

2. Softmax

1. Softmax

$$① \quad P(y=y | X=x) = \text{softmax}(W \cdot x + b)_y$$

$$= \frac{e^{W_i \cdot x + b_i}}{\sum_j e^{W_j \cdot x + b_j}}$$

(the probability of class i given x)

$$② \quad \text{Loss}(W, b) = \sum_t -\log(P(y=y_t | x_t))$$

$$= - \sum_t \log(P(y=y_t | x_t))$$

$$= - \sum_t \log(\text{softmax}(W x_t + b)_{y_t})$$

$$= - \sum_t \log\left(\frac{e^{W_{y_t} \cdot x_t + b_{y_t}}}{\sum_j e^{W_j \cdot x_t + b_j}}\right)$$

$$= - \sum_t W_{y_t} \cdot x_t - b_{y_t} - \log\left(\sum_{j=1}^n e^{W_j \cdot x_t + b_j}\right)$$

$$= \sum_t \log\left(\sum_{j=1}^n e^{W_j \cdot x_t + b_j}\right) - W_{y_t} \cdot x_t - b_{y_t}$$

③ SGD: $W^t \leftarrow W^{t-1} - \eta \frac{\partial}{\partial W^t}$

Loss \rightarrow W \rightarrow Loss is cost

$$-2 \log \left(\frac{e^{W_i x_t + b_i}}{\sum_{j=1}^K e^{W_j x_t + b_j}} \right) \rightarrow \text{cost}$$

$$= -\log(e^{W_i x_t + b_i}) + \log\left(\sum_{j=1}^K e^{W_j x_t + b_j}\right)$$

$$= -(W_i x_t + b_i) + \log\left(\sum_{j=1}^K e^{W_j x_t + b_j}\right)$$

b_j \rightarrow W_j \rightarrow cost

$$\frac{\partial}{\partial W_j} \left(-(W_i x_t + b_j) + \log\left(\sum_{j=1}^K e^{W_j x_t + b_j}\right) \right) : W_j \quad (1)$$

$$\Rightarrow \begin{array}{c|c} j=i & j \neq i \\ \hline = -x_t + \frac{e^{W_j x_t + b_j} \cdot x_t}{\sum_{j=1}^K e^{W_j x_t + b_j}} & = \frac{e^{W_j x_t + b_j} \cdot x_t}{\sum_{j=1}^K e^{W_j x_t + b_j}} \end{array}$$

$$\frac{\partial}{\partial b_j} \left(-(W_i x_t + b_i) + \log\left(\sum_{j=1}^K e^{W_j x_t + b_j}\right) \right) : b_j \quad (2)$$

$$= -1 + \begin{array}{c|c} i=j & i \neq j \\ \hline \frac{e^{W_j x_t + b_j}}{\sum_{j=1}^K e^{W_j x_t + b_j}} & = \frac{e^{W_j x_t + b_j}}{\sum_{j=1}^K e^{W_j x_t + b_j}} \end{array}$$

2.10

: i=3 728

$$w_j = -x_t + \frac{e^{w_j x_t + b_j}}{\sum e^{w_j x_t + b_j}} = -x_t + \text{SoftMax}(w x_t + b)_j \cdot x_t$$

: i=3 728

$$b_j = -1 + \frac{e^{w_j x_t + b_j}}{\sum e^{w_j x_t + b_j}} = -1 + \text{SoftMax}(w x_t + b)_j$$

: i=3 728

$$w_j = \frac{e^{w_j x_t + b_j}}{\sum e^{w_j x_t + b_j}} = \text{SoftMax}(w x_t + b)_j \cdot x_t$$

$$b_j = \frac{e^{w_j x_t + b_j}}{\sum e^{w_j x_t + b_j}} = \text{SoftMax}(w x_t + b)_j$$

728 728 728

ex2.py 728

wc 0.23M

