$q_3^1 = \frac{num\ of\ examples\ from\ the\ form\ (x(1), 1, 0)\ and\ x(1) = x(2)}{num\ of\ examples\ from\ the\ form\ (x(1), 1, 0)} = \frac{n}{2n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,2) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$.

הראינו כי $err_{after}(S_0,i) = \frac{1}{2}$ לכל $i \in A$, ומכאן ש $argmax_{i \in A} Gain(S,i) = \{1,2\}$.
נסתכל על המקרה בו האלגוריתם בוחר $j = 2$ עבור המבחן השני.

<u>ענף ימין:</u>
$S_1 = \{(x,y) \in S | x(3) = 1\}$ = המדגם בקריאה הרקורסיבית הנ"ל.
$A = \{1,2\}$.

נבחן את $err_{after}(S_1, i)$ לכל $i \in A$:
<u>$i = 1$:</u>

$$p_1^0 = \frac{num\ of\ examples\ from\ the\ form\ (0, x(2), 1)}{num\ of\ examples\ from\ the\ form\ (x(1), x(2), 1)} = \frac{2n}{4n} = \frac{1}{2}$$

$$q_1^0 = \frac{num\ of\ examples\ from\ the\ form\ (0, x(2), 1)\ and\ x(1) = x(2)}{num\ of\ examples\ from\ the\ form\ (0, x(2), 1)} = \frac{n}{2n} = \frac{1}{2}$$

$$q_1^1 = \frac{num\ of\ examples\ from\ the\ form\ (1, x(2), 1)\ and\ x(1) = x(2)}{num\ of\ examples\ from\ the\ form\ (1, x(2), 1)} = \frac{n}{2n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,1) = \frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$.

<u>$i=2$:</u>

$$p_3^0 = \frac{num\ of\ examples\ from\ the\ form\ (x(1),0,1)}{num\ of\ examples\ from\ the\ form\ (x(1),x(2),1)} = \frac{2n}{4n} = \frac{1}{2}$$

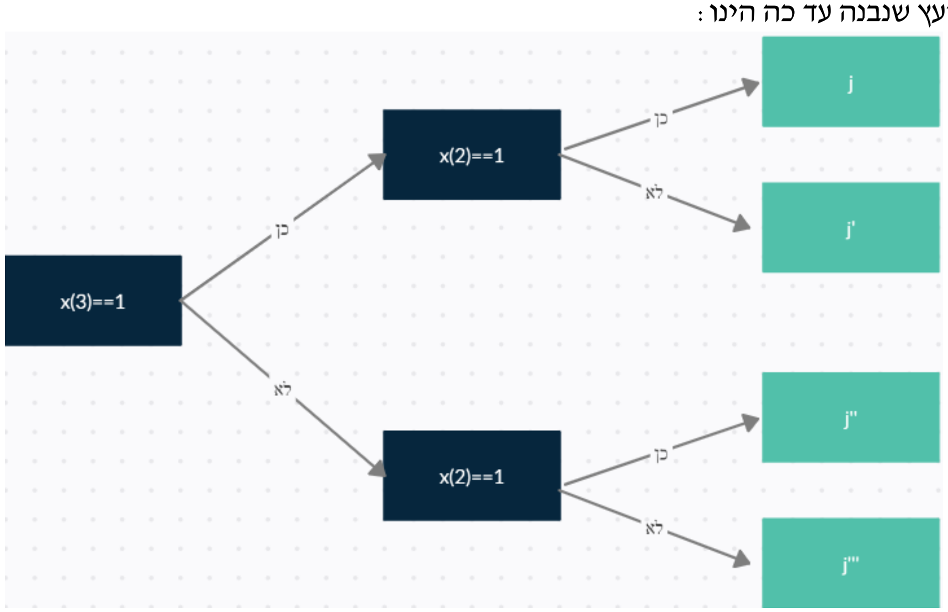
$$q_3^0 = \frac{num\ of\ examples\ from\ the\ form\ (x(1),0,1)\ and\ x(1)=x(2)}{num\ of\ examples\ from\ the\ form\ (x(1),0,1)} = \frac{n}{2n} = \frac{1}{2}$$

$$q_3^1 = \frac{num\ of\ examples\ from\ the\ form\ (x(1),1,1)\ and\ x(1)=x(2)}{num\ of\ examples\ from\ the\ form\ (x(1),1,1)} = \frac{n}{2n} = \frac{1}{2}$$

מכיוון ש $err\left(\frac{1}{2}\right) = Entorpy\left(\frac{1}{2}\right) = Gini\left(\frac{1}{2}\right) = \frac{1}{2}$ נקבל כי $err_{after}(S,2) = \frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$.

