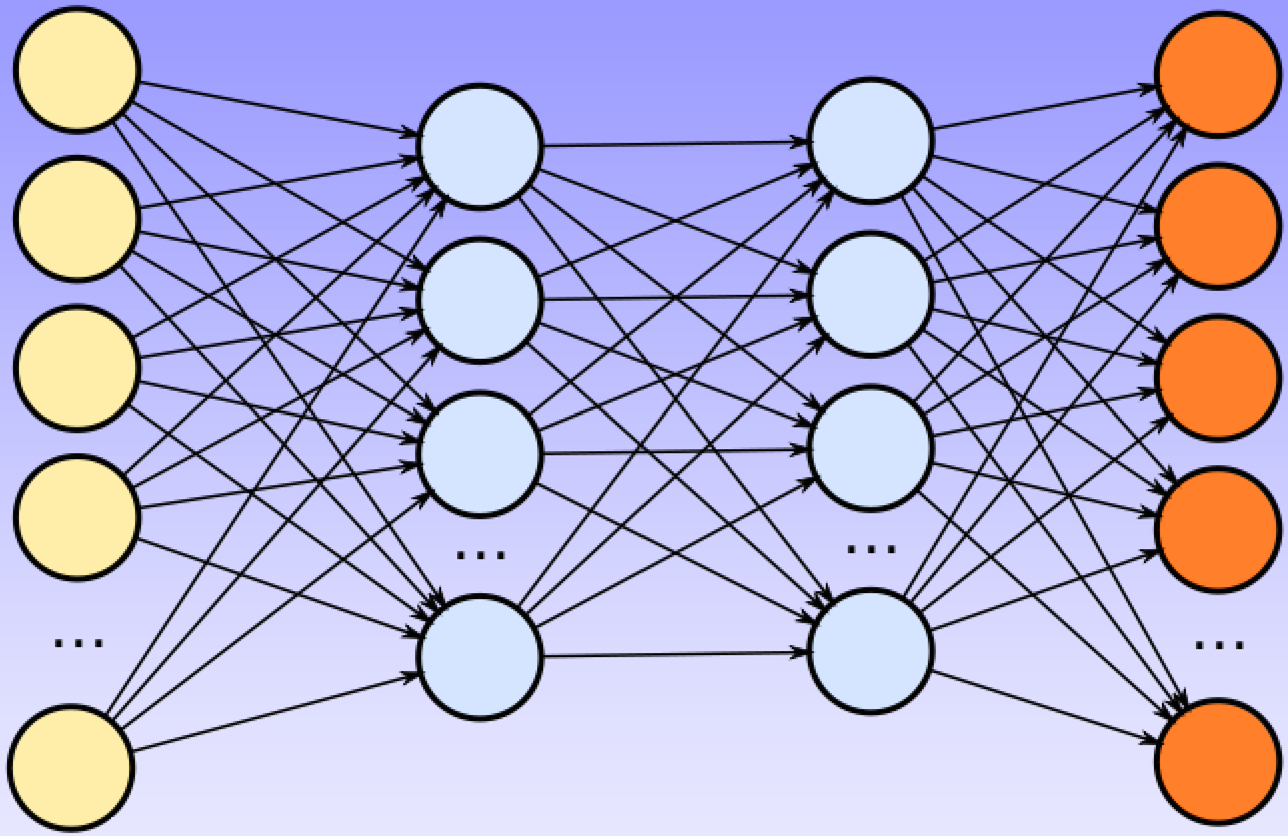
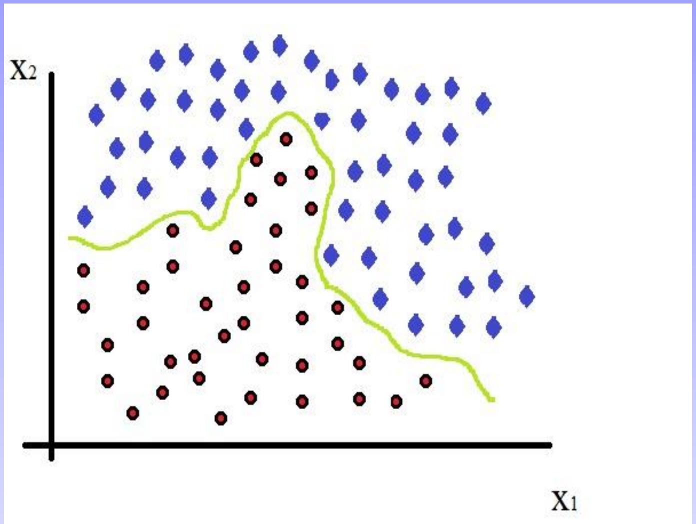
ANN



היפותזות לא ליניאריות

- רשתות עצביות – משמשות במערכות רבות של למידה חישובית.
- סווג לא ליניארי.
- נתבונן בדוגמא הבאה:



- אפשר להשתמש ברגרסיה לוגיסטית,

באמצעות פולינום עם מספיק איברים כמו:

$$h_\theta(x) = g\left(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^{\ 2} + \theta_3 x_2^{\ 2} + \theta_4 x_1^{\ 2} x_2 + \theta_5 x_1 x_2^{\ 2} + \theta_6 x_1^{\ 2} x_2^{\ 2} + \theta_7 x_1^{\ 2} x_2^{\ 3} + \theta_8 x_1^{\ 3} x_2^{\ 2}\right)$$

פתרון כזה יוצר את משטח ההפרדה (--).

- מתאים לוקטור תכונות דו-מימדי.

היפותזות לא ליניאריות

- בבעיות למידה רבות מספר רב של תכונות – הפתרון הקודם לא מתאים.
- לדוגמא –
- בבעיית הבתים עליה דיברנו, נניח שזוהי בעיית סווג – וצריך לסווג את הבתים כבעלי פוטנציאל גבוה או נמוך למכירה –

תכונות רבות –

- גודל הבית,
- מספר חדרי השינה,
- מספר מפלסים,
- שטח הגינה,
- גיל הבית,
- מקומות חנייה וכו'.

היפותזות לא ליניאריות

- בבעיית גילוי ה- spam – (או באופן כללי בבעיית זיהוי טקסט) וקטור התכונות עשוי להכיל מילון של 10000-50000 מלים.

