

מבוא לסדרה מלאכותית - חלק 4

נושאים:

- סוגיית עניין ה-SVM למחנה
- רשתות נוירונים - Backpropagation

SVMs:

רק כדי להיות מכר: ניסוח הבעיה (הנאמר במקרים השונים -

$$\max_{\lambda \in \mathbb{R}^m} \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle : \text{Hard Margin}$$

$$\text{s.t. } \lambda_i \geq 0 \quad \forall i, \quad \sum_{i=1}^m \lambda_i y_i = 0$$

λ הם כובלי סטראונץ' (המקומות האלמנטים בעציה המקורית). יש כאן אינסוף
 של $\lambda_i \geq 0$. לא אכנס לסימול, ממה צריך לקרוא זאת.
 כאמור, הנוסחה Kernel מתחברת עם החלפת $\langle x_i, x_j \rangle$ ב- $K(x_i, x_j)$.
 זו בעיה לא פשוטה עבור אלמנטים של אופטימיזציה למינימום ריבועי.

$$\max_{\lambda \in \mathbb{R}^m} \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle : \text{Soft Margin}$$

$$\text{s.t. } 0 \leq \lambda_i \leq C$$

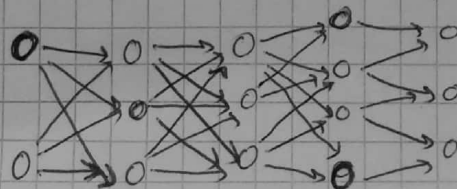
$$\sum_{i=1}^m \lambda_i y_i = 0$$

Box (Constraint) (שמוכנה גם) C כאן, λ_i מאותו סדר מצי C (שמוכנה גם) C
 שוב, אפשר לבדוק עם אלמנטים למינימום ריבועי.

רשתות נוירונים:

רשת נוירונים היא, באופן כללי, אוסף של נוירונים (קוויקווי גוף למקבצים ערכים, מעצבים
 כוויקווי פשוטו עיניה ומציאים פלט) ומחוברים זה לזה בקשתות ממוקצות.
 ואוסף של נוירונים מקבל ערכים מנוירונים אחרים, שממוקצים לפי קשתות, מחשב
 ערך ומציא אותו לנוירונים אחרים, אבל בלבד המערכת של נוירונים (שעושים דברים לא מסובכים)
 מסוגלת ~~לעצם פשוט~~ מסוגלת ולבנות פונקציות לא טריוויאליות.
 עז כאן טיול שבוטרים במבוא. עכשיו עומדים:

רשת נוירונים במבנה היא גרף מכוון (V, E) זר קשתות ממוקצות: $\mathbb{R} \rightarrow \mathbb{R}$
 ספציפית, נוסח ברשתות Feed-Forward - כאמור, מבין ערכים מעצבים.



לקודקודים קוראים גם נירונים. הם מסוגלים להשכבה, כאשר (בישטית פשוטית) כל שכבה מקבלת כשלים מהשכבה הקודמת (כדי להשכבה הראשונה, אינה זקוקה אל הקודם).

השכבה הראשונה היא שכבה הקלט, האחרונה היא שכבה הפלט ואם שנייה מכילה שכבות חבוי (hidden layers).

המטרה שלנו: להינתן גיף ופלט ולחזות את הקודקודים, למצוא את המסלולים והמחירים של הקודקודים. כך שהפלט מהיר יותר ורוב כל האפשרויות.

כמוצאם של הקודקודים:

כל קודקוד מקבל פלט מקודקוד בשכבה הקודמת, משקל לפי הקשר בין שני הקודקודים האלו. $(rel-1) r-d n$

נסמן את השכבה l , את אינדקס הקודקוד בשכבה n לא, הקשתות בין הקודקודים.

