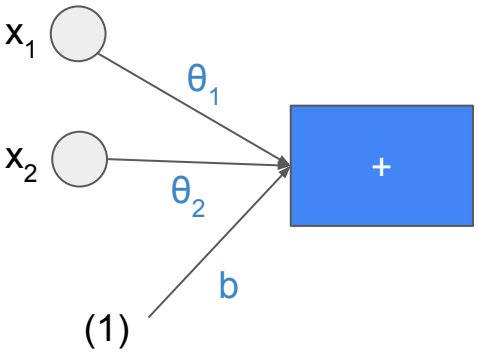
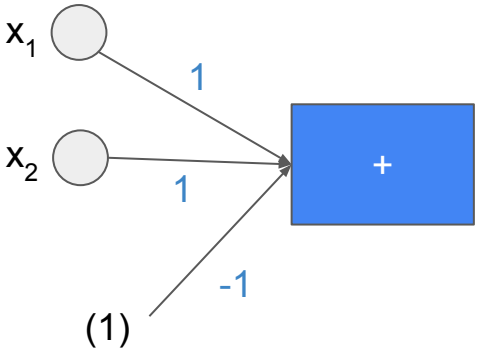
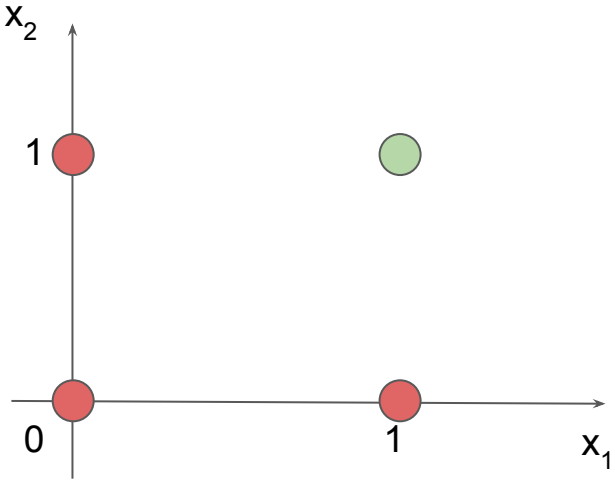
עיבוד שפה טבעית ש8:
תיוג רצפים נוירוני
(SLP 9)

רשתות נוירונים פשוטות - תזכורת

- יחידת הבסיס (״נוירון״) - פרספטרון (ציון לינארי, חיזוי לפי > 0)
- נדמיין פרספטרון שצריך לחשב AND

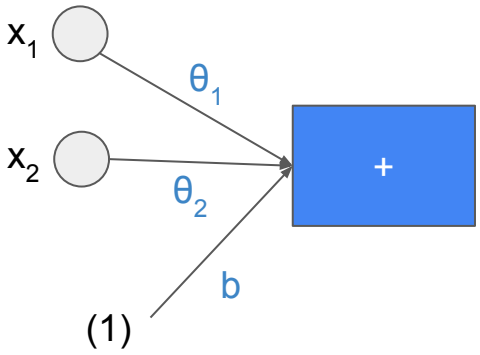


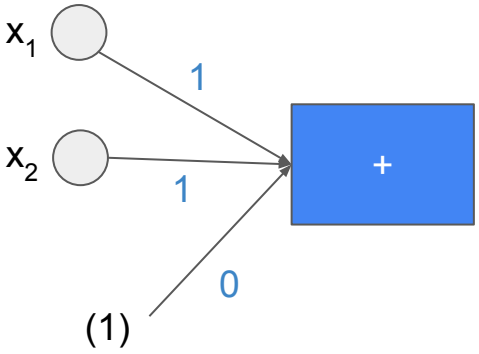


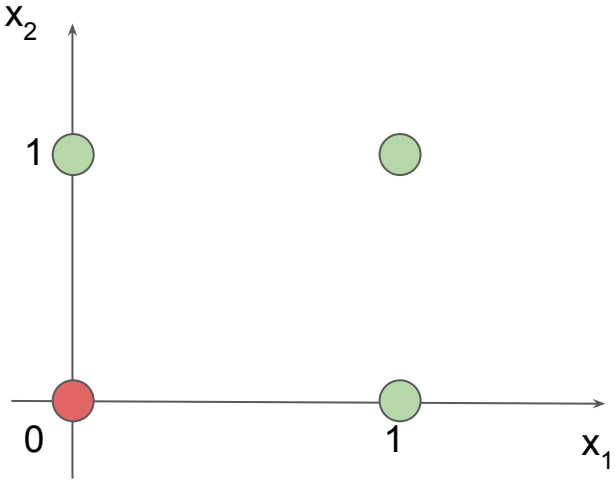


רשתות נוירונים פשוטות - תזכורת

- יחידת הבסיס (״נוירון״) - פרספטרון
- נדמיין פרספטרון שצריך לחשב AND
- וכעת פרספטרון שצריך לחשב OR