AARDVARK
AARDWOLF AARON
ABACK ABACUS
ABAFT ABALONE
ABANDON
ABANDONED
ABANDONMENT
ABANDONS ABASE
ABASED
ABASEMENT ABASH
ABASHED ABATE
ABATED
ABATEMENT
ABATES ABATTOIR
ABATTOIRS ABBE
ABBESS ABBEY
ABBEYS ABBOT
ABBOTS
ABBREVIATE
ABBREVIATED
ABBREVIATES
ABBREVIATING
ABBREVIATION
ABBREVIATIONS
ABDICATE
ABDICATED
…

היפותזות לא ליניאריות

- עבור 100 תכונות –

אם ננסה לכלול את כל התכונות הקוודרטיות בלבד –

כלומר:

$$x_1^{\,2},\ x_1x_2,\ \ x_1x_3,\ldots,\ x_1x_{100},\ldots,\ x_1^{\,2},\ \ x_2x_3,\ldots\ x_{100}^{\ \ 2}$$

$$\approx o\left(n^2\right)\quad,\quad \approx n^2/2=5000\ \ features$$

היפותזות לא ליניאריות

- אפשר לכלול רק תת-קבוצה של התכונות המוצעות, לדוגמא:

$$x_1^{\ 2},\ \ x_2^{\ 2}, \ldots, x_{100}^{\ \ 2}$$

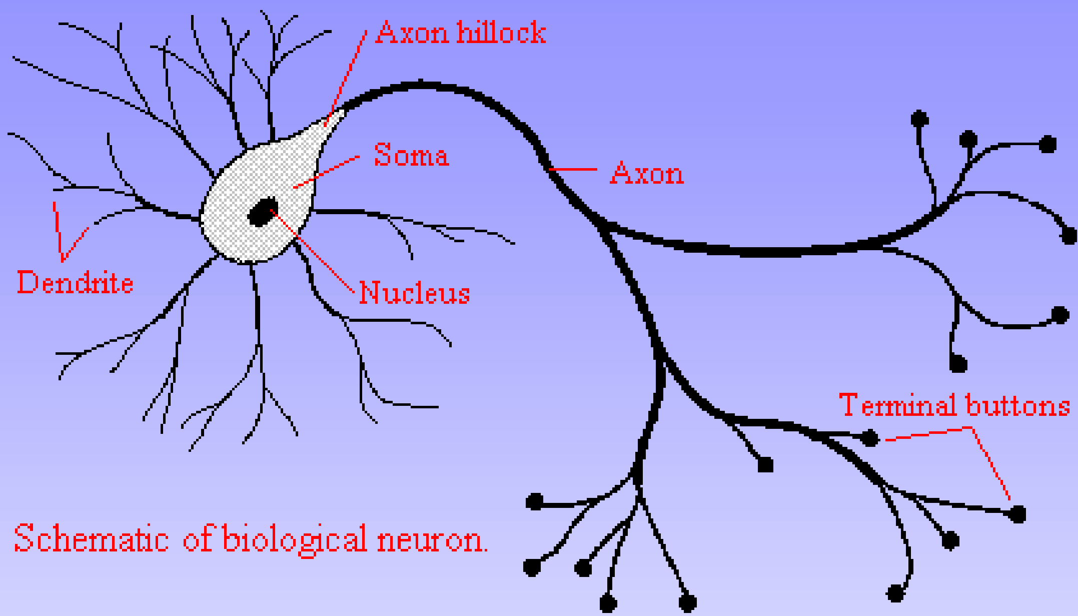
- התאמת משטחי הפרדה כמו אליפסות, אך לא משטחים יותר מורכבים הדרושים לביצוע מוצלח של הסווג.
- הכללת פרמטרים מדרגה שלישית: עבור n=100 תכונות מקוריות מספר התכונות עשוי לעלות על 150,000.

$$\approx o\left(n^3\right) \quad , \quad \approx 170,000 \ \ features$$

- יותר מדי תכונות: 1. התאמת יתר 2. חישוביות

האם יש דרך מוצלחת יותר לבנות מסווג לא ליניארי?

ייצוג המודלים ברשתות עצביות



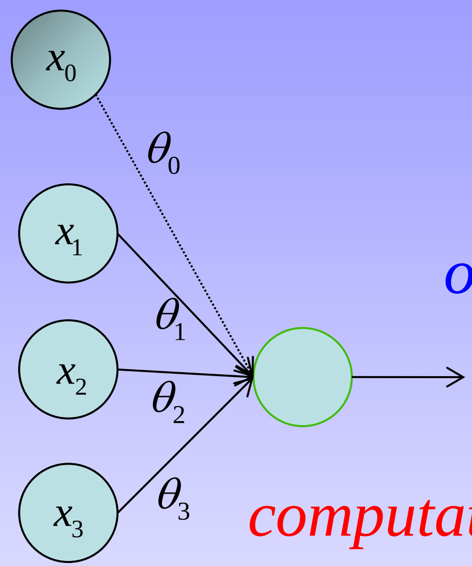
תיאור של נוירון בודד:
מבנה של נוירון אופייני.

-
- ניתן להבחין במרכיבים הבאים:
- 1. דנדריטים: "ערוצי כניסה" מנוירונים אחרים Dendrites
- 2. גוף התא: המרכיב ה"חישובי" Soma
- 3. אקסונים וחיבורים סינפטיים: ערוצי היציאה לנוירונים אחרים (Axons)

מודל של תא עצב (נוירון) בודד

$bias\ unit\ x_0 = 1$

- מודל של נוירון – מימוש פשטני:



output

$h_\theta(x)$

computation

input

(dendrites)

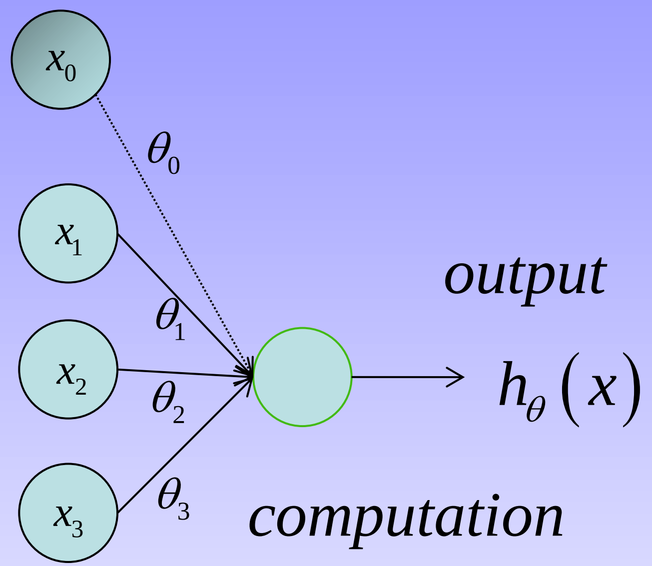
$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

מודל של תא עצב (נוירון) בודד

bias unit $x_0 = 1$



input

(dendrites)

