

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

1η Εργασία

Στην παρακάτω εικόνα φαίνονται οι μαθηματικές σχέσεις των παραγώγων που χρησιμοποίησα για τον πίνακα $W^{(1)}$.

$$\text{Συνάρτηση κόστους: } E(w) = \sum_{n=1}^N \sum_{k=1}^K t_{nk} \log y_{nk} - \lambda \|w\|^2$$

$$\text{απλοποιείται σε: } E(w) = \sum_{n=1}^N \left[\left(\sum_{k=1}^K t_{nk} w_k^{(2)T} z_n \right) - \log \left(\sum_{j=1}^K e^{w_j^{(2)T} \cdot z_n} \right) \right] - \frac{\lambda}{2} \sum_{k=1}^K \|w_k\|^2$$

Έχουμε η h είναι η συνάρτηση ενεργοποίησης του κρυφού επιπέδου.

$$\text{Θέτω: } a = w_j^{(1)T} x_n$$

Από τον κανόνα της αλυσίδας η παράγωγος της $E(w)$ ως προς $w^{(1)}$ είναι η εξής:

$$\frac{dE}{dw^{(1)}} = \frac{dE}{dz} \cdot \frac{dz}{da} \cdot \frac{da}{dw^{(1)}}$$

Όπου αναλυτικά έχουμε:

$$\frac{dE}{dz} = t_k w_k^{(2)T} - y_k w_k^{(2)T} = (t_k - y_k) w_k^{(2)T} = (\tau - \gamma) w^{(2)T}$$

$$\frac{dz}{da} = H', \text{ όπου } H' \text{ ο πίνακας με τις παραγώγους των εισόδων του κρυφού επιπέδου με την συνάρτηση } h.$$

$$\frac{da}{dw^{(1)}} = X.$$

$$\text{Άρα: } \frac{dE}{dw^{(1)}} = ((\tau - \gamma) w^{(2)T} \cdot H')^T \cdot X - \lambda w^{(1)}$$

Οι παράγωγοι για τις παρακάτω συναρτήσεις ενεργητικότητας είναι οι εξής:

$$\bullet \quad h(a) = \log(1+e^a) \Rightarrow h'(a) = \frac{1}{1+e^a} (1+e^a)' = \frac{e^a}{1+e^a}$$

$$\bullet \quad h(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \Rightarrow h'(a) = \frac{(e^a - e^{-a})'(e^a + e^{-a}) - (e^a - e^{-a})(e^a + e^{-a})'}{(e^a + e^{-a})^2}$$

$$\Rightarrow h'(a) = \frac{(e^a - e^{-a})^2 - (e^a - e^{-a})^2}{(e^a + e^{-a})^2} \Rightarrow h'(a) = 1 - \frac{(e^a - e^{-a})^2}{(e^a + e^{-a})^2}$$

$$\bullet \quad h(a) = \cos(a) \Rightarrow h'(a) = -\sin(a)$$