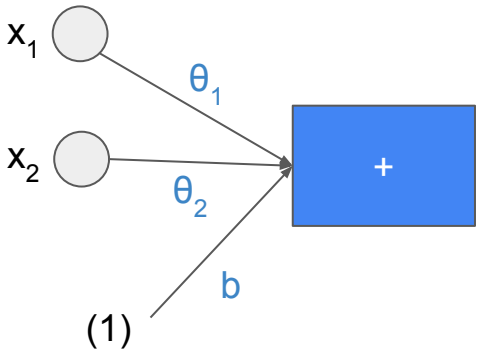


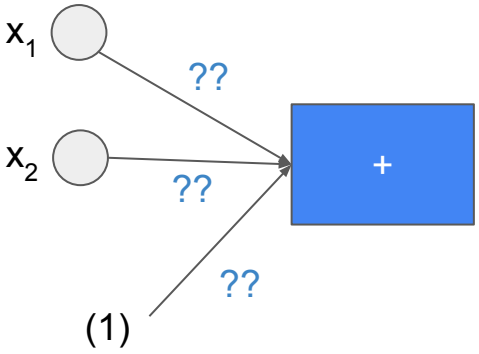


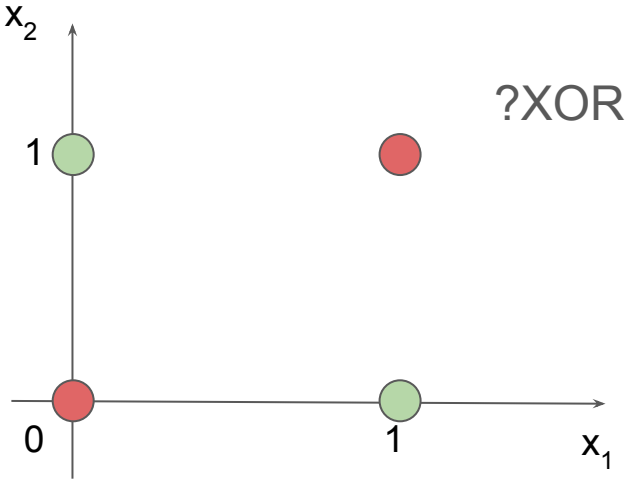


רשתות נוירונים פשוטות - תזכורת

- יחידת הבסיס (״נוירון״) - פרספטרון
- נדמיין פרספטרון שצריך לחשב AND
- וכעת פרספטרון שצריך לחשב OR
- מה לגבי XOR?