- למעשה זוהי יחידה לוגיסטית

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

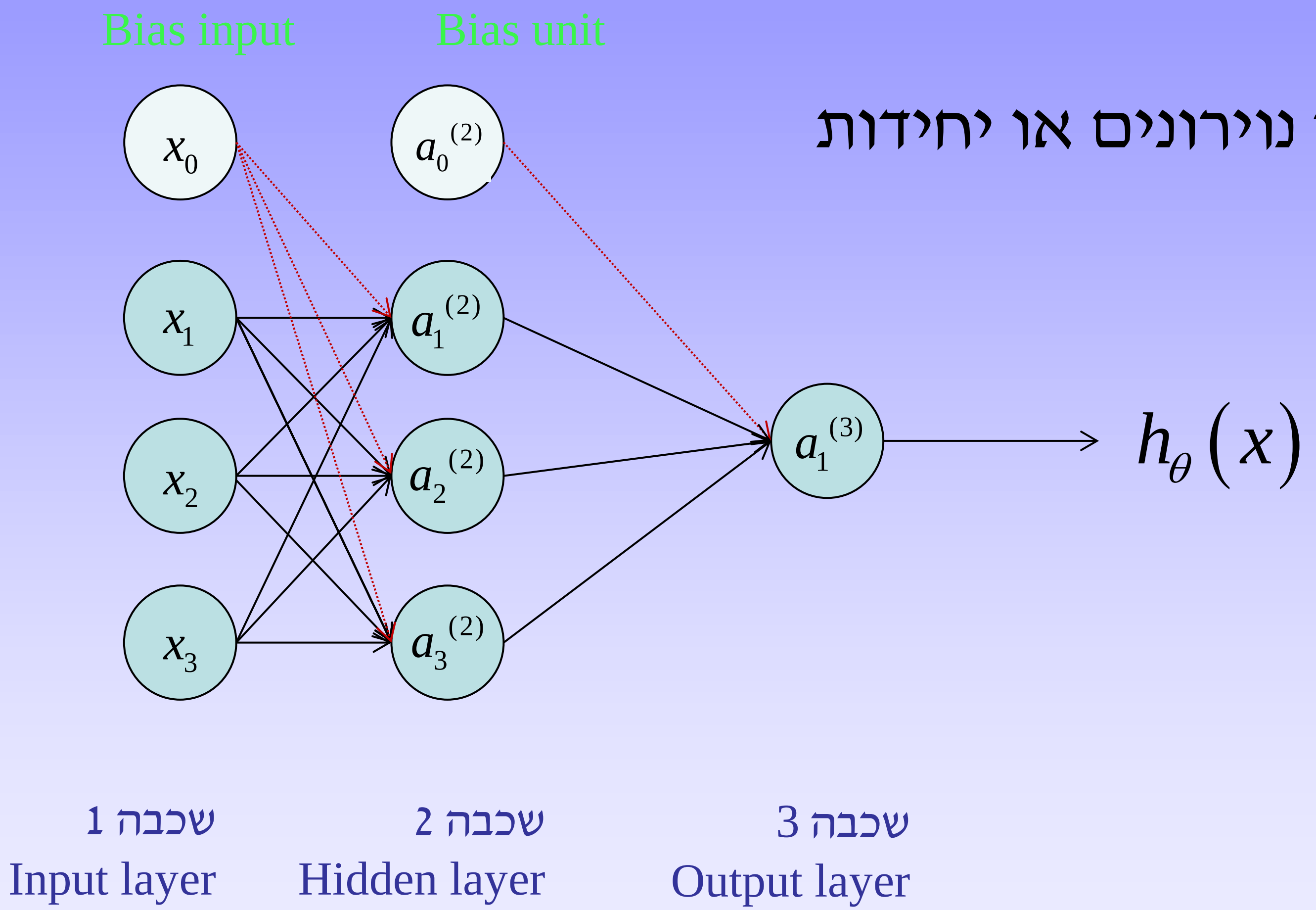
- פונקציית האקטיבציה היא פונקציה סיגמואידית

$$g(z) = \frac{1}{1+e^{-z}}$$

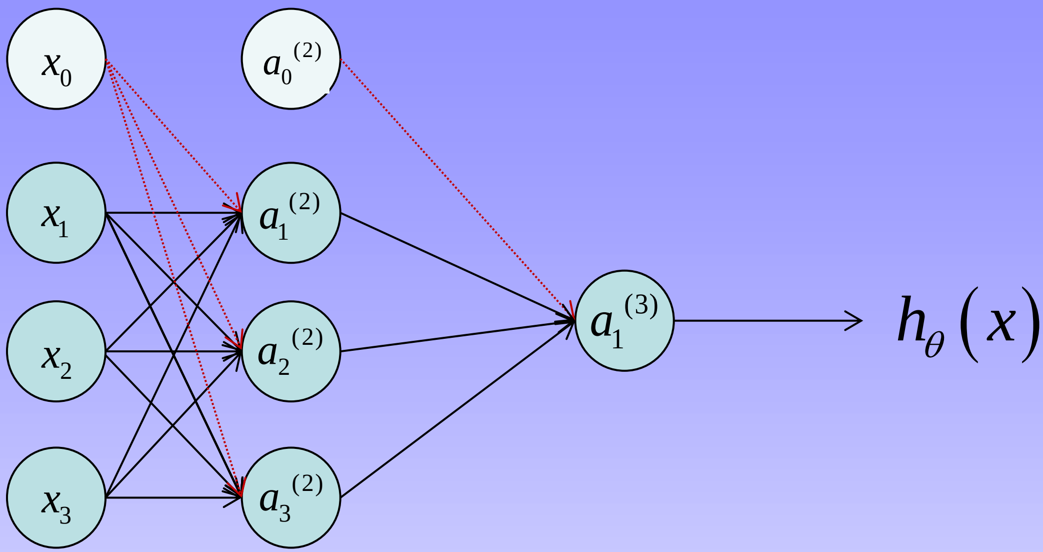
- הפרמטרים נקראים לעתים משקלות.

רשת עצבית

רשת עצבית:
חיבור של מספר נוירונים או יחידות



רשת עצבית

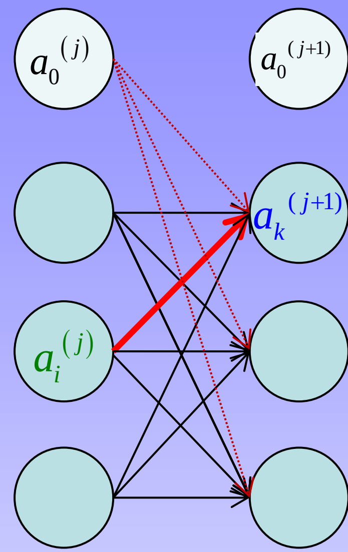


עבור הרשת הנ״ל נשתמש בסימונים הבאים :

- האקטיבציה של הנוירון ה- i בשכבה ה- j – $a_i^{(j)}$
- מטריצת משקלות (או פרמטרים השולטים על המיפוי) - $\theta^{(j)}$
(מהשכבה ה- j לשכבה ה- j+1)

לדוגמא : האקטיבציה של הנוירון הראשון בשכבה השניה
(ה- hidden layer) – $a_1^{(2)}$

רשת עצבית סימון המשקלות (weigts)

