



לא לשכוח להפעיל הקלטה!

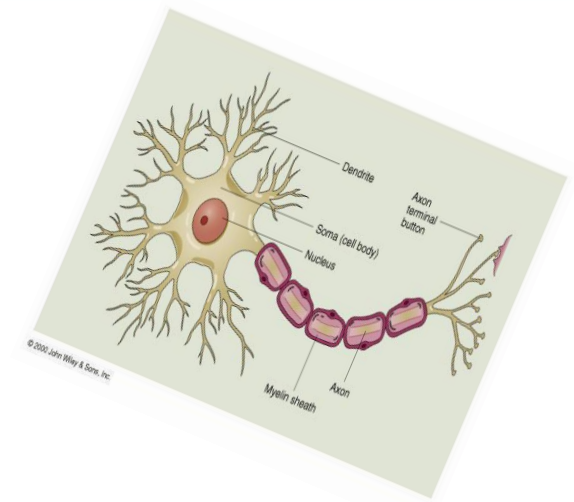
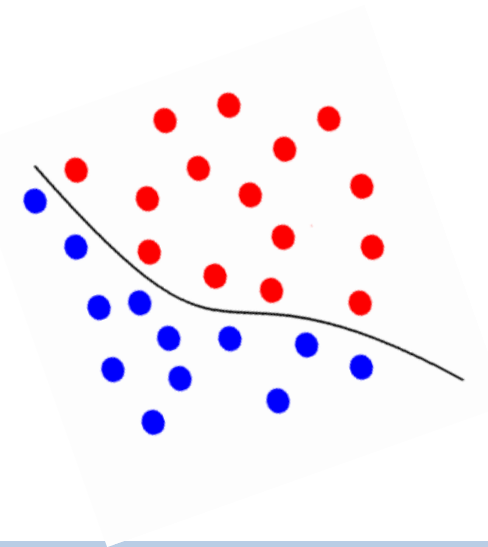
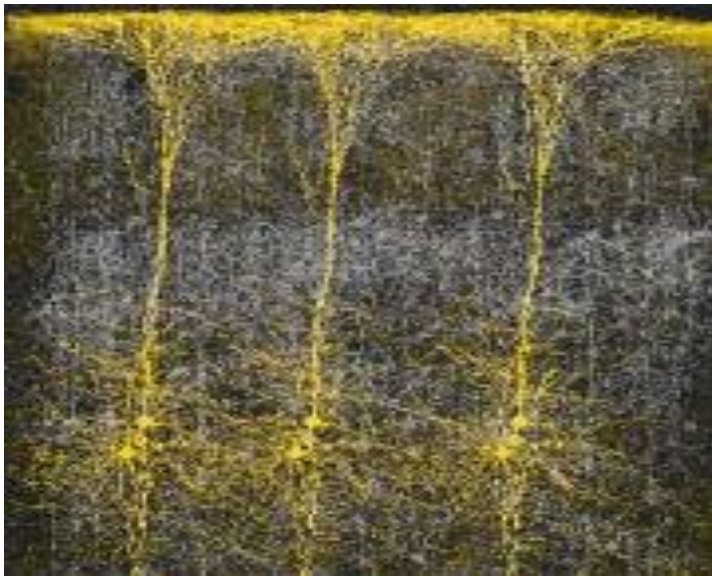
✓ מבנה ומימדי רשתות

✓ רשת Feed forward

✓ פונקציית מעבר

✓ Back propagation

מבנה וחישובים ברשתות נוירונים

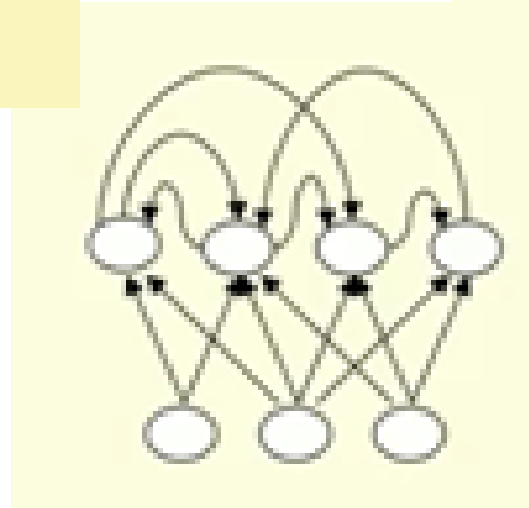
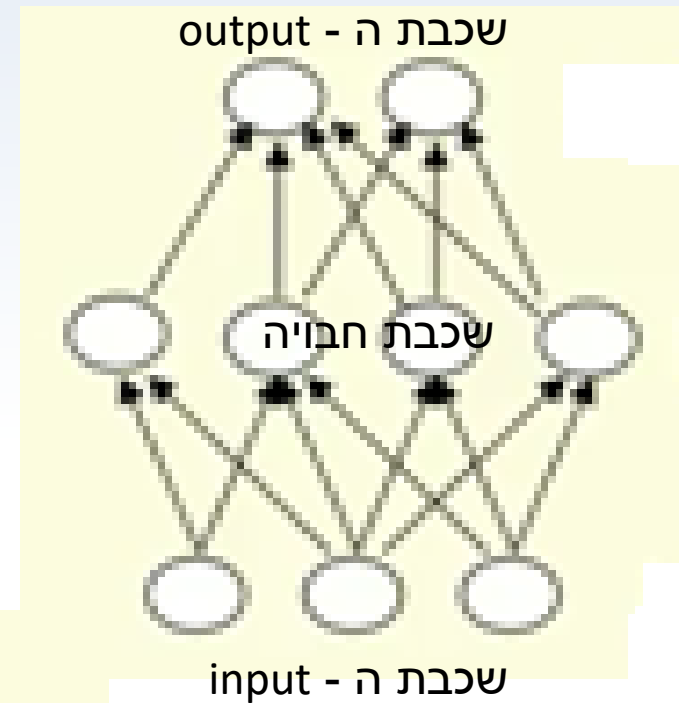
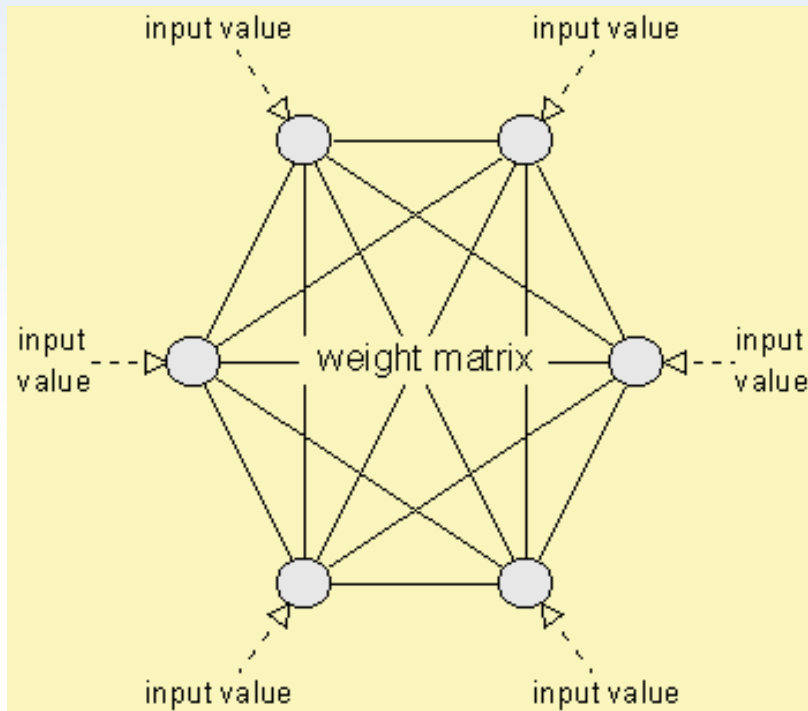