הראינו כי $err_{after}(S_1,i) = \frac{1}{2}$ לכל $i\in A$, ומכאן ש $argmax_{i\in A}Gain(S,i) = \{1,2\}$.
נסתכל על המקרה בו האלגוריתם בוחר $j=2$ עבור המבחן השני.

בסה"כ העץ שנבנה עד כה הינו:



עבור $j,j',j'',j'''\in\{-1,1\}$.

- נסתכל על דוגמאות מהצורה $(x(1),1,1)$. אם $j=1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(0,1,1)$. אחרת, אם $j=-1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(1,1,1)$.
- נסתכל על דוגמאות מהצורה $(x(1),0,1)$. אם $j'=1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(1,0,1)$. אחרת, אם $j'=-1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(0,0,1)$.
- נסתכל על דוגמאות מהצורה $(x(1),1,0)$. אם $j''=1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(0,1,0)$. אחרת, אם $j''=-1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(1,1,0)$.
- נסתכל על דוגמאות מהצורה $(x(1),0,0)$. אם $j'''=1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(1,0,0)$. אחרת, אם $j'''=-1$ אזי העץ טועה עבור n הדוגמאות מהצורה $(0,0,0)$.

בסה"כ הראינו כי העץ טועה עבור $4n$ מתוך $8n$ הדוגמאות, ולכן בעל שגיאה של 50%.

שאלה 5

א. קיים פתרון יחיד $w=(X^TX)^{-1}X^Ty$ לבעיית squared loss על המדגם הנתון, אז X^TX הפיכה.
$X^TX\in\mathbb{R}^{d\times d}$ הפיכה אז $rank(XX^T)=d$.
ע״פ משפט מאלגברה ליניארית מתקיים $rank(X^TX)\le \min\{rank(X^T),\ rank(X)\}$.
מכאן, $rank(X^TX)\le rank(X)$, כלומר $d\le rank(X)$.
מכיוון ש $X\in\mathbb{R}^{m\times d}$ מתקיים כי $rank(X)\le \min\{m,d\}$, ומכאן $rank(X)\le d$.
$rank(X)\le d$ וגם $d\le rank(X)$ ונקבל בסה״כ כי $\boldsymbol{rank(X)=d}$.

ב. עבור w הפתרון האופטימלי עבור S, קבוצת הפתרונות עבור S' הינה $A=\left\{\begin{pmatrix}w\\a\end{pmatrix}:\ a\in\mathbb{R}\right\}$.

טענה ראשית: $u'\in\mathbb{R}^{d+1}$ פתרון אופטימלי עבור S' $\Leftrightarrow$ $u'\in A$.

טענת עזר: u הינו פתרון אופטימלי עבור S $\Leftrightarrow$ $u'=\begin{pmatrix}u\\a\end{pmatrix}$ הינו פתרון אופטימלי עבור S' לכל $a\in\mathbb{R}$.

אבחנה: מכיוון שלכל $1\le i\le m$ מתקיים $x_i'(d+1)=0$, נקבל כי לכל $u'=\begin{pmatrix}u\\a\end{pmatrix}$ כאשר $a\in\mathbb{R}, u\in\mathbb{R}^d$ מתקיים כי $f(u)=\sum_{i=1}^m(\langle u,x_i\rangle-y_i)^2=\sum_{i=1}^m(\langle u',x_i'\rangle-y_i)^2=f'(u')$.

הוכחת טענה ראשית: $\Leftarrow$ יהי $\begin{pmatrix}u\\a\end{pmatrix}==$ שu' כאשר $a\in\mathbb{R}, u\in\mathbb{R}^d$ פתרון אופטימלי עבור S', ונניח בשלילה ש $u'\notin A$. כלומר קיים $1\le i\le d$ כך ש $w(i)\ne u(i)$. ע״פ טענת העזר u הינו פתרון אופטימלי עבור S, ומכך ש $u\ne w$ נקבל סתירה לכך ש w הינו הפתרון האופטימלי היחיד עבור S.
$\Rightarrow$ יהי $u'\in A$ ונניח בשלילה ש u' אינו פתרון אופטימלי עבור S'. מכיוון ש $u'=\begin{pmatrix}w\\a\end{pmatrix}$ עבור $a\in\mathbb{R}$ כלשהו, נקבל ע״פ טענת העזר כי w אינו פתרון אופטימלי עבור S, בסתירה לכך ש w הינו פתרון אופטימלי עבור S.