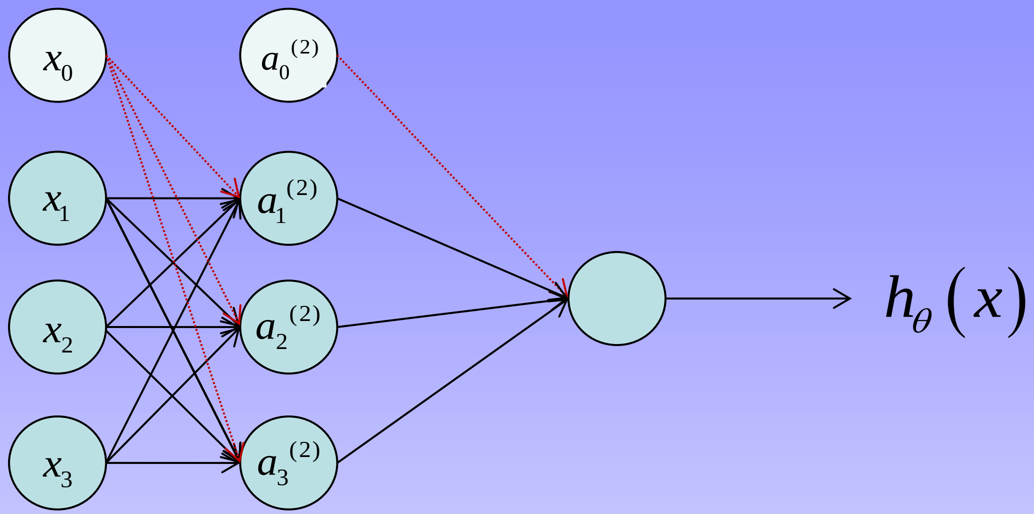


לדוגמא: סימון המשקל מהנוירון ה- i בשכבה ה- j

לנוירון ה- k בשכבה ה- j+1 : $\theta_{ki}^{(j)}$

$$a_i^{(j)} \rightarrow a_k^{(j+1)}$$

רשת עצבית



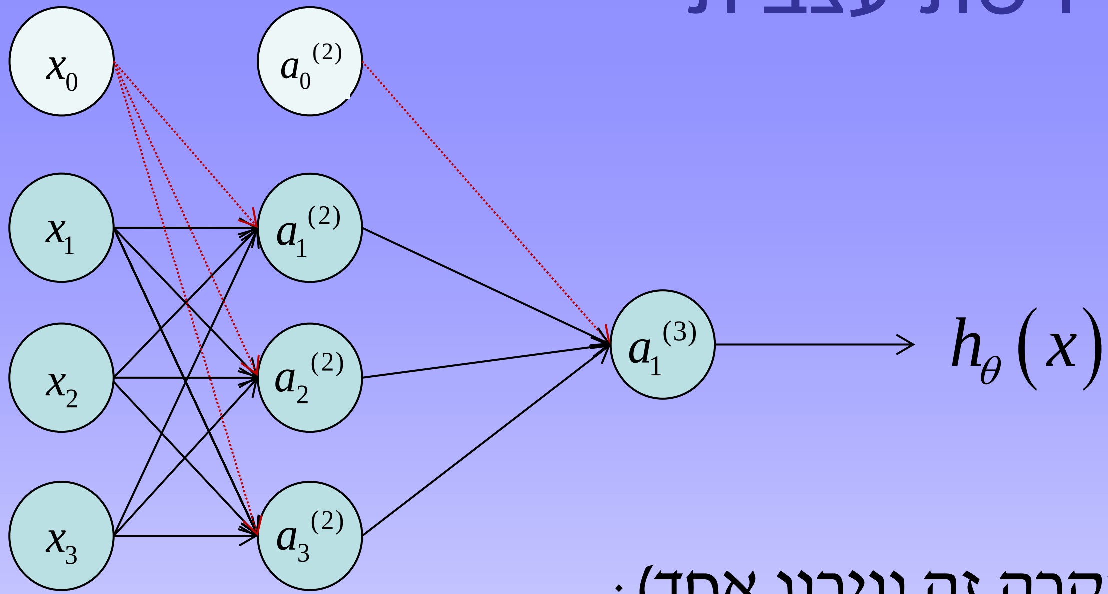
נציג את החישובים שמתבצעים ברשת:

$$a_1^{(2)} = g\left(\theta_{10}^{(1)}x_0 + \theta_{11}^{(1)}x_1 + \theta_{12}^{(1)}x_2 + \theta_{13}^{(1)}x_3\right)$$

$$a_2^{(2)} = g\left(\theta_{20}^{(1)}x_0 + \theta_{21}^{(1)}x_1 + \theta_{22}^{(1)}x_2 + \theta_{23}^{(1)}x_3\right)$$

$$a_3^{(2)} = g\left(\theta_{30}^{(1)}x_0 + \theta_{31}^{(1)}x_1 + \theta_{32}^{(1)}x_2 + \theta_{33}^{(1)}x_3\right)$$

רשת עצבית



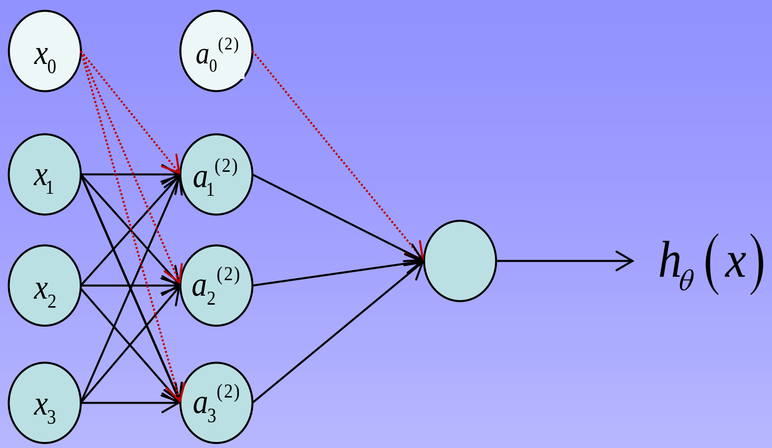
שכבת היציאה (מכילה במקרה זה נוירון אחד):

$$h_\theta(x) = a_1^{(3)} = g\left(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)}\right)$$

הרשת מכילה 3 יחידות כניסה ו-3 יחידות חבויות ולכן:

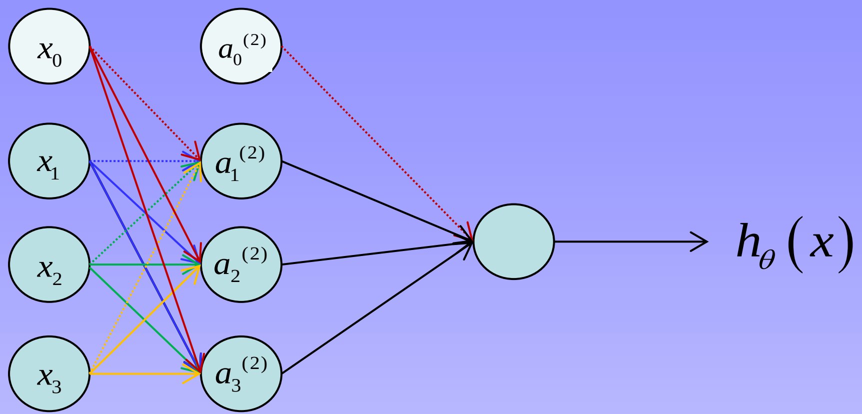
$$\Theta^{(1)} \in R^{3x4}$$

רשת עצבית



באופן כללי, אם לרשת Sj יחידות בשכבה ה- j ו- S_{j+1} יחידות
בשכבה j+1 אז מטריצת המשקלות תהיה מטריצה עם : S_{j+1} x S_j $+1$
מימדים.