מבנה הרשת

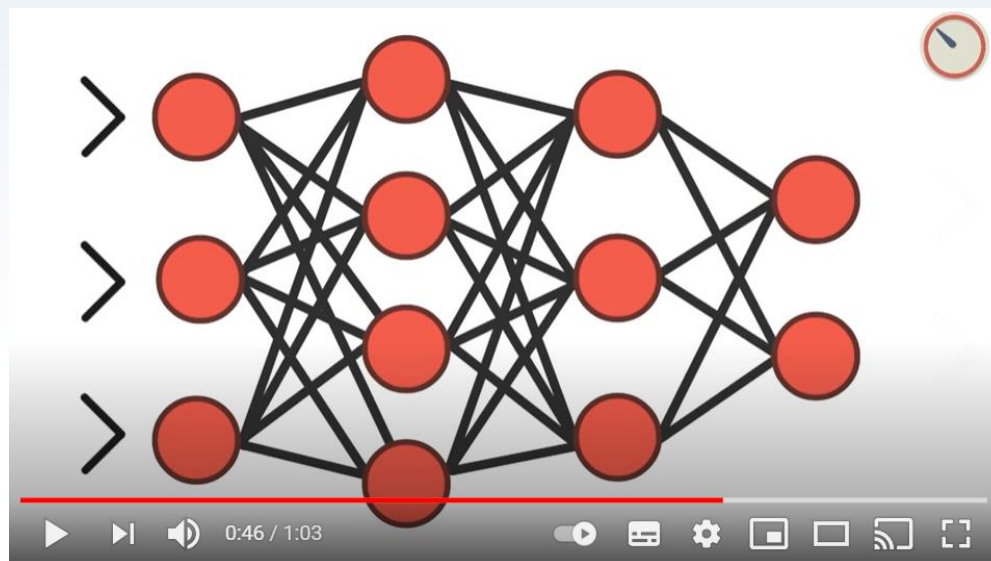
ארכיטקטורה של רשתות

- כל רשת מורכבת מנוירונים המקבלים קלט משוקלל, מעדכנים את הסף ומפעילים פונקציית מעבר.
- הארכיטקטורה של הרשת מורכבת מנוירונים רבים המחוברים ביניהם בקשרים סינפטיים.
- קיימים מספר סוגים של ארכיטקטורה, למשל:
 - Feed-forward networks –
 - Recurrent networks –
 - Symmetrically connected networks –