הוכחת טענת עזר: $\Leftarrow$ יהי u פתרון אופטימלי עבור S, ונניח בשלילה שקיים $a\in\mathbb{R}$ כך ש $u'=\begin{pmatrix}u\\a\end{pmatrix}$ אינו פתרון אופטימלי עבור S'. כלומר, קיים $v'=\begin{pmatrix}v\\b\end{pmatrix}$ כאשר $b\in\mathbb{R}, v\in\mathbb{R}^d$ כך ש $f'(v')<f'(u')$.
ע״פ האבחנה נקבל כי $f(v)<f(u)$ בסתירה לכך ש u הינו פתרון אופטימלי עבור S.
$\Rightarrow$ יהי $u'=\begin{pmatrix}u\\a\end{pmatrix}$ כאשר $a\in\mathbb{R}, u\in\mathbb{R}^d$ פתרון אופטימלי עבור S', ונניח בשלילה כי u אינו פתרון אופטימלי עבור S. כלומר, קיים v כך ש $f(v)<f(u)$. ע״פ האבחנה נקבל כי עבור $v'=\begin{pmatrix}v\\a\end{pmatrix}$ מתקיים כי $f'(v')<f'(u')$ בסתירה לכך ש u' הינו פתרון אופטימלי עבור S'.

שאלה 6

א.

```python
import numpy as np

X = np.array([[1, 2, 3, 4], [3, 4, 1, -5], [-10, 1, 4, 6]])
d = 4
k = 2
distortion = 0

XT = np.transpose(X)
XTX = np.matmul(XT, X)
[w, v] = np.linalg.eig(XTX) # (X^T)X eigenvalues and eigenvectors
w.sort() # sorting (X^T)X eigenvalues

for i in range(d-k): # last d-k eigenvalues of (X^T)X, 1<=i<=d-k
  distortion = distortion + w[i] # adding the i'th eigenvalue of (X^T)X

print(f'The distortion is {distortion}')
```

```
The distortion is 21.272812377486858
```

ב.

```python
import numpy as np

X = np.array([[1, 2, 3, 4], [3, 4, 1, -5], [-10, 1, 4, 6]])
d = 4
k = 2
U = np.zeros((d, k)) # dXk matrix
orthonormal = True

XT = np.transpose(X)
XTX = np.matmul(XT, X)
[w, v] = np.linalg.eig(XTX) # (X^T)X eigenvalues and eigenvectors
ws = np.sort(w) # (X^T)X sorted eigenvalues

for i in range((d-k), d): # top k eigenvalues of (X^T)X, 1<=i<=k
  index = np.argmax(w == ws[i]) # index of the i'th eigenvalue at w
  U[:, (d-k)-i] = np.transpose(v[:, index]) # insert to U the i'th eigenvector

print(f'U^T = {np.transpose(U)}') # U^T

for i in range(k):
  if(np.round(np.linalg.norm(UT[i])) != 1): # rounding is needed for numeric reasons
    orthonormal = False # the norm of the i'th eigenvector of U^T is not 1!
  for j in range(i+1, k):
    if(np.round(np.dot(np.transpose(UT[i]), UT[j])) != 0): # rounding is needed for numeric reasons
      orthonormal = False # the i'th and the j'th eigenvectors of U^T are not orthogonal!

if(orthonormal):
  print('U^T is Orthonormal!')
else:
  print('U^T is Not Orthonormal!')
```

```
U^T = [[-0.12805216 -0.79647712 -0.58328883  0.0948735 ]
 [ 0.74704792  0.01643167 -0.28415459 -0.60075418]]
U^T is Orthonormal!
```