Forward Propagation: מימוש מטריצי - וקטורי



$$a_1^{(2)} = g\left(\theta_{10}^{(1)}x_0 + \theta_{11}^{(1)}x_1 + \theta_{12}^{(1)}x_2 + \theta_{13}^{(1)}x_3\right)$$

$$a_2^{(2)} = g\left(\theta_{20}^{(1)}x_0 + \theta_{21}^{(1)}x_1 + \theta_{22}^{(1)}x_2 + \theta_{23}^{(1)}x_3\right)$$

$$a_3^{(2)} = g\left(\theta_{30}^{(1)}x_0 + \theta_{31}^{(1)}x_1 + \theta_{32}^{(1)}x_2 + \theta_{33}^{(1)}x_3\right)$$

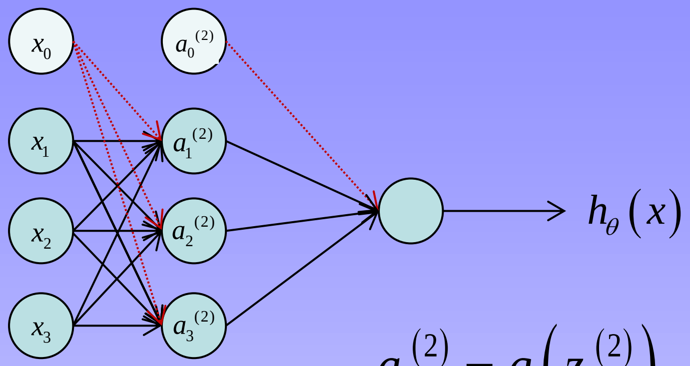
נסמן:

$$z_1^{(2)} = \theta_{10}^{(1)}x_0 + \theta_{11}^{(1)}x_1 + \theta_{12}^{(1)}x_2 + \theta_{13}^{(1)}x_3$$

$$z_2^{(2)} = \theta_{20}^{(1)}x_0 + \theta_{21}^{(1)}x_1 + \theta_{22}^{(1)}x_2 + \theta_{23}^{(1)}x_3$$

$$z_3^{(2)} = \theta_{30}^{(1)}x_0 + \theta_{31}^{(1)}x_1 + \theta_{32}^{(1)}x_2 + \theta_{33}^{(1)}x_3$$

מימוש מטריצי - וקטורי Forward Propagation



כאשר:

$$a_1^{(2)} = g\left(z_1^{(2)}\right)$$

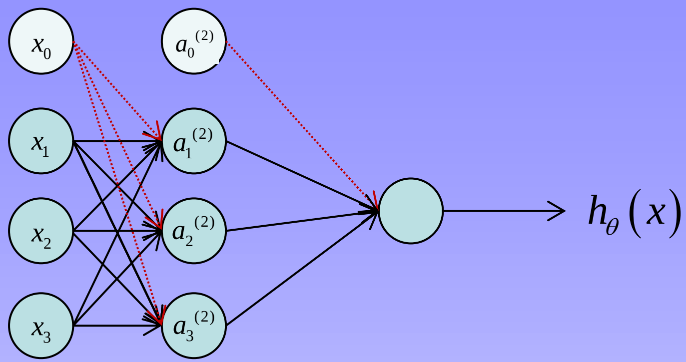
$$a_2^{(2)} = g\left(z_2^{(2)}\right)$$

$$a_3^{(2)} = g\left(z_3^{(2)}\right)$$

$$Z_i^{(j)}$$

כל אחד מה- z_i הוא קלט לנוירון ה- i
בשכבה ה- j, ובמקרה זה לשכבה 2

מימוש מטריצי - וקטורי Forward Propagation



נגדיר באופן מטריצי:

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad z^{(2)} = \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{pmatrix}$$

$$z^{(2)} = \Theta^{(1)} x = \begin{pmatrix} \theta_{10}^{(1)} & \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ \theta_{20}^{(1)} & \theta_{21}^{(1)} & \theta_{22}^{(1)} & \theta_{23}^{(1)} \\ \theta_{30}^{(1)} & \theta_{31}^{(1)} & \theta_{32}^{(1)} & \theta_{33}^{(1)} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$a_2^{(2)} = g\left(z_2^{\;2}\right) \quad \in R^3$$

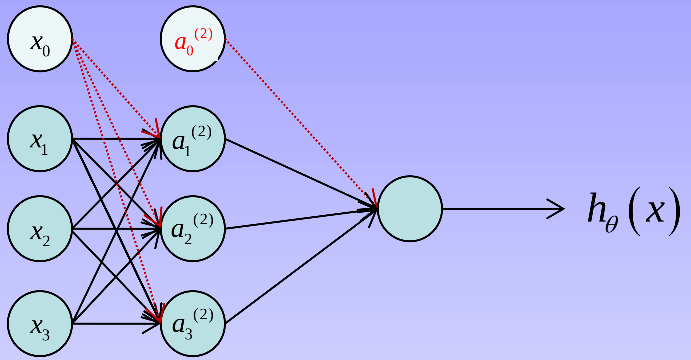
כאשר הפונקציה g פועלת על האלמנטים של z (כל אחד לחוד באמצעות הסיגמואיד)

מימוש מטריצי - וקטורי Forward Propagation

מטעמי עקביות אפשר לסמן את שכבת הכניסה $a^{(1)}$, כלומר $x=a^{(1)}$

ולכן:

$$z^{(2)} = \Theta^{(1)} a^{(1)}$$



נוסיף לשכבה החבויה את:

$$a_0^{(2)} = 1$$

$$z^{(3)} = \Theta^{(2)} a^{(2)}$$

כאשר:

$$a^{(2)} = \begin{pmatrix} a_0^{(2)} \\ a_1^{(2)} \\ a_2^{(2)} \\ a_3^{(2)} \end{pmatrix}$$