ארכיטקטורה של רשתות

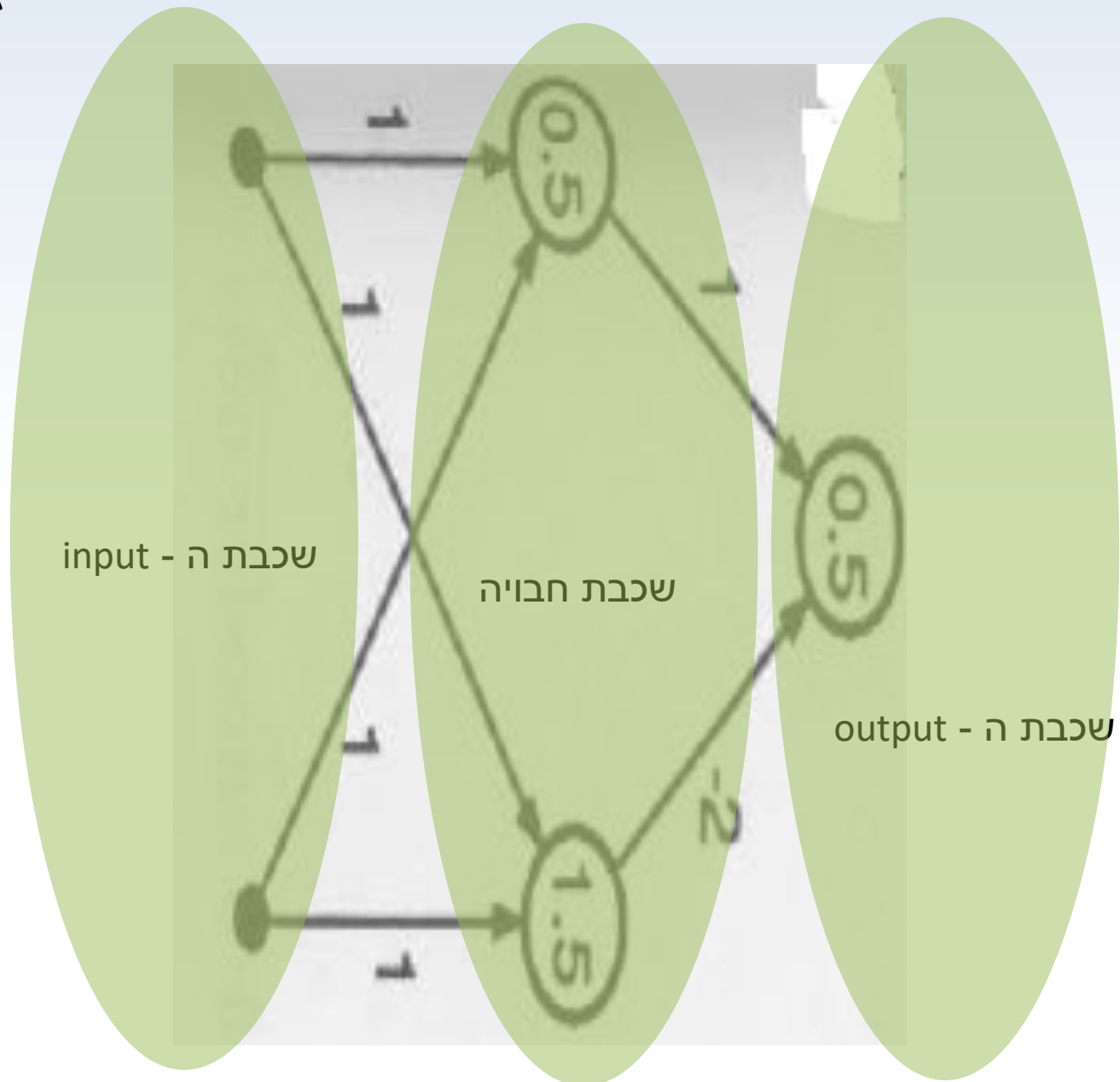


נצפה בסרטון



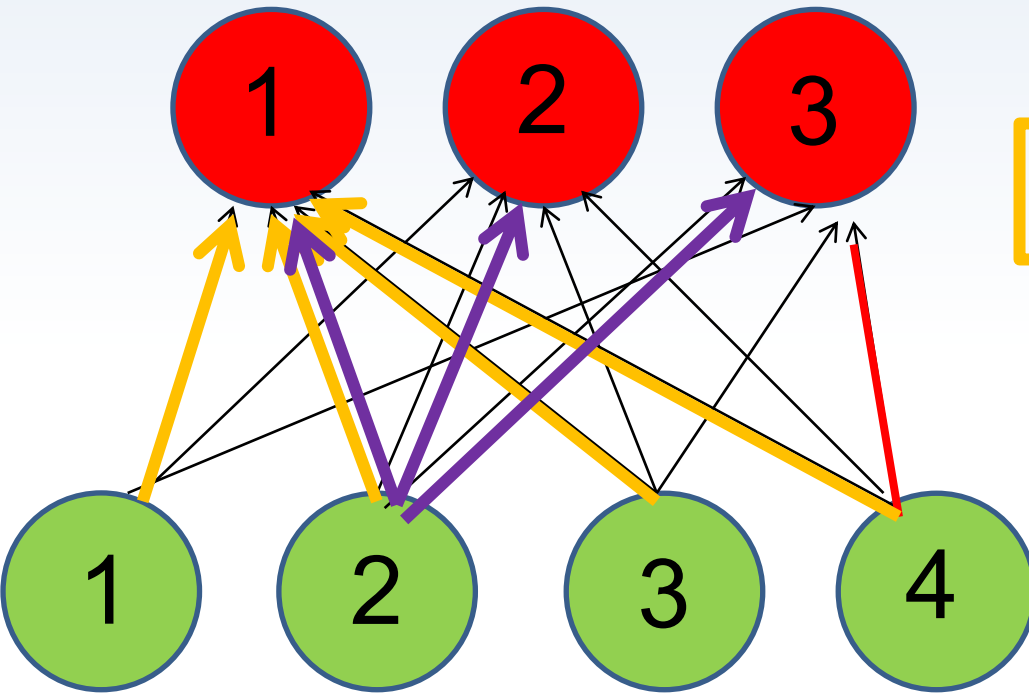
<https://www.youtube.com/watch?v=rEDzUT3ymw4>