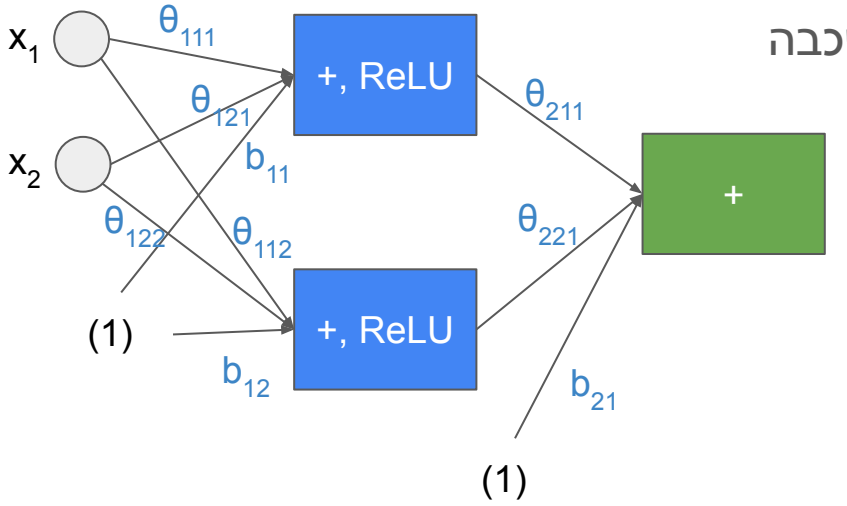


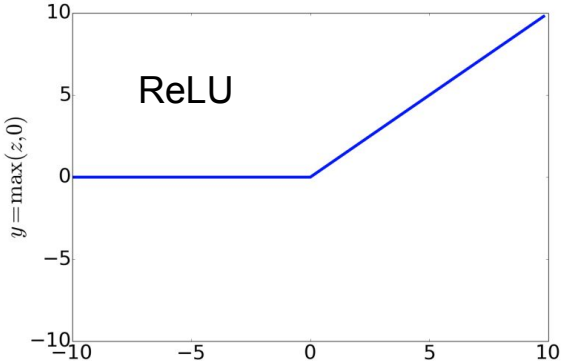




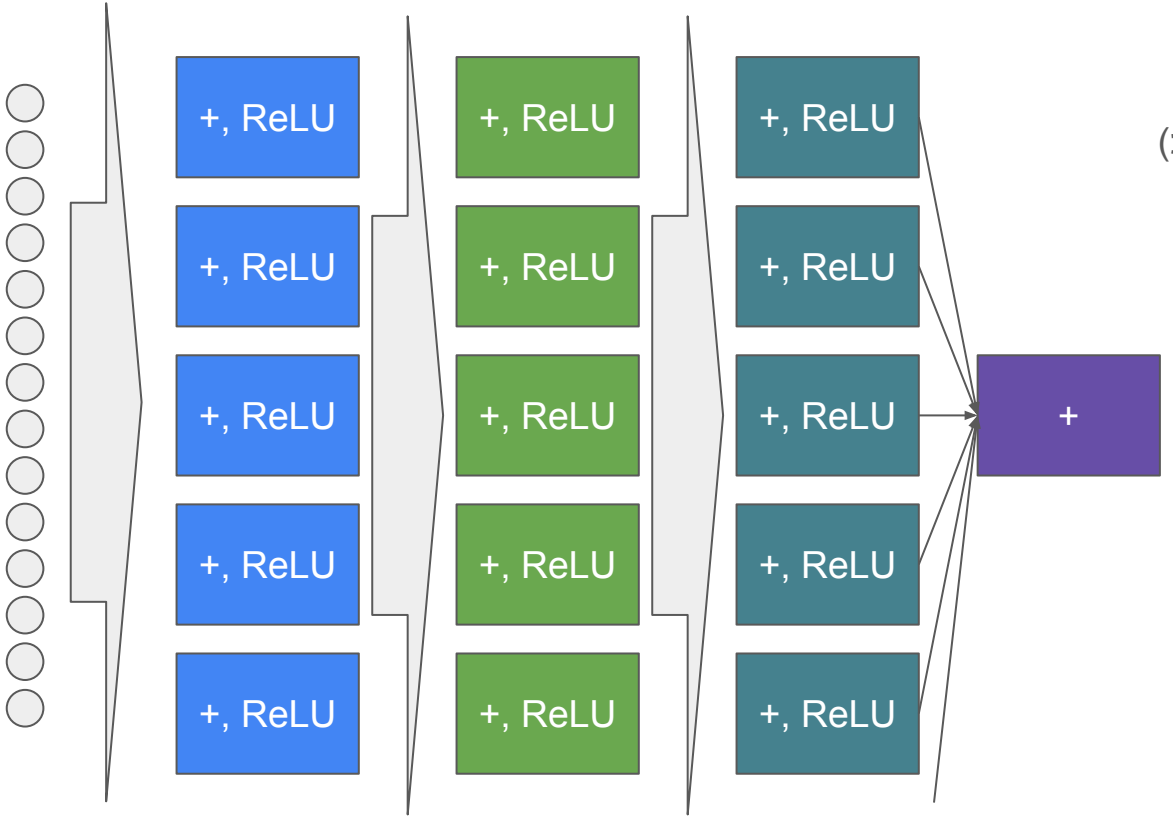
רשתות נוירונים פשוטות - תזכורת

- נרכיב שלושה פרספטרונים אחד על-גבי השני
- סתם חיבור לא יעזור (למה?) ולכן נוסיף **אקטיבציה** בשכבה האמצעית (השכבה הנחבאת, hidden layer)





רשתות נוירונים בהיזן קדמי (Feedforward NNs, FFN, MLP)

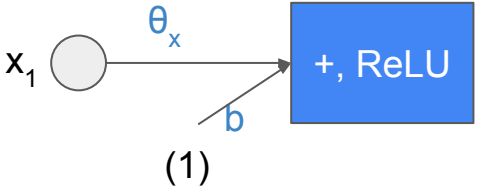


- אותו דבר רק בהמון
  - המון שכבות (= רשתות **עמוקות**)
  - המון פיצ׳רים
  - המון נוירונים חבויים בכל שכבה
- איך מאמנים אותן?
  - פעפוע אחורה - backpropagation
  - איך מאתחלים את המשקלות?
- איזו אקטיבציה בוחרים?
  - ReLU
  - סיגמויד
  - טנגנס היפרבולי tanh
- אילו פיצ׳רים מכניסים?