מימוש מטריצי - וקטורי Forward Propagation

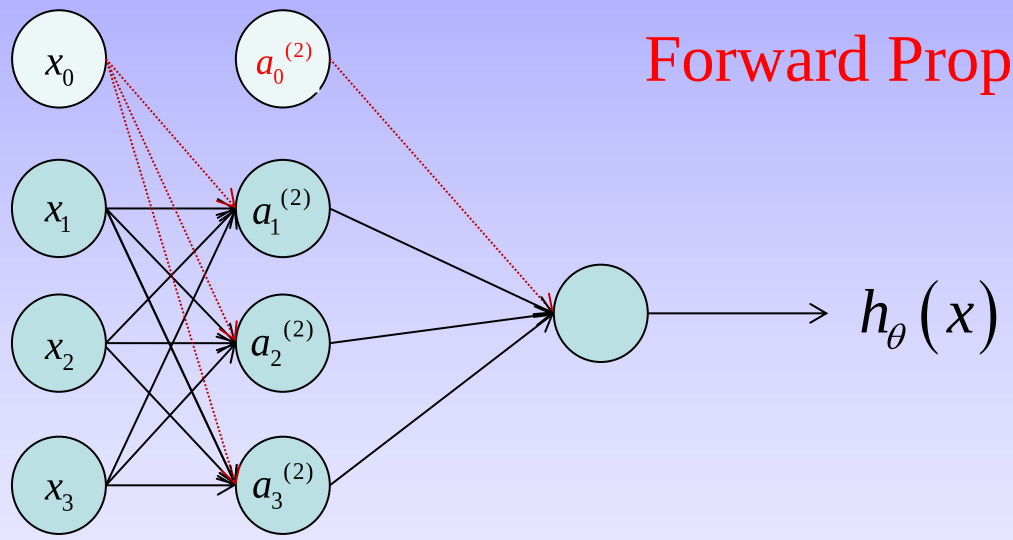
והיציאה של הרשת היא אם כן:

$$h_\theta(x) = a_1^{(3)} = g\left(z^{(3)}\right)$$

כאשר:

$$z^{(3)} = \Theta^{(2)} a^{(2)}$$

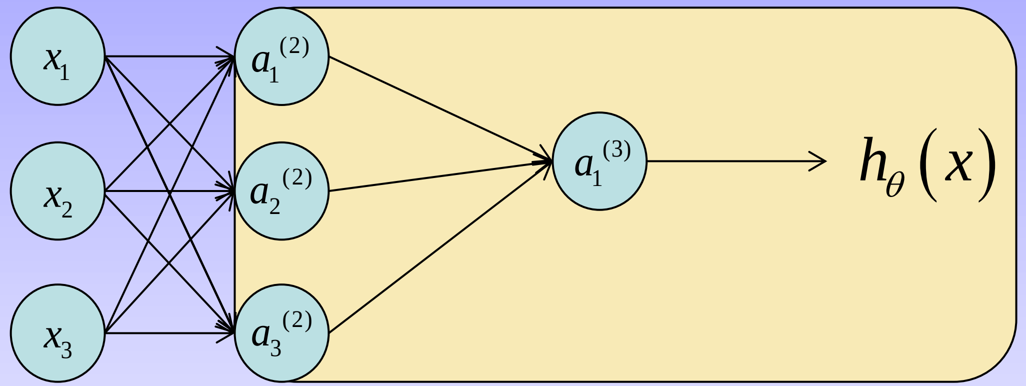
תהליך זה נקרא Forward Propagation



Forward Propagation

ה- Forward Propagation עוזר להבין מה הרשת מבצעת, ואיך היא מאפשרת ליצור היפותיזות לא ליניאריות

נתבונן בחלק השמאלי של הרשת בלבד

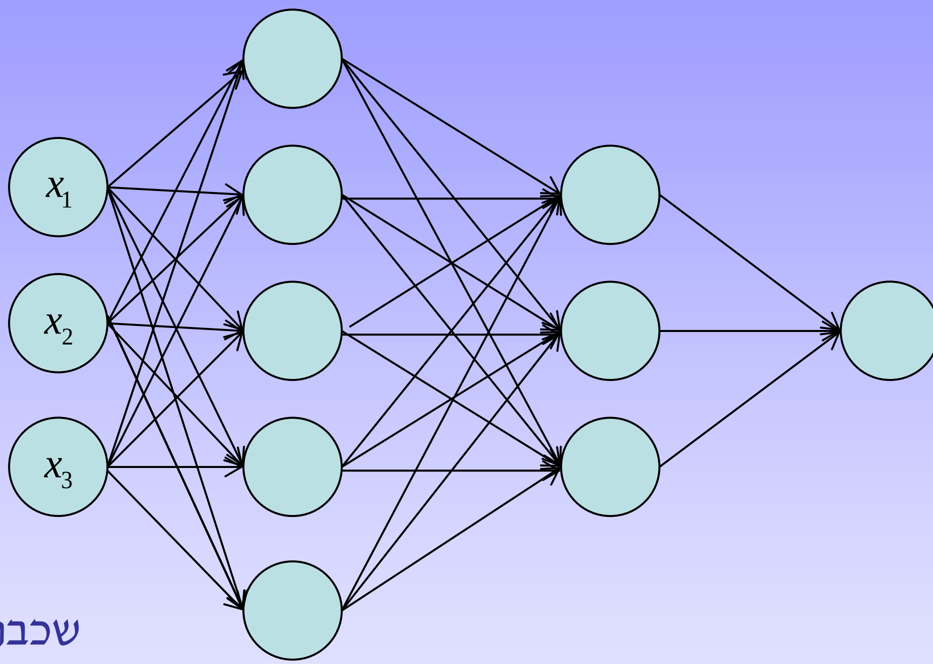


מה שהרשת מבצעת כאן זהה לרגרסיה לוגיסטית:

$$h_\theta(x) = a_1^{(3)} = g\left(z^{(3)}\right) = g\left(\Theta_{10}{}^{(2)}a_0{}^{(2)} + \Theta_{11}{}^{(2)}a_1{}^{(2)} + \Theta_{12}{}^{(2)}a_2{}^{(2)} + \Theta_{13}{}^{(2)}a_3{}^{(2)}\right)$$

תצורות רשת נוספות

אפשר לבנות רשת בתצורות שונות :