XOR



מטריצת הקשרים ברשת (ff) Feed-Forward

פלט



קלט

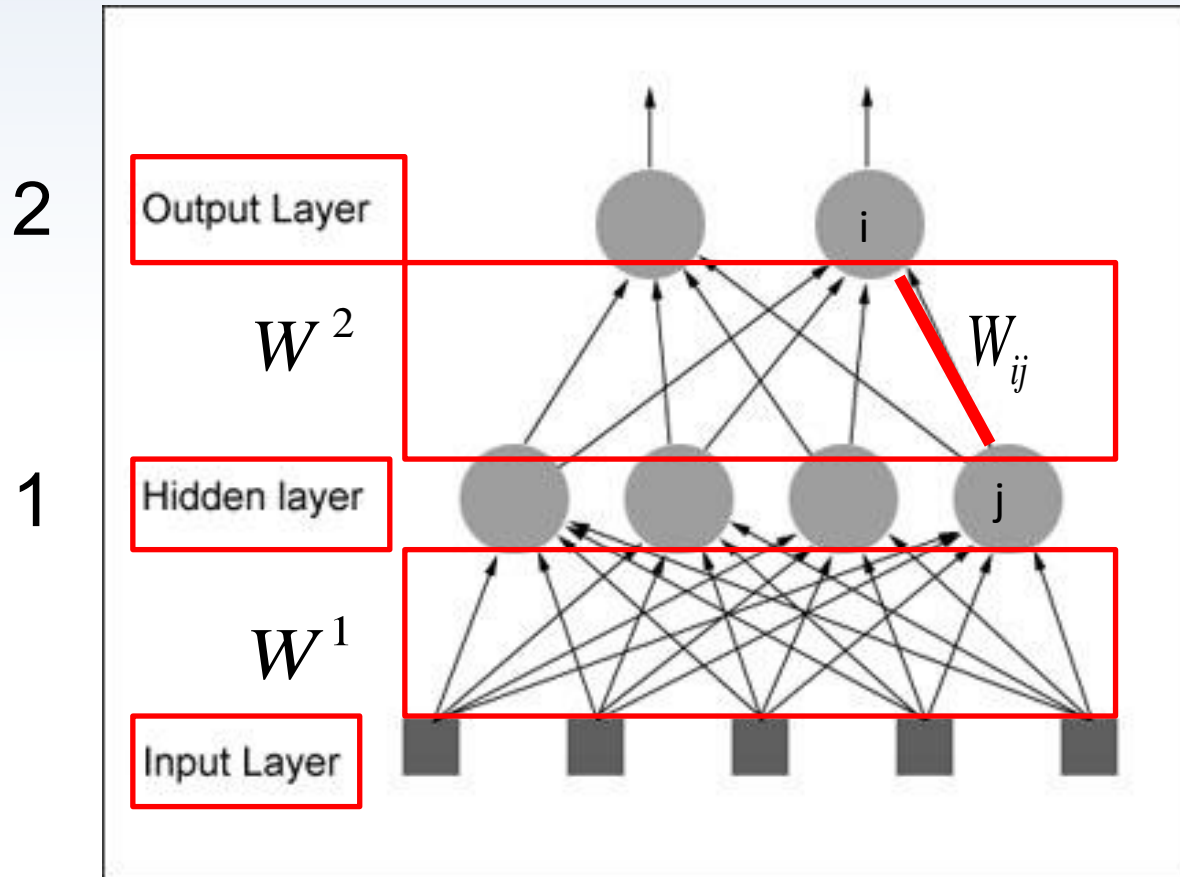
W_{11}	W_{12}	W_{13}	W_{14}
W_{21}	W_{22}	W_{23}	W_{24}
W_{31}	W_{32}	W_{33}	W_{34}

מקלט לפלט

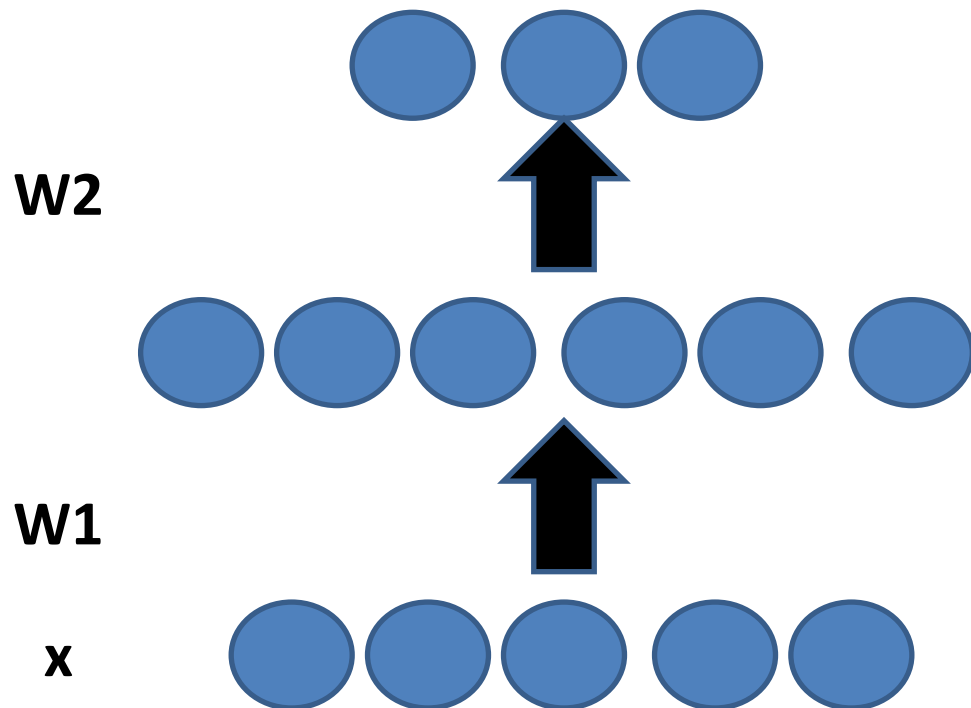
מימדי המטריצה: 3×4
3 ניורונים בפלט
4 ניורונים בקלט