כ- $W_{r,n}$. אלו:

$$output_{l,n} = \sigma \left(\sum_{rel-1} W_{r,n} \cdot output_{l,r} \right)$$

הערות - יש גם $W_{0,l-1}$ (או, פשוט, b_l) - משקל הנוצר והפונקציה σ , או לא אכזרוב ככה, ובמקום זאת אנחנו שכל שכבה יש נירון בוסס שער הפלט שלו.

הוא נמצא b_l , שקבוצת b_l קראים פאזמים bias units.

הפונקציה σ היא פונקציה אקטיבציה (activation function), שמטרתה לקבוע כמה הנירון

"פעיל" בהיחסים לעצמים בהכרח והכאן.

פונקציות אקטיבציה שוליות: - סימטריה - $\frac{1}{1+e^{-x}}$ (ערכים בין 0 ל-1)

- טאנגנס היפרבולי (ערכים בין -1 ל-1)

- פונקציה זוגית

- Rectified Linear Unit (ReLU) - $\max(0, x)$.

- Exponential Linear Unit (ELU) - עבור $x < 0$, $\alpha(e^x - 1)$; עבור $x \geq 0$, x .

יש עוד המון. הדרך שבה יש גזירות (פאזמים), גזירות פירט גזירות (אם יש).

רשת נירונים מהווה מחלקת השלמות - $H_{V,E,\sigma} = \{h_{V,E,\sigma,w} : w \text{ שוקדים}\}$.

ולמשפט מיינר VC של רשת כאלה סופי. נוח לסמן, בהינתן גזירות (x, y) כ- x מציגה n ושכבה (פלט) מציגה k , או $h: \mathbb{R}^n \rightarrow \mathbb{R}^k$ פה n מציגה ופירט מציגה k הריש.

הכוון של רשת נירונים הוא שגזירותם אפשר לתאר פונקציות רבות אפילו וגם גזירותם.

~~הפונקציות הללו הן פונקציות רבות~~

איתון הרשת ע' Backpropagation

כאשר, המטרה שלנו היא למצוא את $L = \text{loss}(h_w(x), y)$ (כל- x ו- y כל בעקבות התנאים שיש לנו), נתאים ה- w . [אשר אם למצוא loss עם רגולריזציה בצורה דומה]

באיתון ע' Backpropagation, מה שמבצעים בפועל הוא:

בהינתן נתונים (x_i, y_i) או (x_i, y_i) ~~מסומנים~~ ^{מסומנים}, $\nabla L(\text{נתון } (x_i, y_i) \text{ מסומן})$ ואם מצביעים את w :

כאשר η גודל הלמידה (learning rate)

$$w^{(new)} = w^{(old)} - \eta \nabla L(\text{נתון } (x_i, y_i) \text{ מסומן})$$

איך מצביעים את הירידה?

1. התקדמות קדימה לאורך הרשת - בהינתן x , $\hat{y}$ מחשבים את כל הערכים של הנוירונים ברשת, נסמך $\hat{y} = \sigma(z_j)$

2. התקדמות אחורה - בשכבה האחרונה: $\delta_i = (y_i - \hat{y}_i)$ (תיוג - פסל)

ומחשבים לכל נירון את δ_j שלו באופן הבא:

$$\delta_j = \left(\sum_{k \text{ שכבה הסוקבת}} w_{jk} \delta_k \right) \cdot \sigma'(a_j)$$

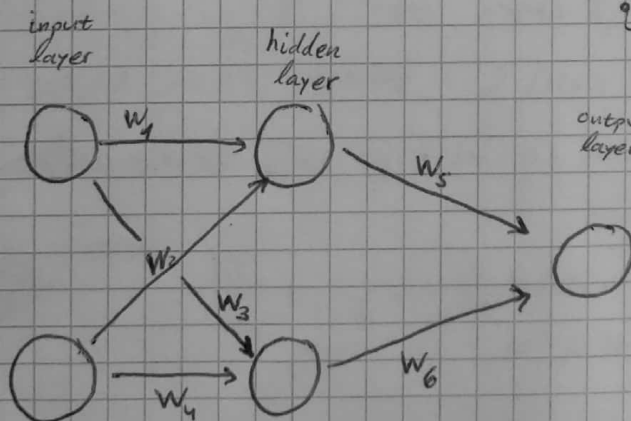
שכבה i לשכבה $j+1$ $w_{ij}^{(l)}$ הוא שקל קלט בין)

$$\frac{\partial L}{\partial w_{ij}^{(l)}} = z_j^{(l)} \delta_i^{(l+1)}$$

צוואתה בשורה:

תהי הרשת: (כלל bias)

עם פ' אקטביזם שהיא פ' הזוג.



הניחו משקלים מתחלתים של $w_1=0.11, w_2=0.21, w_3=0.12, w_4=0.08, w_5=0.14, w_6=0.15$

נתונה קדימה מהצגה: $x=(2,3), y=1$

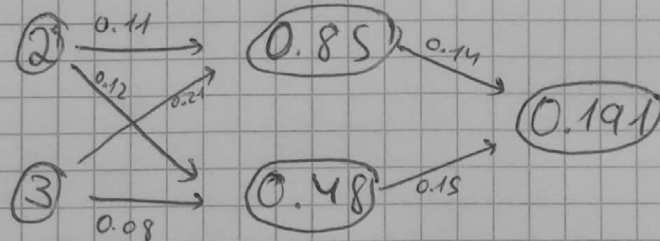
השתמשו ה- loss ריבועי ה- learning rate של 0.05 ומחצית את המוצא שלם אחת.

סדרון

1. מעבר קצונה - בעברת האינטוס נכנסים את הערכים הנתונים. יש לנו W היחסי, עכשיו אין

בעלות:

$$\begin{pmatrix} w_1 & w_2 \\ w_3 & w_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.11 & 0.21 \\ 0.12 & 0.08 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.85 \\ 0.48 \end{pmatrix}$$



2. מעבר אחורה - המקרה שלנו קדם: $\sigma'(x) = 1$ לכל x .

יש לנו 6 משקלים כחשב מחזק, קודם נחשב $\frac{\partial L}{\partial w_6}$.

$$\frac{\partial L}{\partial w_6} = -0.809 \cdot 0.48 \approx -0.388$$

המקרה של 5, 6, $\delta = (0.191 - 1)$

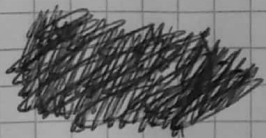
$$\frac{\partial L}{\partial w_5} = -0.809 \cdot 0.85 \approx -0.68765$$

המקרה של 1, 2, 3, 4, הערך z הוא ערך הקדם.

הערך δ הממדים הוא: $w_j \cdot \delta$ כל-כ-5 הוא הערך מהשכבה הקודמת

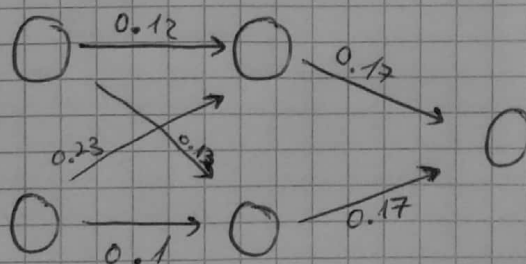
1- w_j המשקל של היות היות את המקודמים (בעברת הבאר, שישו עם אים

ית, אחת כל 150 עכשיו קודם כה).



$$\frac{\partial L}{\partial w_4} = w_6 \cdot \delta \cdot x_2 = 0.15 \cdot (-0.809) \cdot 3 \approx -0.364$$

עכשיו, מעשינו את w : $w^{new} = w^{old} - 0.05 \frac{\partial L}{\partial w}$ (משקלים בקורות):



כל מעשים מחזק את אותו אינטוס, הערך האינטוס הוא 0.26, פאור
התקדמו בכיוון הנכון!