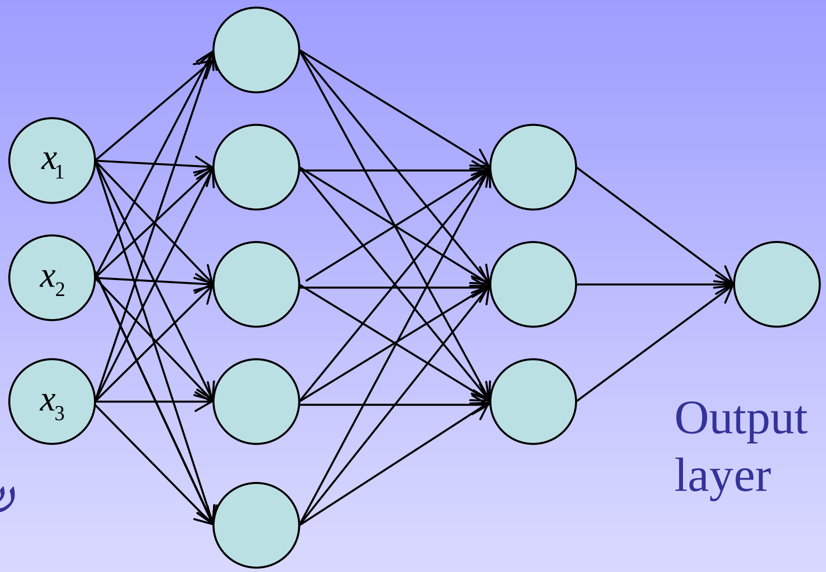
Output layer

שכבה 1
Input layer

שכבות חבויות
Hidden layers

תצורות רשת נוספות

רשתות אלה נקראות Feedforward Networks



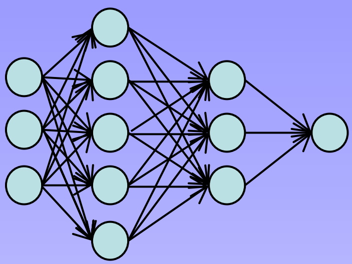
Output
layer

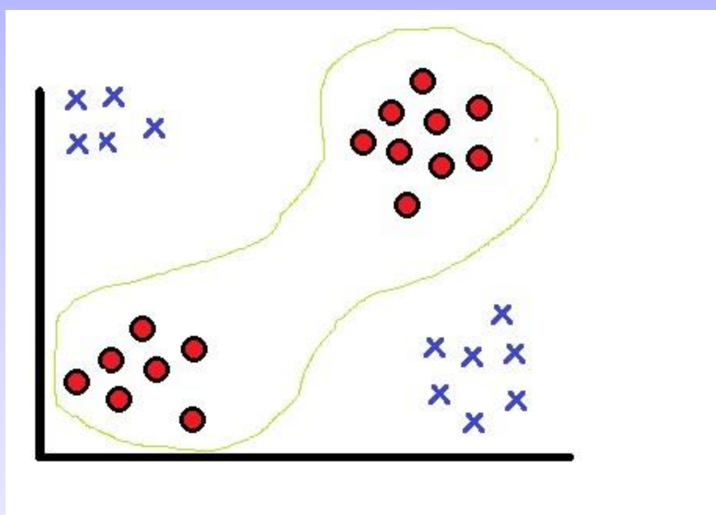
שכבה 1
Input
layer

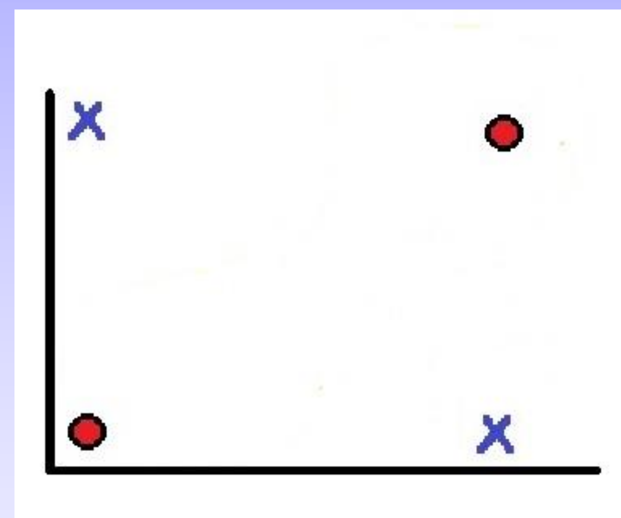
שכבות חבויות
Hidden layers

דוגמא לעבודת הרשת על קלט לא ליניארי

כיצד רשת עצבית עובדת על קלט מורכב לא ליניארי , וכיצד רשת יכולה ללמוד היפותיזות לא ליניאריות מורכבות?



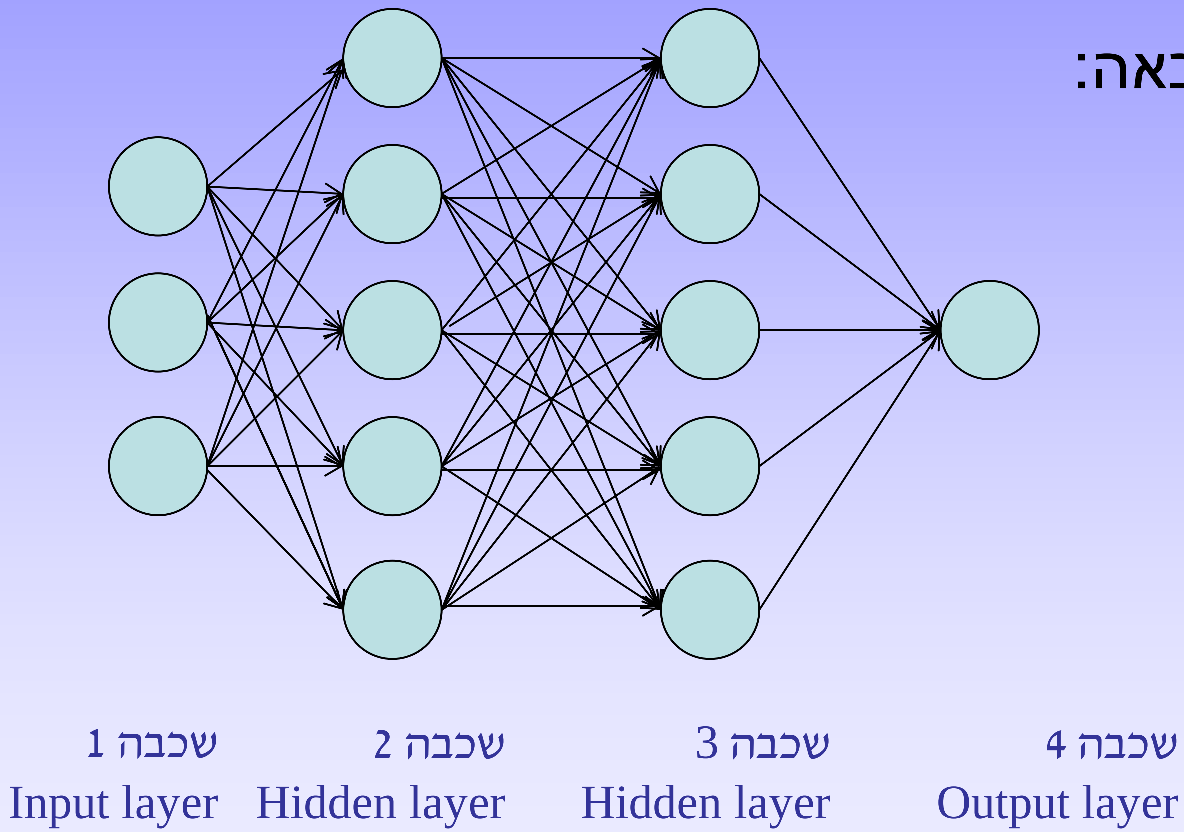




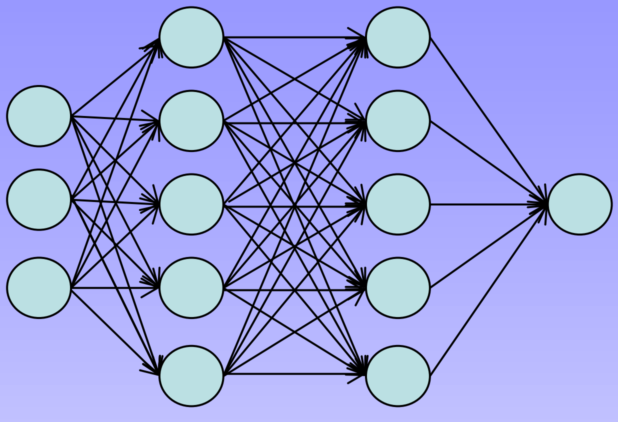
זוהי דוגמא פשוטה של ה- data משמאל

אלגוריתם הלימוד להתאמת פרמטרים

- נדבר עתה על פונקציית המחיר.
- נסמן את הרשת הבאה:



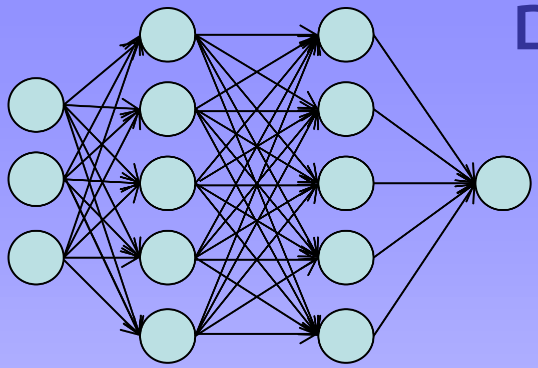
אלגוריתם הלימוד להתאמת פרמטרים



נניח שנתונות דוגמאות האימון הבאות:

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots (x^{(m)}, y^{(m)})\}$$

אלגוריתם הלימוד להתאמת פרמטרים



נסמן ב- L – סה"כ השכבות ברשת. במקרה הזה L=4

נסמן ב- s_l - מספר היחידות (ללא ה- Bias) בשכבה ה- l

לדוגמא: $s_1 = 3, \quad s_2 = 5, \quad s_4 = s_L = 1$

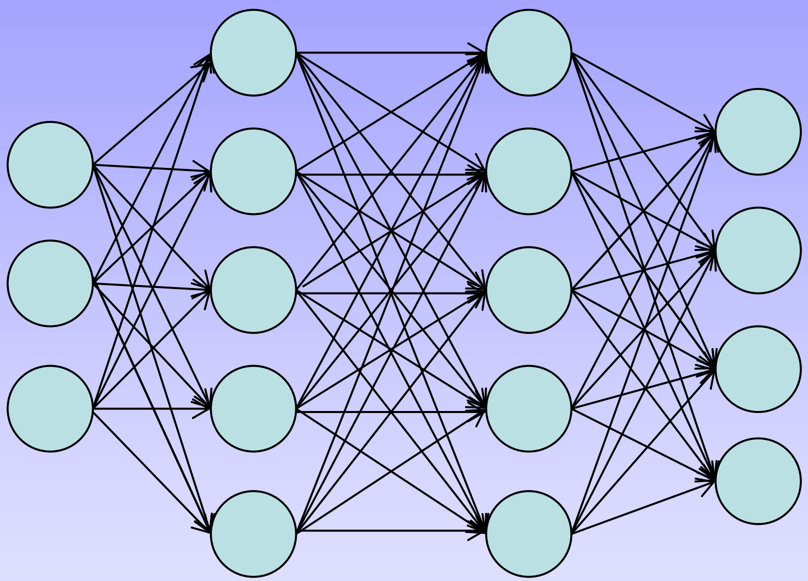
במקרה של סווג בינארי דרושה יחידת פלט אחת, והתיוג הוא $y = 0 \quad or \quad y = 1$

$$h_\Theta(x) \in R$$

$$S_L = 1$$

Multiclass classification סווג רב מחלקתי

כאשר הסווג הוא ל- k מחלקות, לדוגמא – סווג ל- k=4 מחלקות



$$y \in R^k, \quad y = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

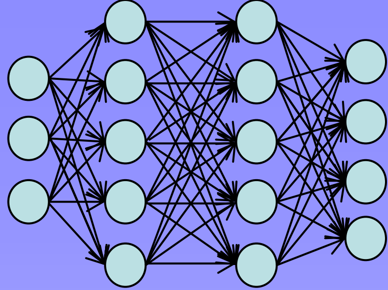
לדוגמא – מצב מזג האוויר:

שמשי, מעונן, גשום, מושלג

(סה"כ צריך k יחידות פלט)

$$h_\Theta(x) \in R^4$$

$$S_L = 4$$



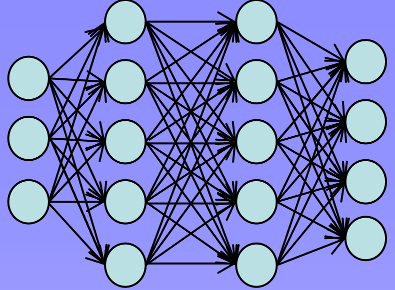
פונקציית המחיר עבור רשת עצבית

פונקציית המחיר עבור הרגרסיה הלוגיסטית:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log((h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)})) + \underbrace{\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^{\,2}}$$

(איבר הרגולריזציה
(ללא θ_0)

עבור רשת עצבית נבצע הכללה של פונקציית המחיר.



פונקציית המחיר עבור רשת עצבית

- עבור רשת עצבית נבצע הכללה של פונקציית המחיר:

$$h_\Theta(x) \in R^K \quad (h_\Theta(x))_i = i^{th}\ output$$

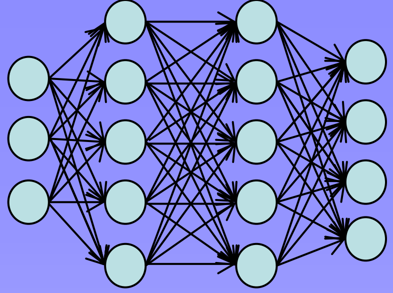
$$J(\theta) = -\frac{1}{m}\left(\sum_{i=1}^{m} \sum_{k=1}^{K} y_k{}^{(i)} \log((h_\theta(x^{(i)}))_k + (1 - y_k{}^{(i)}) \log(1 - h_\theta(x^{(i)})_k) \right)$$

$$\underbrace{+\frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{S_l} \sum_{j=1}^{n} (\Theta_{ji}{}^{l})^2}_{\text{איבר הרגולריזציה}}$$

עבור כל יחידות הפלט

עבור כל הדוגמאות

בהמשך נבצע אופטימיזציה עבור פונקציית המחיר הנ"ל



אלגוריתם ה- Backpropagation

נדבר על אלגוריתם למיזעור פונקציית המחיר: $\min_{\Theta} J(\Theta)$

כדי למזער את פונקציית המחיר נצטרך לחשב את: $J(\Theta)$

$$\frac{\partial}{\partial \Theta_{ij}{}^{l}} J(\Theta), \quad \Theta_{ij}{}^{l} \in R$$

איך נחשב את הנגזרת החלקית לפי $\Theta_{ij}{}^{l}$?