רשת ff עם יותר משכבה אחת

רשת Feed-forward

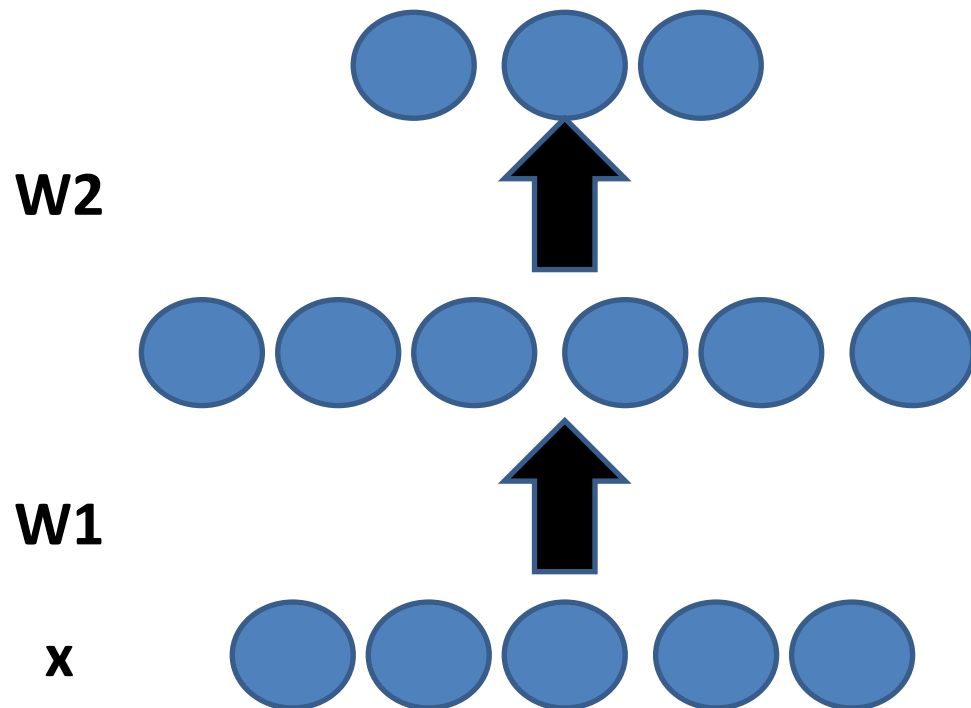


תרגול - חישובי רשת נוירונים כמכפלת וקטורים ומטריצות



- מהם מימדי:
– $w1$?
– $w2$?
- כמה נוירונים בשכבת הקלט?
- כמה נוירונים בשכבת הפלט?

תרגול - חישובי רשת נוירונים כמכפלת וקטורים ומטריצות



• מהם מימדי:

$6 * 5$? $W1$ –

$3 * 6$? $W2$ –

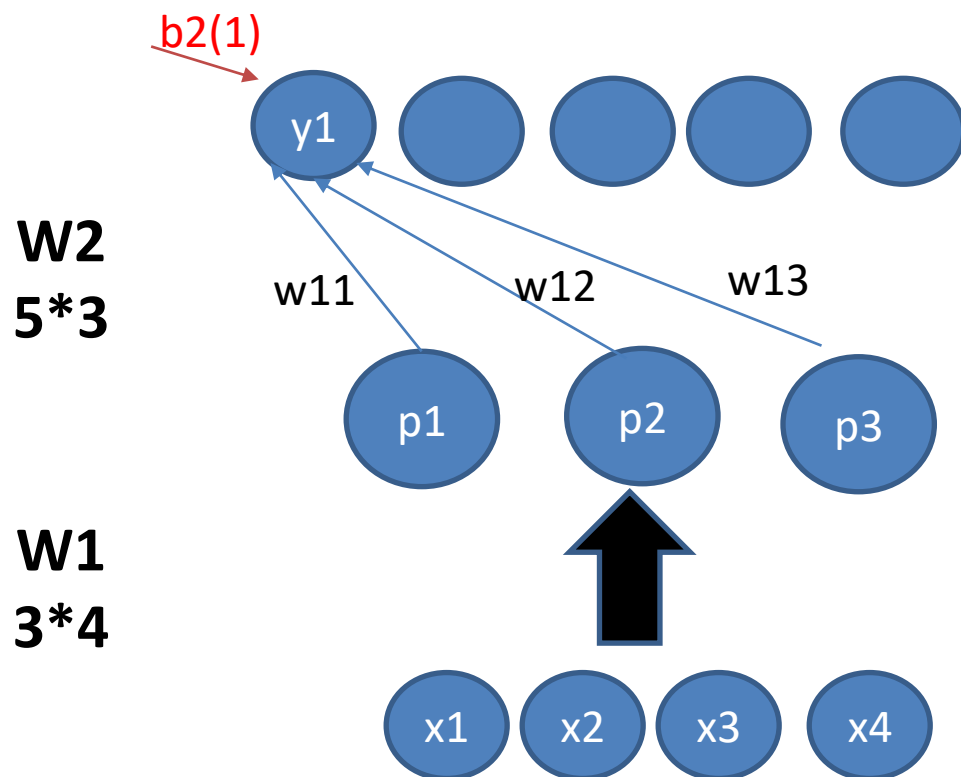
מקלט לפלט

כמה נוירונים בשכבת הקלט? **5**

כמה נוירונים בשכבת הפלט? **3**

חישובי רשת נוירונים כמכפלת וקטורים ומטריצות

נתונה הרשת:



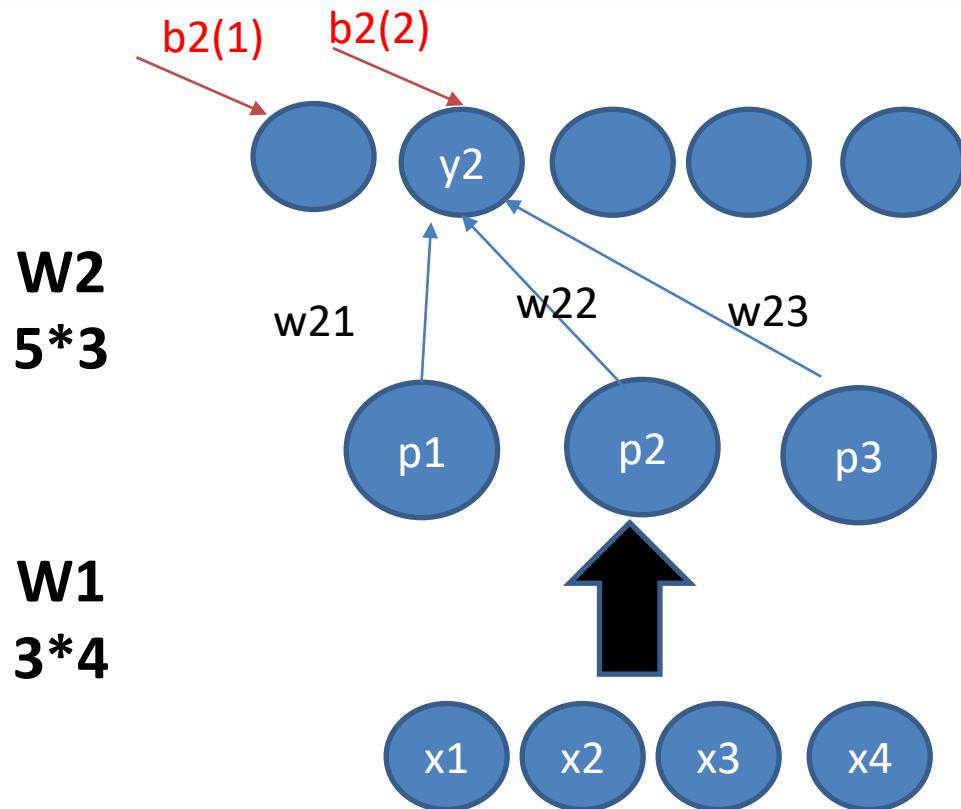
4 נוירוני קלט
3 נוירוני ביניים
5 נוירוני פלט.