מגבלות ה-FFN

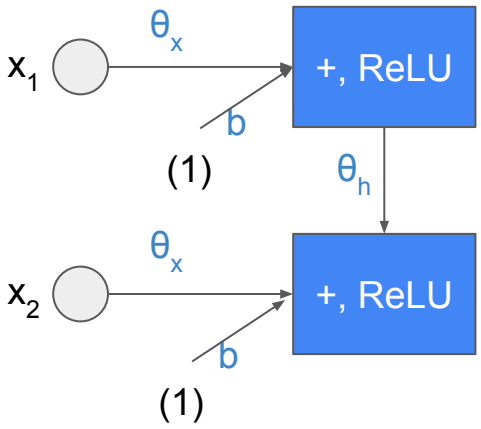
- התייחסות לכל הקלט בבת אחת
- חישובים שאינם תלויים ב**סדר** הקלט (bag-of-words)
- מוגבל לחיזוי תג אחד (מה נעשה עם תיוג רצפים?)

רשת נוירונים נשנית (Recurrent Neural Net - RNN)



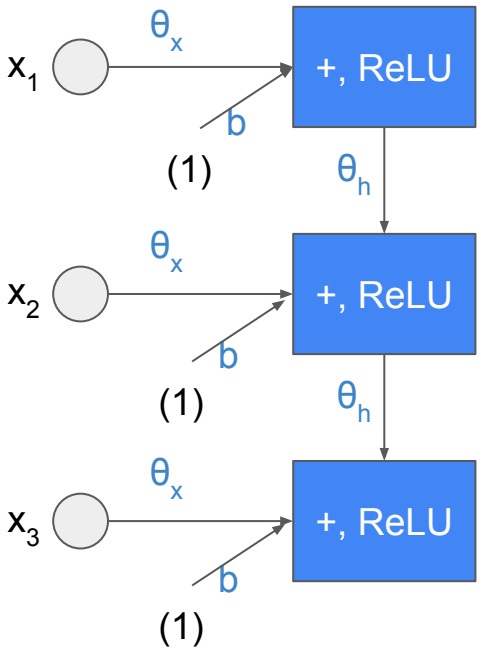
- נתחיל בדוגמה מקלט בעל נוירון אחד: המצב בתא הכחול **נשמר**

רשת נוירונים נשנית (Recurrent Neural Net - RNN)



- נתחיל בדוגמה מקלט בעל נוירון אחד: המצב בתא הכחול **נשמר**
- ו**מועבר** למצב הבא

רשת נוירונים נשנית (Recurrent Neural Net - RNN)



- נתחיל בדוגמה מקלט בעל נוירון אחד: המצב בתא הכחול **נשמר**
- ו**מועבר** למצב הבא
- וכך הלאה
- חשוב: הפרמטרים הם אותם פרמטרים
  - צורת ההצגה הזאת נקראת פרישה, גלילה unrolled

רשת נשנית - פורמלית

- קלטים x על-פני זמן 1,...,T
- מצב חבוי h

- קשר ההישנות recurrence relation:

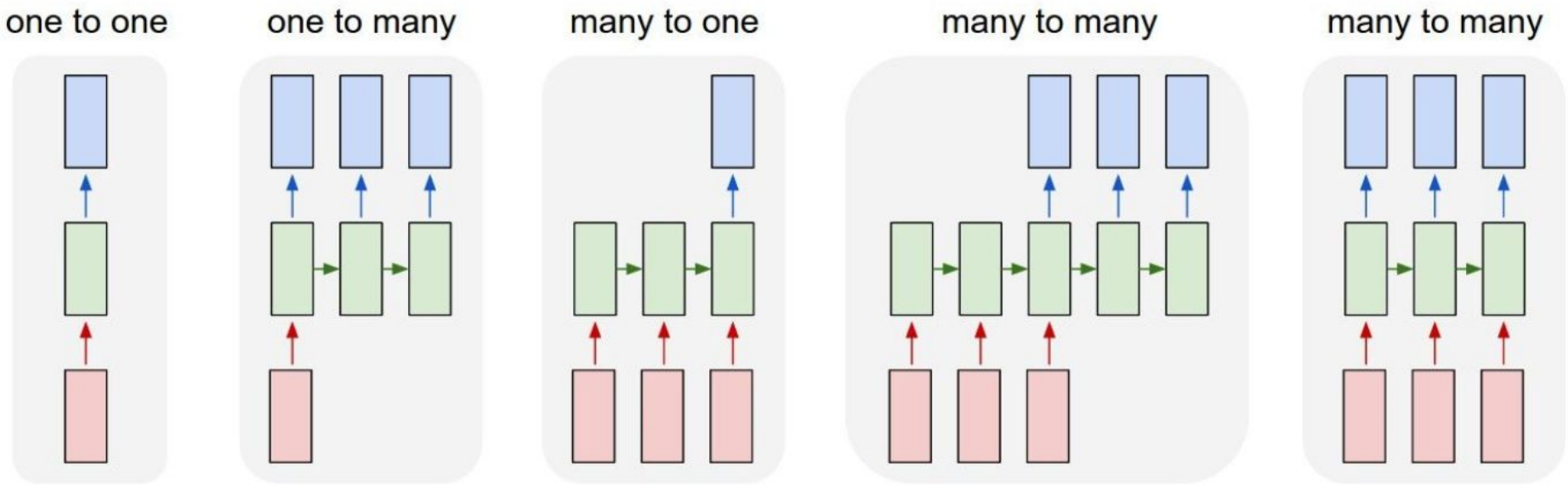
- $h_t = f(x_t, h_{t-1};\boldsymbol{\theta})$

- במקרה שראינו:

- $\boldsymbol{\theta}=\{\theta_x,\theta_h,b\}$; $f(x,h) = \theta_x x+\theta_h h+b$

- מקרים יותר סבוכים:
  - קלט שהוא בעצמו וקטור (כמעט תמיד)
    - ואז שכבת הפרמטרים היא...?
  - רשת עם f שדואג לא "לשכוח" מצבים רחוקים

מקרי שימוש ברשת נשנית



(לא רשת נשנית) סיווג

כתיבת הגדרה במילון

סיווג (שאינו BoW)

תרגום, תמצות

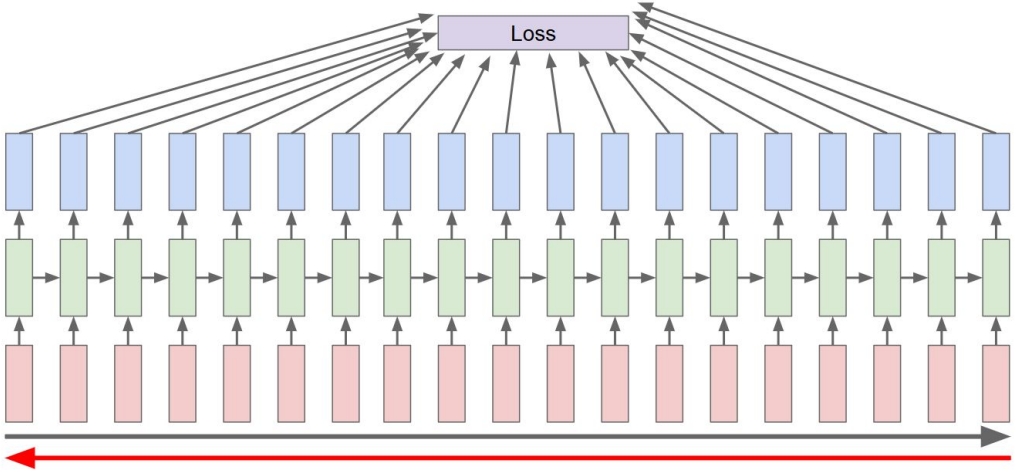
תיוג רצפים

(איזו עוד משימה מתחבאת כאן?)