ג.

```python
distortion2 = 0

for i in range(3):
  print(f'x_{i+1} = {X[i]}')
  restoredX = np.matmul(U,np.matmul(UT, X[i])) # restored x_i = U*U^T*x_i
  print(f'restored x_{i+1} = {restoredX}')
  diffNorm = np.linalg.norm(np.subtract(X[i], restoredX)) # diffNorm = ||x_i - restored x_i||
  distortion2 = distortion2 + diffNorm**2 # distortion2 = distortion2 + diffNorm^2

print(f'\ndistortion (= {distortion}) and distortion2 (= {distortion2})\n \
      are identical taking into account numerical errors, as expected!\n \
      Because the formula for calculation distortion was calculated\n \
      according to U^T as optimal solution.')
```

```
x_1 = [1 2 3 4]
restored x_1 = [-1.45351109  2.42153483  2.50661114  1.19391862]
x_2 = [ 3  4  1 -5]
restored x_2 = [ 4.34761613  3.76846767  1.27099889 -3.45873102]
x_3 = [-10   1   4   6]
restored x_3 = [-8.94650184  0.81899973  4.21185323  7.20488616]

distortion (= 21.272812377486858) and distortion2 (= 21.272812377486886)
       are identical taking into account numerical errors, as expected!
       Because the formula for calculation distortion was calculated
       according to U^T as optimal solution.
```

שאלה 7

א. **לא מקיים אקסיומת SI.**

לדוגמא עבור $S = \{0,\ 2r\} \subseteq \mathbb{R}^1$ כאשר פונקציית המרחק הינה המרחק האוקלידי הסטנדרטי.
$F(S,\rho) = \big\{\{0\},\{2r\}\big\}$ מכיוון שכבר בתחילת האלגוריתם המרחק בין כל שני קלאסטרים הינו גדול מ r, אולם $F\left(S,\frac{1}{4}\rho\right) = \big\{\{0,2r\}\big\}$ מכיוון שבתחילת האלגוריתם המרחק בין שני הקלאסטרים הינו $\frac{1}{2}r$, ורק לאחר האיחוד שלהם תנאי העצירה מתקיים.

ב. **מקיים אקסיומת RI.**

יהי מדגם סופי $S \subset X$ וחלוקה $C = \{C_1, \dots, C_k\}$ של S.
נגדיר את המטריקה $\rho: X \times X \to \mathbb{R}_+$ הבאה:

$$\rho(x,x') = \begin{cases} 0, & x = x' \\ r, & x \neq x' \in C_i, 1 \leq i \leq k \\ 2r, & x \in C_i, x' \in C_j, 1 \leq i \neq j \leq k \end{cases}$$

טענה: ρ הינה מטריקה חוקית
הוכחה:

- לכל $x, y \in X$ מתקיים כי $\rho(x,y) \geq 0$
- לכל $x \in X$ מתקיים כי $\rho(x,x) = 0$
- לכל $x, y \in X$ מתקיים כי $\rho(x,y) = \rho(y,x)$
- לכל $x, y, z \in X$ שונים מתקיים כי $\rho(x,y) \leq \rho(x,z) + \rho(y,z)$ מכיוון ש $\rho(x,y) \leq 2r$ וגם $2r \leq \rho(x,z) + \rho(y,z)$.

טענה: $F(S,\rho) = C$
הוכחה: $\supseteq$ תהי $C_i \in C$ עבור $1 \leq i \leq k$ ונניח בשלילה $C_i \notin F(S,\rho)$.
I. קיים $C^* \in F(S,\rho)$ כך ש C_i מוכל ממש ב C^*:
מתקיים כי קיימים $x \in C_i, x' \in C_j$ כאשר $i \neq j$ כך ש $x, x' \in C^*$.
$\rho(x,x') = 2r$, ומכיוון שהאלגוריתם הינו single-linakge נבין כי החיבור בין הקלאסטרים של x, x' התבצע כשהמרחק המינימלי בין שני קלאסטרים היה גדול מ r בסתירה להגדרת האלגוריתם.
II. קיים $C^* \in F(S,\rho)$ כך ש C^* מוכל ממש ב C_i:
מתקיים כי קיימים $x, x' \in C_i$ כך ש $x \in C^*$ וגם $x' \notin C^*$.
$\rho(x,x') = r$, ולכן נבין כי כאשר האלגוריתם הפסיק לרוץ, היו שני קלאסטרים שהמרחק המינימלי ביניהם היה לכל היותר r, בסתירה להגדרת האלגוריתם.
$\subseteq$ תהי $C^* \in F(S,\rho)$ ונניח בשלילה כי $C^* \notin C$.
I. קיים $C_i \in C$ כך ש C^* מוכל ממש ב C_i: זהה למקרה II בכיוון הקודם.
II. קיים $C_i \in C$ כך ש C_i מוכל ממש ב C^*: זהה למקרה I בכיוון הקודם.