16 1/1/17

נושאים מתקדמים בראייה חישובית - הרצאה #12

נושאי השיעור:

- Back prop
- Alex net

תזכורת: רשת היא אוסף של משקולות - $w_{ij}^{(l)}$ (פירוט בשיעור קודם)

עדכון למשקולות נעשה במטרה להקטין את השגיאה שמתקבלת לאחר העדכון.

$$\Delta w = \frac{\partial e(w)}{\partial w_{ij}^{(l)}} = ?$$

$$e(w) = (x_1^{(L)} - y)^2$$

$$\frac{\partial e(w)}{\partial w_{ij}^{(l)}} = \underbrace{\frac{\partial e(w)}{\partial s_j^{(l)}}}_{\delta_j^{(l)}} \cdot \underbrace{\frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}}}_{x_i^{(l-1)}}$$

(תזכורת: $s_j^{(l)} = \sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} x_i^{(l-1)}$)

$$\Rightarrow \Delta w = \delta_j^{(l)} \cdot x_i^{(l-1)}$$

* עבור השכבה האחרונה: $(l=L, j=1)$:

$$\delta_j^{(l)} = \frac{\partial e(w)}{\partial s_j^{(l)}}$$

$$\Rightarrow \delta_1^{(L)} = \frac{\partial (e(w))}{\partial s_1^{(L)}}$$

נתון:

$$e(w) = (x_1^{(L)} - y)^2 = (\theta(s_1^{(L)}) - y)^2$$

מכיוון ש: $\theta'(s) = 1 - \theta^2(s)$ נציב ונקבל:

$$\delta_1^{(L)} = \left(1 - x_1^{(L)2}\right)\left(x_1^{(L)} - y\right)$$

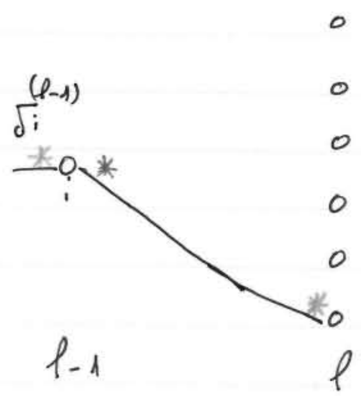
* הביטוי אמור להיות מוכפל פי 2 (מהנגזרת) אבל זה לא קריטי כי אפשר לשחק עם ה- learning rate.

16/1/17

(*) עבור שכבת ביניים כלשהי :

$$\delta_i^{(l-1)} = \frac{\partial(e(w))}{\partial s_i^{(l-1)}} = \dots$$

$$\dots = \sum_{j=1}^{d^{(l)}} \underbrace{\frac{\partial e(w)}{\partial s_j^{(l)}}}_{\text{השגיאה בשכבה ה-}l.} \cdot \underbrace{\frac{\partial s_j^{(l)}}{\partial x_i^{(l-1)}}}_{\text{השפעה של } x_i^{(l-1)} \text{ במוצא הנוירון ה-}i \text{ בשכבה } l-1.} \cdot \underbrace{\frac{\partial x_i^{(l-1)}}{\partial s_i^{(l-1)}}}_{\text{מעבר דרך הפונק' הלא ליניארית.}} =$$



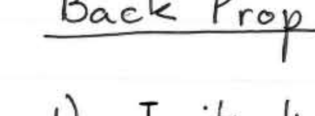
$$= \sum_{j=1}^{d^{(l)}} \delta_j^{(l)} \cdot w_{ij}^{(l)} \, \theta'(s_i^{(l-1)}) =$$

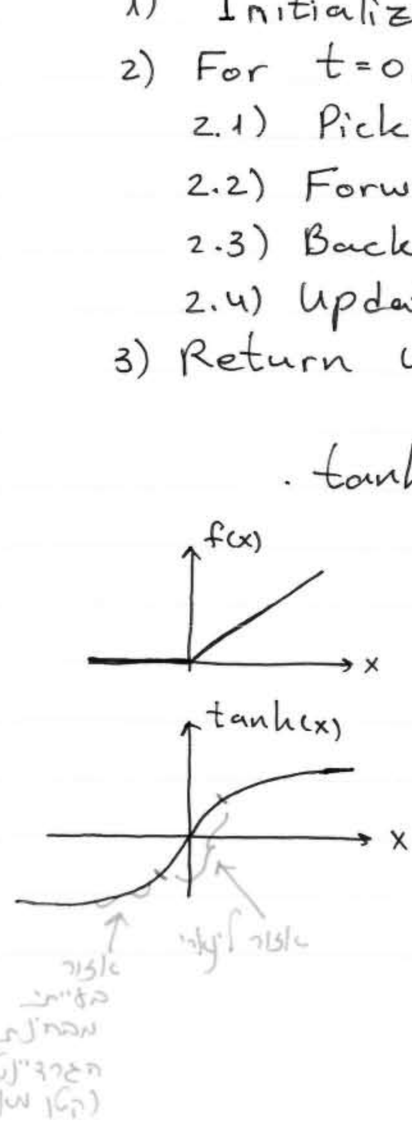
$$= \left(1 - x_i^{(l-1)^2}\right) \sum_{j=1}^{d^{(l)}} w_{ij}^{(l)} \delta_j^{(l)}$$

Back Prop Algorithm:

1) Initialize $w_{ij}^{(l)}$
2) For $t = 0 \dots T$
   2.1) Pick $n \in \{1, \dots, N\}$
   2.2) Forward : compute $x_j^{(l)}$
   2.3) Backward : compute $\delta_j^{(l)}$
   2.4) Update : $w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta x_i^{(l-1)} \delta_i^{(l)}$
3) Return $w_{ij}^{(l)}$

ב-AlexNet עבדו עם ReLU במקום tanh.
$f(x) = \max(x, 0)$





בעיה של tanh היא Vanashing gradients