אימון רשת נשנית

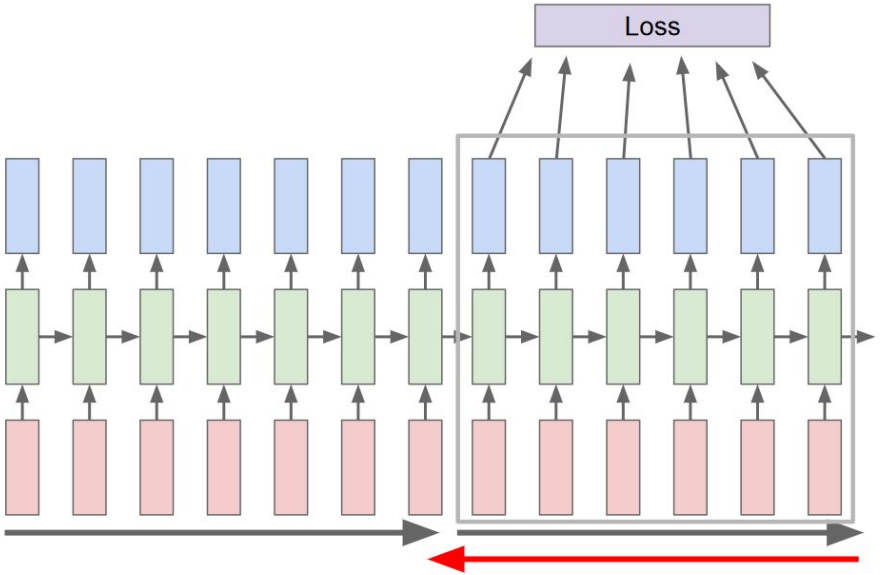
- אלגוריתם העדכון של רשת רגילה - backpropagation: מריצים את הרשת, מחשבים הפסד, מפעפעים לכל הפרמטרים ע״י כלל השרשרת

- ברשת נשנית, נריץ את כל הרשת ונפעפע את כל ההפסדים עבור רצף קלט **בבת אחת**: backpropagation through time



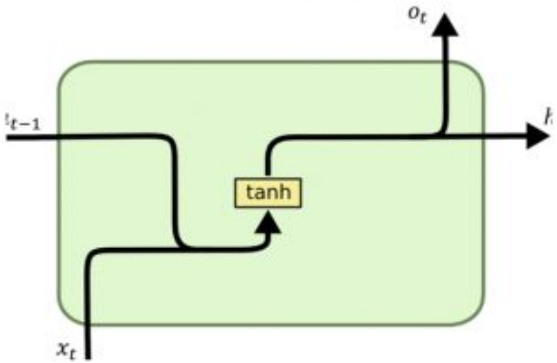
אימון רשת נשנית

- הרבה פעמים נחתוך את ההרצה באמצע ונפעפע בשלבים - truncated backpropagation through time



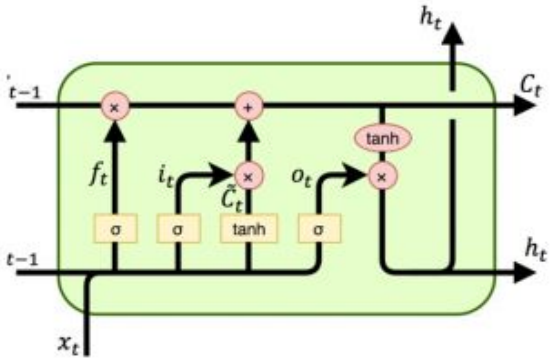
גרסאות לפונקציית ההישנות

RNN



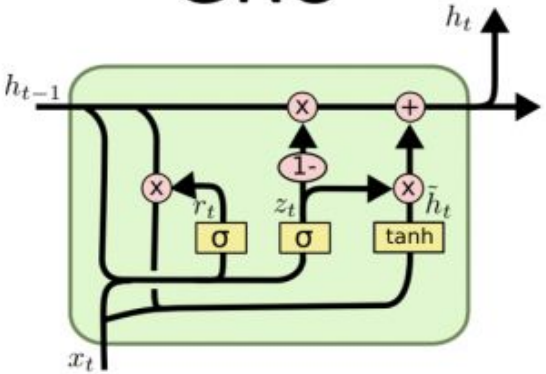
מה שראינו עד עכשיו, הכי פשוט
(ולכן נקרא גם "וניל" Vanilla RNN)
(נקרא גם על-שם ההוגה Elman RNN)

LSTM



מכיל מרכיבים שנועדו לשמור על מצבים
רחוקים (Long Short-Term Memory)
הכי נפוץ בשימוש כיום

GRU



פשרה בין שני הקיצוניים - Gated
Recurrent Unit שמצליח להסתכל
אחורה אבל מהיר בהרבה מ-LSTM