$$y_1 = W_{11} * p_1 + W_{12} * p_2 + W_{13} * p_3 + b_{2(1)}$$

איך נחשב y_1 ?

חישובי רשת נוירונים כמכפלת וקטורים ומטריצות

נתונה הרשת:



4 נוירוני קלט
3 נוירוני ביניים
5 נוירוני פלט.

$$y_1 = W_{11} * p_1 + W_{12} * p_2 + W_{13} * p_3 + b_{2_1} \quad \text{Y1 מחושב כך:}$$

$$y_2 = W_{21} * p_1 + W_{22} * p_2 + W_{23} * p_3 + b_{2_2} \quad \text{איך נחשב } y_2 ?$$

חישובי רשת נוירונים כמכפלת וקטורים ומטריצות

- קלט לרשת נוירונים מיוצג כוּקטור עמודה. נסמן אותו כ- p .
- וקטור הפלט מחושב כמכפלה בין מטריצת הקשרים לבין וקטור הקלט היא $W_2 * p$:

$$\begin{array}{ccc} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \\ W_{41} & W_{42} & W_{43} \\ W_{51} & W_{52} & W_{53} \end{array} * \begin{array}{c} p_1 \\ p_2 \\ p_3 \end{array} + \begin{array}{c} b2_1 \\ b2_2 \\ b2_3 \\ b2_4 \\ b2_5 \end{array} = \begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{array}$$

- לפי חוקי הכפלת מטריצות:

$$y = W * p + b2$$

- חישוב W^*p
- עדכון **ערך הסף bias**
- הפעלת פונקציית מעבר

הפלט של שכבת הנוירונים עבור הקלט p :

$$\text{Output} = \text{transfer_function}(\overbrace{W^*p}^y + b2)$$

פלט שכבה אחרונה:

$$\text{Output} = \text{trans_f2} (W2 * p + b2)$$

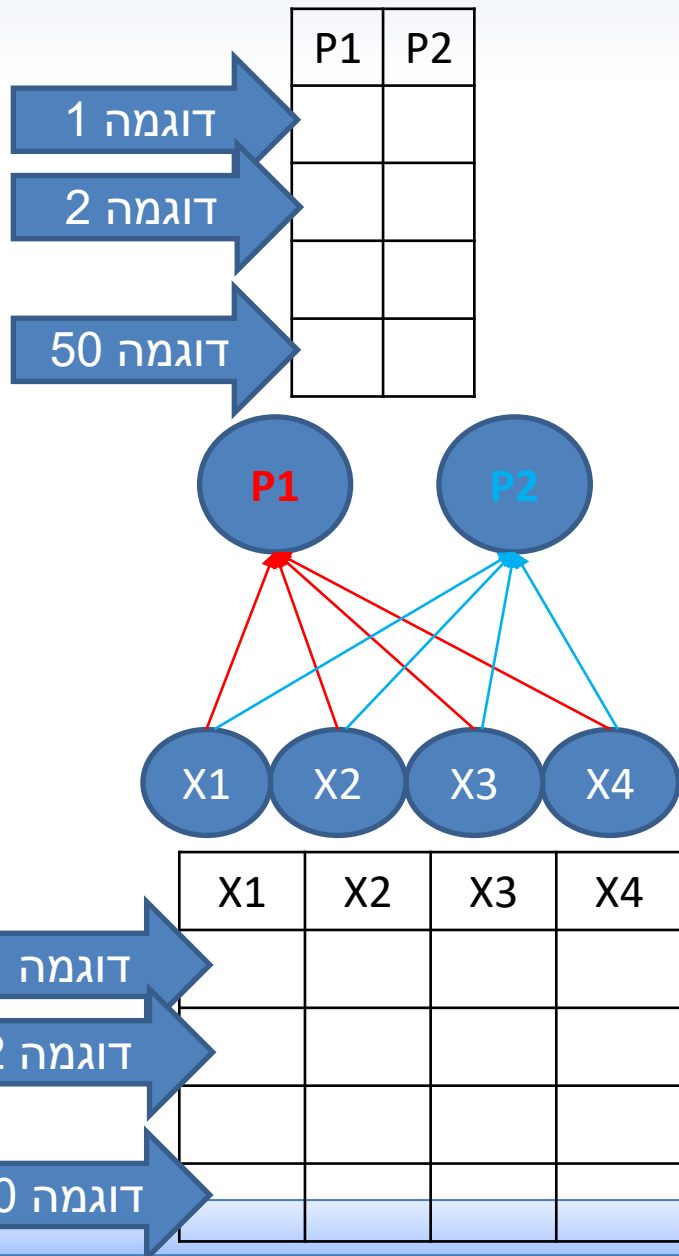
פלט שכבת הביניים:

$$p = \text{trans_f1} (W1 * x + b1)$$

לסיכום, העברת קלט דרך חישוב של רשת בעלת
שכבת ביניים אחת:

$$\text{Output} = \text{trans_f2} (W2 * \text{trans_f1} (W1 * x + b1) + b2)$$

חישובי רשת נוירונים כמכפלת וקטורים ומטריצות



כיצד נחשב את הפלט עבור מספר גדול של דוגמאות בבת אחת?

