

# Ori Cohen

ID: 207375783

$$1) \text{ a) } P(y=i | x_t) = \frac{\text{softmax}(w_i x_t + b_i)}{\sum_j^k e^{w_j x_t + b_j}}$$

לפנינו מופיעות  $j$  ו  $i$  כפונקציות של  $x_t$

מינימום מושג בדעתם (במקרה של מינימום אחד)

(?con) bias  $\rightarrow$  עליה למטה

b)

לפנינו מושג מינימום של האנרגיה

likelihood  $\rightarrow$  לוג  $\rightarrow$  מינימיזציה

loss  $\rightarrow$  מינימיזציה של האנרגיה

לפנינו מושג מינימום של האנרגיה

$$\min_w \arg \frac{1}{m} \sum_{i=1}^m -\log(P(Y_i=y | X_i=x)) =$$

$$\min_w \arg \sum_i -\log(\text{softmax}(w x + b)_y) =$$

...  
...  
 $w = w[y], b = b[y]$

$$= \min_w \arg \sum_i -\log \left( \frac{e^{w x + b}}{\sum_j e^{w[j] x + b[j]}} \right)$$

1)

6)

የኢትዮጵያውያንድ በኋላ እንደሚከተሉ ስለመስጠት የሚገኘውን የሚከተሉት ደንብ ነው:

$$\begin{aligned}
 \nabla_{w_i} \text{loss}(w, b) &= \nabla_{w_i} \frac{1}{m} \sum_{t=1}^m -\log(p(Y_t=y | X_t=x)) \\
 &= \frac{1}{m} \sum_{t=1}^m \nabla_{w_i} -\log(p(Y_t=y | X_t=x)) = \frac{1}{m} \sum_{t=1}^m \nabla_{w_i} -\log\left(\frac{e^{w_i x + b_y}}{\sum_{j=1}^n e^{w_j x + b_j}}\right) \\
 &= \frac{1}{m} \sum_{t=1}^m -\nabla_{w_i} \log(e^{w_i x + b_y}) + \nabla_{w_i} \log\left(\sum_{j=1}^n e^{w_j x + b_j}\right) \\
 &= \frac{1}{m} \sum_{t=1}^m -\nabla_{w_i} (w_i x + b_y) + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \nabla_{w_i} \left(\sum_{j=1}^n e^{w_j x + b_j}\right) \\
 &\quad " \lambda \in \mathbb{R}" \\
 t \text{ (גנָה שולחן ורְאֵבָה שֶׁ (יְהוָה) וְסִדְרָה וְזַהֲרָה)}
 \end{aligned}$$

$$= \frac{1}{m} \sum_{t=1}^m \text{loss}_+(w; b),$$

$$\begin{aligned}
 \text{loss}_f(w; b) &= -\nabla_{w_y} (w_y x + b_y) + \frac{1}{\sum_{j=1}^n w_j x + b_j} \nabla_{w_y} \left( \sum_{j=1}^n e^{w_j x + b_j} \right) \\
 &= -x + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \left( \sum_{j=1}^n \nabla_{w_y} e^{w_j x + b_j} \right) = -x + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \left( \nabla_{w_y} e^{w_0 x + b_0} \right. \\
 &\quad \left. + \dots + \nabla_{w_y} e^{w_n x + b_n} \right) = x \left( -1 + \frac{e^{w_0 x + b_0}}{\sum_{j=1}^n e^{w_j x + b_j}} \right) = x (\text{softmax}_{w,y}(w x + b) - 1)
 \end{aligned}$$

$$\text{loss}(w; b) = -\nabla_{w_i} (\omega_y x + b_y) + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \nabla_{w_i} \left( \sum_{j=1}^n e^{w_j x + b_j} \right)$$

:(2. ကျမ်းမားပါ

$$= \frac{e^{w_i x + b_i}}{\sum_{j=1}^n e^{w_j x + b_j}} x = \text{softmax}(wx + b)_{[i]} \cdot x$$

:(မြန်မာစာပေါ် ၁၂။၃)

$$\nabla_{w_i} \text{loss}(w; b) = \begin{cases} x(\text{softmax}_{y \neq i}(wx + b) - 1) & : i=y \\ \text{softmax}(wx + b)_{[i]} x & : i \neq y \end{cases}$$

: bias ကျမ်းမားပါ၏ မှတ်မှတ်

$$\nabla_{b_i} \text{loss}(w; b) = \nabla_{b_i} \frac{1}{m} \sum_{t=1}^m -\log(P(Y_t = y | X_t = x))$$

$$= \frac{1}{m} \sum_{t=1}^m \nabla_{b_i} -\log(P(Y_t = y | X_t = x)) = \frac{1}{m} \sum_{t=1}^m \nabla_{b_i} -\log\left(\frac{e^{w_y x + b_y}}{\sum_{j=1}^n e^{w_j x + b_j}}\right)$$

$$= \frac{1}{m} \sum_{t=1}^m -\nabla_{b_i} \log(e^{w_y x + b_y}) + \nabla_{b_i} \log\left(\sum_{j=1}^n e^{w_j x + b_j}\right)$$

$$= \frac{1}{m} \sum_{t=1}^m -\nabla_{b_i} (\omega_y x + b_y) + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \nabla_{b_i} \left( \sum_{j=1}^n e^{w_j x + b_j} \right)$$

ပြုလုပ်ရန် ပါ။

t မြန်မာစာပေါ် ၁၂။၃ မှတ် အောင် ၁

t မြန်မာစာပေါ် ၁၂။၃ မှတ် အောင် ၂

$$-\nabla_{b_y} (\omega_y x + b_y) + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \nabla_{b_y} \left( \sum_{j=1}^n e^{w_j x + b_j} \right)$$

:(1 ကျမ်းမားပါ)

$$= -1 + \frac{1}{\sum_{j=1}^n e^{w_j x + b_j}} \left( \nabla_{b_y} e^{w_0 x + b_0} + \dots + \nabla_{b_y} e^{w_n x + b_n} \right)$$

$$= -1 + \frac{e^{w_y x + b_y}}{\sum_{j=1}^n e^{w_j x + b_j}} = \text{softmax}(wx + b)_{[y]} - 1$$

-1 o  $\text{softmax}$  ת'ם קיימ' לילו ; (2 מילון  
: loss ~ $\epsilon \cdot \delta^2$ )

$$\nabla_{b_i} \text{loss}_+ (\omega; b) = \begin{cases} \text{softmax}(\omega x + b)_{[y]} - 1 & : i=y \\ \text{softmax}(\omega x + b)_{[y]} & : i \neq y \end{cases}$$

: loss () SB זלען זילען לא

$$SGD: \omega_i = \eta \nabla_{\omega_i} \text{loss}(\omega; b)$$

W פונקציית נספ'ת בפ' נספ'ת

$$b_i = \eta \nabla_{b_i} \text{loss}(\omega; b)$$

... bias o מינימיזציה ב- loss

ר' ג' נספ'ת : loss () SB זלען זילען לא

$$\omega_i = \begin{cases} \eta \times (\text{softmax}_{[y]}(\omega x + b) - 1) & : i=y \\ \eta \cdot \text{softmax}(\omega x + b)_{[y]} \cdot x & : i \neq y \end{cases}$$

$$b_i = \begin{cases} \eta (\text{softmax}(\omega x + b)_{[y]} - 1) & : i=y \\ \eta \text{softmax}(\omega x + b)_{[y]} & : i \neq y \end{cases}$$

