ECE C247 Homework 02

Chih-En Lin

2.

Let $a_i(x) = W_i^T x + b_i$, $\theta = \{W_i, b_i\}$, $i = 1, \cdots, c$

$$\underbrace{\mathcal{L}(\theta)}_{\text{likelihood}} = \prod_{i=1}^{m} p(y^{(i)} | x^{(i)}, \theta)$$

$$= \prod_{i=1}^{m} \text{softmax}_{y^{(i)}}(x^{(i)})$$

$$= \prod_{i=1}^{m} \frac{e^{a_{y^{(i)}}(x^{(i)})}}{\sum_{j=1}^{c} e^{a_j(x^{(i)})}}$$

$$\Rightarrow \log \mathcal{L}(\theta) = \sum_{i=1}^{m} \log \left[\frac{e^{a_{y^{(i)}}(x^{(i)})}}{\sum_{j=1}^{c} e^{a_j(x^{(i)})}} \right]$$

$$= \sum_{i=1}^{m} \left[a_{y^{(i)}}(x^{(i)}) - \log \left(\sum_{j=1}^{c} e^{a_j(x^{(i)})} \right) \right]$$

#

$$\Rightarrow \underset{\theta}{\text{argmax}} \log \mathcal{L}(\theta) = \underset{\theta}{\text{argmin}} \underline{-\log \mathcal{L}(\theta)}$$

negative log-likelihood ($\mathcal{L}$) we want in this question

Let $J(\theta) = -\log \mathcal{L}(\theta)$

$$\nabla_{w_i} J(\theta) = \sum_{k=1}^{m} \frac{e^{a_i(x^{(k)})}}{\sum_{j=1}^{c} e^{a_j(x^{(k)})}} x^{(k)} - \sum_{\{k : y^{(k)} = i,\ \forall k \in \{1,\cdots,m\}\}} x^{(k)}$$

$$= \sum_{k=1}^{m} \mathrm{softmax}_i(x^{(k)})\, x^{(k)} - \sum_{\{k : y^{(k)} = i,\ \forall k \in \{1,\cdots,m\}\}} x^{(k)}$$

#

$$\nabla_{b_i} J(\theta) = \sum_{k=1}^{m} \frac{e^{a_i(x^{(k)})}}{\sum_{j=1}^{c} e^{a_j(x^{(k)})}} - \sum_{\{k : y^{(k)} = i,\ \forall k \in \{1,\cdots,m\}\}} 1$$

$$= \sum_{k=1}^{m} \mathrm{softmax}_i(x^{(k)}) - \sum_{\{k : y^{(k)} = i,\ \forall k \in \{1,\cdots,m\}\}} 1$$

#