נניח שיש 4 נוירוני קלט: X_1, X_2, X_3, X_4

2 נוירוני ביניים: P_1, P_2

50 דוגמאות

חישוב ערכי שכבת ביניים ברשת:

$$W * X + b = P$$

$$(2, 4) (4, 50) \quad (2,) \quad (2, 50)$$

תרגיל 1 – Feed Forward

$$W^{[1]} = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & 0 & 0 \\ -0.5 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

• נתונה רשת בעלת הפרמטרים הבאים:

1. כמה מאפיינים יש לקלט של הרשת?

2. כמה שכבות יש לרשת?

3. כמה נוירונים יש בכל אחת מהשכבות?

4. ציירו את הרשת

5. חשבו את תוצאת הרשת עבור הקלטים הבאים, כאשר פונקצית האקטיבציה היא הפונקציה הלוגיסטית:

$$x^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, x^{(2)} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

$$W^{[2]} = [1 \quad -1 \quad 1 \quad -1]$$

$$b^{[1]} = \begin{bmatrix} 0 \\ 0 \\ -0.5 \\ 0.5 \end{bmatrix}$$

$$b^{[2]} = [0.5]$$

תרגיל 2 – יישום בפיתון

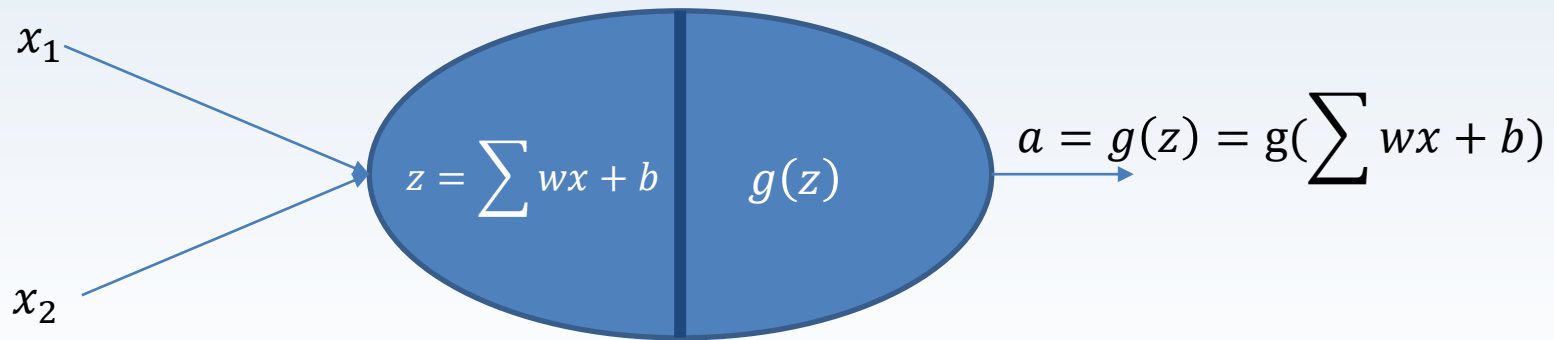
- ממשו פונקציה $\text{feed_forward}(x, W1, b1, W2, b2)$ המקבלת וקטור קלט יחיד x ומטריצות פרמטרים של רשת $W1, b1, W2, b2$. הפונקציה מחשבת את הפלט של הרשת על הוקטור x .
- בדקו את פעולת הפונקציה על הרשת מתרגיל 1
- חשבו את \hat{y} של הרשת מתרגיל 2 עבור הקלט הבא, עדכנו את הפונקציה $\text{feed_forward}(X, W1, b1, W2, b2)$ כך שתקבל מטריצת קלט X במקום וקטור קלט יחיד:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

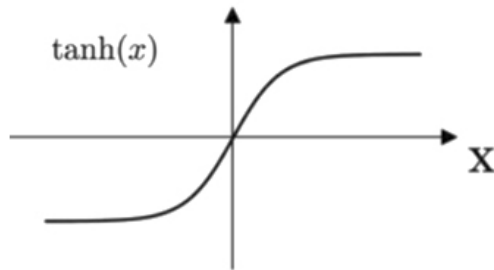
פונקציית מעבר

הפרדת תפקידים בנוירון

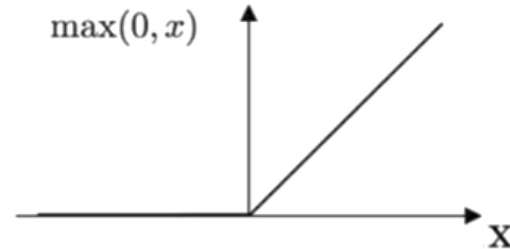
חלק לינארי ואקטיבציה



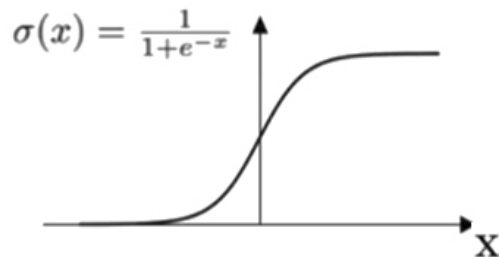
Tanh



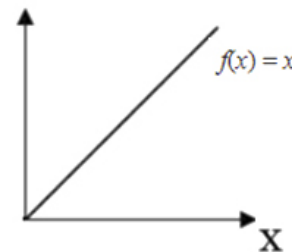
ReLU



Sigmoid



Linear



- לכל שכבה יכולה להיות פונקציית מעבר שונה.
- בד"כ נבחר בפונקציית מעבר לא ליניארית לשכבות הביניים.
- פונקציית המעבר של שכבת הפלט תיבחר בהתאם לבעיה, למשל:
 - מה יתאים לבעיית קלסיפיקציה?
 - מה יתאים לבעיית רגרסיה?
- אם נרצה לאמן את הרשת בשיטת gradient descent, לא נוכל להשתמש בפונקציות לא גזירות.

forward pass



error function



backprop



result

Back propagation

איך נמצא את המשקלות של הרשת?