מקור התרשים:
http://dprogrammer.org/rnn-lstm-gru

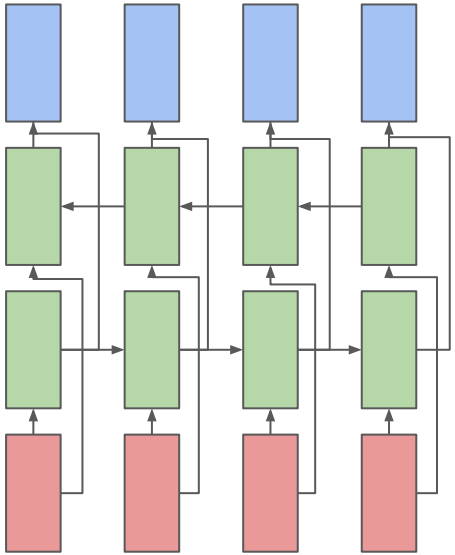
קלט וקטורי

- שיכוני מילים (Word embeddings) - נלמד באריכות עוד שבועיים
- בינתיים נתייחס לזה כאל "קופסה שחורה" - בהינתן מילה, מישהו חישב עבורה וקטור
- הווקטורים "מאוחסנים" בטבלאות
  - ויכולים בעצמם להיחשב לפרמטרים של המערכת

	e_1	e_2	e_3	e_4	e_5	e_6
שלום	0.07	-0.33	-0.73	0.95	1.10	0.64
עולם	0.11	-0.63	0.23	-0.12	-0.75	1.33
המונים	0.27	-0.12	0.11	0.43	-0.31	-0.21
4563	4.03	-1.13	0.92	1.34	0.09	-1.50
מד"ב	0.55	0.01	0.48	7.01	0.44	0.02

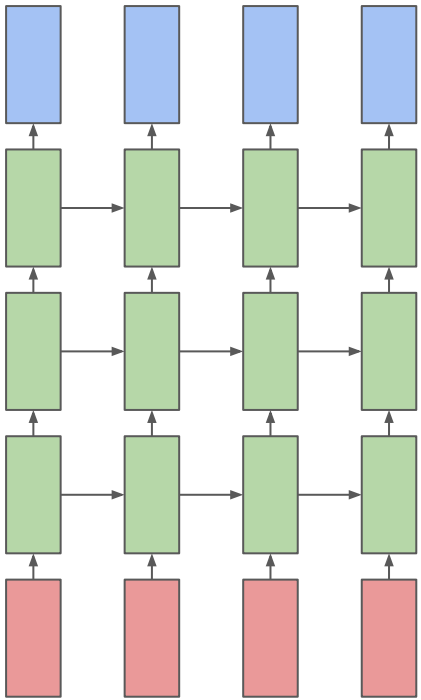
רשת נשנית דו-כיוונית (Bidirectional RNN)

- עבור משימות תיוג, הקשר "עתידי" חשוב כמו הקשר "מהעבר"
  - (עברית: אם המילה הבאה היא בבירור מילת יחס, הנוכחית כמעט בוודאות אינה מילת יחס)
- המבנה מבדיל בין הכיוונים - לכל כיוון יש את הפרמטרים שלו וה"חיבור" נעשה רק ברמה הבאה

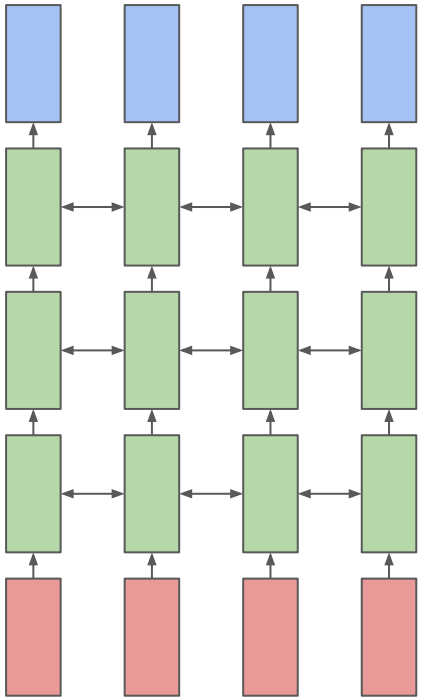


שכבות ברשת נשנית

- פלט משכבה אחת עולה כקלט לשכבה הבאה
- ברשת חד-כיוונית, אפשר למקבל את העיבוד
- ברשת דו-כיוונית חייבים לחכות לסוף המקטע עבור כל שכבה

