

אוניברסיטת ת"א, ביה"ס להנדסת חשמל, למידת מכונה סטטיסטית

תרגיל בית 6

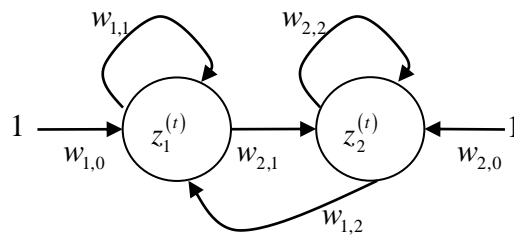
תרגיל בית זה עוסק ברשתות עצביות. התרגיל מורכב משני חלקים – חלק א' תיאורטי וחלק ב' שהינו תרגיל מחשב. יש להגיש כל חלק בנפרד, כל עבודה תיבדק בנפרד.

הגשה: עליכם להגיש קובץ **zip** שמכיל **PDF** עבור התרגיל התיאורטי, וקובץ **py**. עבור תרגיל המחשב. חובה לציין מספר ת.ז. בקבצי ההגשה (בקוד – בהערה בתחילתו).
תזכורת: מי שמגיש בזוג, יש להגיש פעם אחת בלבד תרגיל תיאורטי ופעם אחת תרגיל מחשב (ניתן לערבב זוגות).

חלק א' - שאלות תיאורטיות:

שאלה 1

Recurrent Backpropagation. נתונה רשת עיצבית רקורסיבית בעלת שני תאים כמתואר בציור.



מצבי התאים בזמן t הם $z_1^{(t)}$ ו- $z_2^{(t)}$. הכניסה לרשת היא הוקטור (x_1, x_2) , ואתחול מתבצע ע"י

$(z_1^{(0)}, z_2^{(0)}) = (x_1, x_2)$. לאחר מכן מריצים אותה 5 צעדי עדכון עפ"י הנוסחאות הבאות, עבור

$$t = 0, 1, 2, 3, 4$$

$$z_1^{(t+1)} = \sigma(w_{1,0} + w_{1,1}z_1^{(t)} + w_{1,2}z_2^{(t)})$$

$$z_2^{(t+1)} = \sigma(w_{2,0} + w_{2,1}z_1^{(t)} + w_{2,2}z_2^{(t)})$$

כאשר $\sigma(x) \triangleq \frac{1}{1+e^{-x}}$. המוצא של הרשת הוא הוקטור ממימד 2, $(y_1, y_2) \triangleq (z_1^{(5)}, z_2^{(5)})$. נסמן את

המוצא המבוקש של הרשת ע"י (t_1, t_2) .

רשמו אלגוריתם backpropagation לשערוך פרמטרי המודל מתוך L סדרות אימון,

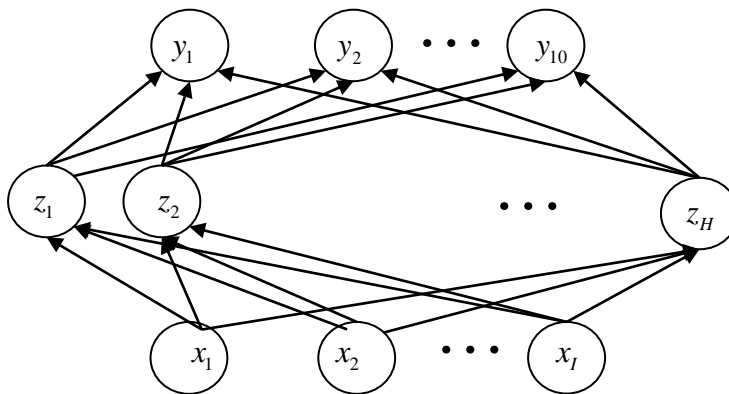
כאשר המטרה היא להביא למינימום את $\{x_1^{(l)}, x_2^{(l)}, t_1^{(l)}, t_2^{(l)}\}_{l=1}^L$

$$E(\mathbf{w}) \triangleq \frac{1}{2} \sum_{l=1}^L \left\{ \left(y_1^{(l)} - t_1^{(l)} \right)^2 + \left(y_2^{(l)} - t_2^{(l)} \right)^2 \right\}$$

הדרכה לפתרון התרגיל: נשים לב שניתן לתאר את פעולת הרשת באופן אקווילנטי ע"י רשת feedforward (כמו שלמדנו בכיתה) בעלת שכבת כניסה (2 תאים), 4 שכבות נסתרות (2 תאים כ"א) ושכבת מוצא (2 תאים). יש להראות כיצד ניתן לחשב באופן יעיל את הגרדיאנט של פונקציית המטרה המתאימה לרשת הזאת (נניח למשל אלגוריתם גרדיאנט מסוג batch). ההבדל בין הרשת הזאת לבין הרשת שלמדנו בכיתה הוא שעכשיו יש אילוצי שוויון בין משקולות של השכבות שונות. בעצם כל מה שצריך להראות הוא כיצד אלגוריתם הגרדיאנט שלמדנו בכיתה משתנה תחת אילוצי השוויון הללו בין המשקולות. מה שמתקבל זה שהמעבר הקדמי והמעבר האחורי זהים למה שלמדנו בכיתה. מה שמתשתנה זה נוסחת חישוב הגרדיאנט מתוך הערכים המתקבלים מהמעבר הקדמי ומהמעבר האחורי.