כל משקולת תאומן לפי כלל ה-gradient descent

בואו ניזכר איך חישבנו את השינוי במשקולות:

$$w(j) = w(j) - (lr/n) \sum_{i=1}^m \underbrace{(\hat{y}(i) - y(i))}_{\text{השגיאה בין הערך החזוי לאמיתי}} x(j)(i)$$

השגיאה בין הערך החזוי לאמיתי

$\hat{y}(i)$ – הערך החזוי בדגימה i

$y(i)$ – הערך האמיתי בדגימה i

m – מספר הדגימות

n – מספר התכונות, כולל ההטייה

$w(j)$ – משקולת של התכונה j

$x(j)(i)$ – התכונה j בדגימה i

lr – קצב הלמידה

כל עוד שכבת הפלט היא היחידה שצריך לאמן $y(i)$ נתון לנו
אבל מה קורה כשצריך לאמן שכבות ביניים?

שלב 1 – Feed Forward & Back propagation

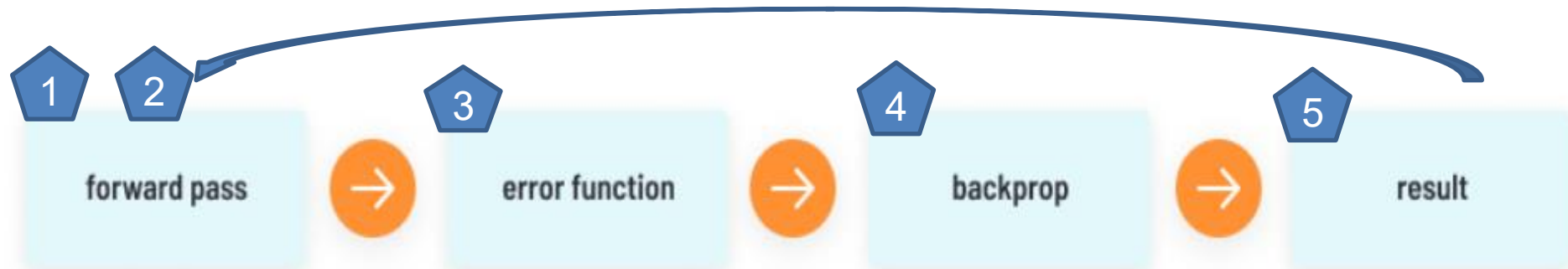
שלב 1 – איתחול הרשת עם ערכי משקולות ראשוניים

שלב 2 - חישוב הניורונים בכל השכבות בהינתן ערכי משקולות התחלתיים וערכי הפיצ'רים

שלב 3 – חישוב הטעות באיטרציה הראשונה

שלב 4 – תיקון ערכי המשקולות מהשכבה האחרונה, דרך השכבות החבויות עד לשכבת הקלט

שלב 5 – תוצאה – ערכי משקולות מתוקנים
חזרה על שלב 2 (הפעם עם ערכי משקולות מתוקנים)
עד שמתכנסים לערך טעות מינימלי



Feed Forward – ברשת עם שכבת ביניים אחת

$$Z^{[1]} = W^{[1]}X + b^{[1]} \quad (1)(2) \text{ חישוב סכום מכפלת המשקולות בערכי הפיצ'רים}$$

$$A^{[1]} = \tanh(Z^{[1]}) \quad (2)(2) \text{ הפעלת פונקציית אקטיבציה לערכי Z}$$

$$Z^{[2]} = W^{[2]}A^{[1]} + b^{[2]} \quad (3)(2) \text{ חישוב סכום מכפלת המשקולות בערכי רשת הביניים}$$

$$\hat{Y} = A^{[2]} = \sigma(Z^{[2]}) \quad (4)(2) \text{ הפעלת פונקציית אקטיבציה לערכי רשת הביניים}$$

(3) חישוב פונקציית העלות בהתחשב בערכי שכבת הביניים:

$$J = -\frac{1}{m} \sum_{i=1}^m \left(y^{(i)} \log(a^{[2](i)}) + (1 - y^{(i)}) \log(1 - a^{[2](i)}) \right)$$

Back propagation – ברשת עם שכבת ביניים אחת

(4.1) חישוב כל ערכי הנגזרות החלקיות:

$$dZ^{[2]} = A^{[2]} - Y$$

$$dW^{[2]} = \frac{1}{m} dZ^{[2]} A^{[1]T}$$

$$db^{[2]} = \frac{1}{m} \text{np.sum}(dZ^{[2]}, \text{axis} = 1, \text{keepdims} = \text{True})$$

$$dZ^{[1]} = W^{[2]T} dZ^{[2]} * g^{[1]'}(Z^{[1]})$$

$$dW^{[1]} = \frac{1}{m} dZ^{[1]} X^T$$

$$db^{[1]} = \frac{1}{m} \text{np.sum}(dZ^{[1]}, \text{axis} = 1, \text{keepdims} = \text{True})$$

Back propagation – ברשת עם שכבת ביניים אחת

(4.1) עדכון כל המשקולות w בין השכבה האחרונה לשכבת הביניים ובין שכבת הביניים לשכבת הכניסה:

$$W1 = W1 - \text{learning_rate} * dW1$$

$$b1 = b1 - \text{learning_rate} * db1$$

$$W2 = W2 - \text{learning_rate} * dW2$$

$$b2 = b2 - \text{learning_rate} * db2$$