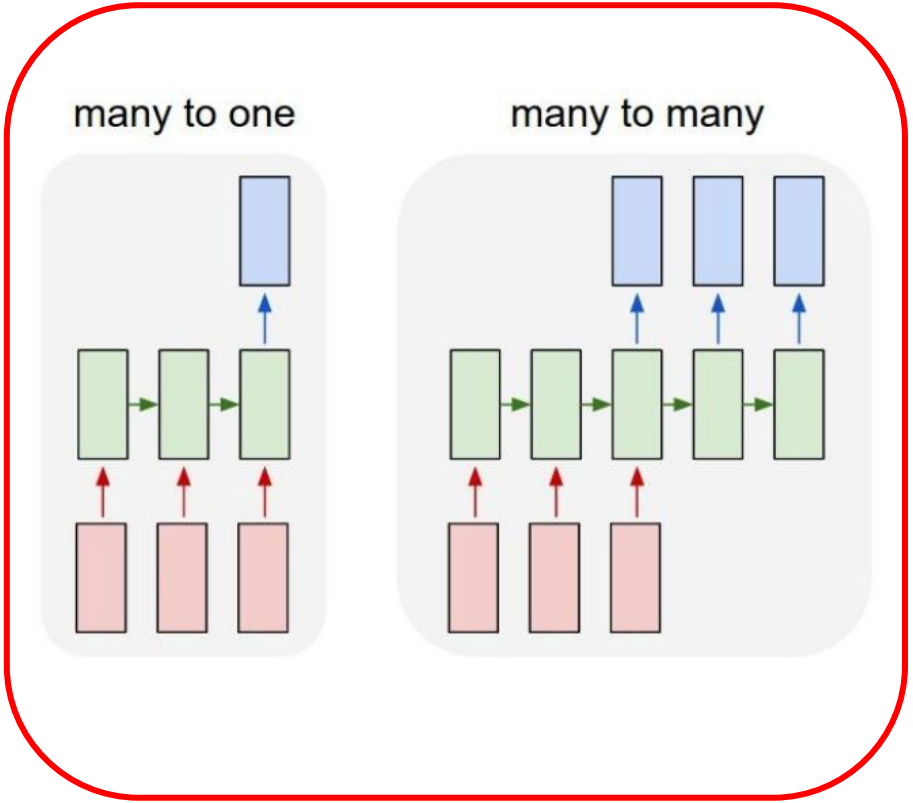


מימוש עבור תיוג



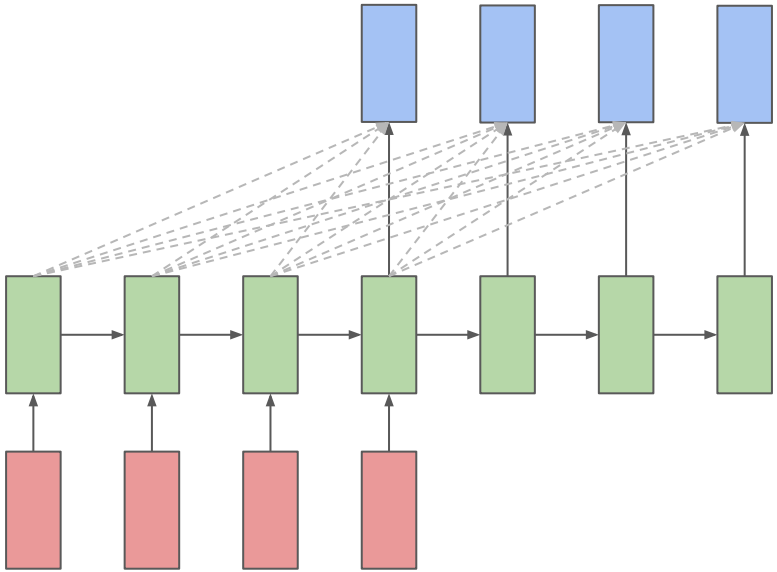
טיזר לנושא הנוירוני הבא-הבא

בשני המקרים האלה
הפרמטר החבוי ברשת
הנשנית צריך "לזכור" המון



טיזר לנושא הנוירוני הבא-הבא

- נלמד "לתת תשומת לב" לכל הקלט בשלבי החיזוי או היצירה
- רשתות צומי Attention models



למימושים - PyTorch

- https://pytorch.org/

- תמיכה בהרבה מאוד סוגי רשתות נוירונים (במודול `torch.nn`)
- בניה מודולרית של רשתות, הרצת הסקה (קדימה) ולמידה (אחורה) מובנית
- הדרכת פייטורץ' עם דוגמת סיווג מסמכים