שאלה 2

נתונה רשת עיצבית feedforward לזיהוי תמונות סרוקות של 10 הספרות 0, 1, 2, ..., 9 שנכתבו בכתב יד. ארכיטקטורת הרשת מתוארת בציור הבא.



הרשת כוללת שכבת כניסה של I תאים, x_1, x_2, \dots, x_I , שהם הפיקסלים של התמונה הסרוקה, שכבה נסתרת של H תאים, בעלי מוצאים z_1, z_2, \dots, z_H , עם קישוריות מלאה לשכבת הכניסה (כולל איבר היסט לכל אחד מ- H התאים), ושכבה מוצא של 10 תאים, y_1, y_2, \dots, y_{10} , עם קישוריות מלאה לשכבה הנסתרת (כולל איבר היסט לכל אחד מ-10 התאים). התאים בשכבה הנסתרת מחשבים את המוצא שלהם

כמו שלמדנו בכיתה עם פונקציית אקטיבציה logistic sigmoid, $\sigma(x) \triangleq \frac{1}{1+e^{-x}}$. התאים בשכבת

המוצא מבצעים חישוב דומה, אבל ללא פונקציית אקטיבציה (לחילופין ניתן לאמר שהם משתמשים בפונקציית הזהות כפונקציית אקטיבציה).

נתון בסיס נתונים של N דוגמאות $\{\mathbf{x}^{(l)}, c^{(l)}\}_{l=1}^N$ כאשר $\mathbf{x}^{(l)}$ התמונה ה- l ו- $c^{(l)}$ ה-label המתאר

אותה. רוצים לאמן את משקלות הרשת ע"י אלגוריתם backpropagation שמביא למקסימום את

$$E(\mathbf{w}) \triangleq \sum_{l=1}^N E_l(\mathbf{w}) \triangleq \sum_{l=1}^N \log \frac{\exp(y_{c^{(l)}})}{\sum_{k=1}^{10} e^{y_k}}$$

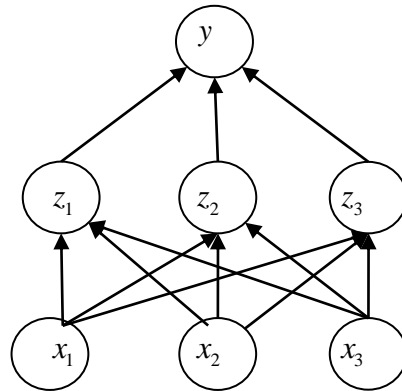
- א. הסבירו את המוטיבציה לקריטריון האימון. כיצד יתבצע הזיהוי?
- ב. איזה שינוי יש לבצע באלגוריתם backpropagation שלמדנו בכיתה כדי שיתאים לקריטריון האימון החדש?

חלק ב' – תרגיל מחשב

יש לציין בשורה הראשונה בקובץ הקוד את ת.ז ושמות המגשים, בצורת הערה.

יש להגיש קובץ אחד אשר כולל את הפתרון עבור שני הסעיפים, כאשר תופיע הערה בתחילת כל פתרון שמציין את מספר הסעיף.

ממשו את אלגוריתם backpropagation בגרסת batch gradient descent פשוט עם מקדם עדכון פרמטרים $\eta = 2$ (ללא גורם מומנטום), עבור בעיית Parity-3. בבעיה הזאת כל וקטור כניסה כולל שלושה משתנים בינאריים $x_1, x_2, x_3 \in \{0, 1\}$. המוצא המבוקש t הוא הזוגיות של וקטור הכניסה, ז"א $t = x_1 \oplus x_2 \oplus x_3$. בסיס הנתונים לאימון הוא כל 8 וקטורי הכניסה האפשריים והזוגיות שלהם. ארכיטקטורת הרשת מתוארת בציור הבא



הרשת כוללת 3 תאים בשכבת המבוא, שכבה נסתרת אחת בעלת 3 תאים ותא מוצא אחד. יש קישוריות מלאה בין השכבות כמתואר בציור. כל תא נסתר וכן תא המוצא משתמשים בפונקציית אקטיבציה logistic sigmoid, $\sigma(x) = (1 + e^{-x})^{-1}$. כל תא גם משתמש בפרמטר היסט שלא מתואר בציור

(לדוגמא, $z_2 = \sigma(w_{20} + w_{21}x_1 + w_{22}x_2 + w_{23}x_3)$).

א. הריצו את התוכנית 100 פעמים למשך 2000 איטרציות, כאשר ההבדל היחיד בין הרצה להרצה הוא הגרלה אחרת של פרמטרי הרשת, וציירו גרף של השגיאה הריבועית הממוצעת על פני 100

ההרצות, $E = \frac{1}{8} \sum_{i=1}^8 (y^{(i)} - t^{(i)})^2$, כפונקציה של אינדקס האיטרציה. המשקולות יאותחלו

בעזרת הגרלה i.i.d של המשקולות (וההיסטים) תוך שימוש בפילוג גאوسي סטנדרטי בעל תוחלת 0 ושונות 1.

ב. חזרו על המשימה עבור המקרה בו יש 6 תאים בשכבה הנסתרת וכל שאר הנתונים ללא שינוי.