Lecture 25: Deep neural networks continued

Professor Ilias Bilionis

Loss functions for classification networks

Training a binary classifier network

Observed data : $x_{1:n} = (x_1, \dots, x_n)$

$y_{1:n} = (y_1, \dots, y_n)$, $y_i \in \{0, 1\}$

DNN w/ param. θ.

$f(x_i; \theta)$

Model :

$$p(y_i = 1 \mid x_i, \theta) = \text{sigm}(f(x_i; \theta)) = \frac{e}{1 + e^{f(x_i;\theta)}}$$

Train : Loss=? Maximum likelihood estimate => Loss

Likelihood : $p(y_i \mid x_i, \theta) = \begin{cases} \text{sigm}(f(x_i;\theta)), & \text{if } y_i = 1 \\ 1 - \text{sigm}(f(x_i;\theta)), & \text{if } y_i = 0 \end{cases}$

$$\Rightarrow p(y_i \mid x_i, \theta) = [\text{sigm}(f(x_i;\theta))]^{y_i} [1 - \text{sigm}(f(x_i;\theta))]^{1-y_i}$$

$$p(y_{1:n} \mid x_{1:n}, \theta) = \prod_{i=1}^{n} p(y_i \mid x_i, \theta)$$

$$= \prod_{i=1}^{n} (\dots)$$

Find params : $\max_\theta \log p(y_{1:n} \mid x_{1:n}, \theta)$

$$L(\theta) = -\log p(y_{1:n} \mid x_{1:n}, \theta)$$

$$= -\sum_{i=1}^{n} \{ y_i \, \text{sigm}(f(x_i, \theta)) + (1 - y_i)[1 - \text{sigm}(f(x_i, \theta))] \}$$

Cross-entropy

$$= -\frac{1}{n} \sum_{i=1}^{n} n \cdot (\dots)$$

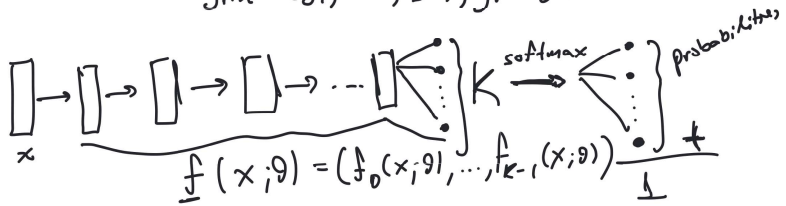
PREDICTIVE SCIENCE LABORATORY

Training a multi-class classifier network

Observed data : $x_{1:n} = (x_1, \dots, x_n)$

$y_{1:n} = (y_1, \dots, y_n)$; $y_i \in \{0, 1, \dots, K-1\}$

Model:



$$p(y_i = k \mid x_i, \vartheta) = \text{softmax}_k(\underline{f}(x_i;\vartheta))$$

$$= \frac{\exp\{f_k(x_i;\vartheta)\}}{\sum_{k'=0}^{K-1} \exp\{f_{k'}(x_i;\vartheta)\}}$$

PREDICTIVE SCIENCE LABORATORY

?

$$\sum_{k=0}^{K-1} p(y_i = k \mid x_i, \vartheta) = 1$$

$$p(y_i \mid x_i, \vartheta) = \text{soft}(\max_{y_i}(\underline{f}(x_i;\vartheta)))$$

$$p(y_{1:n} \mid x_{1:n}, \vartheta) = \prod_{i=1}^{n} p(y_i \mid x_i, \vartheta) = \prod_{i=1}^{n} (\dots)$$

$$L(\vartheta) = -\log p(y_{1:n} \mid x_{1:n}, \vartheta)$$

$$= -\sum_{i=1}^{n} \log \text{soft}(\max_{y_i}(\underline{f}(x_i;\vartheta)))$$

?

$$= -\frac{1}{n} \sum_{i=1}^{n} \log(\dots)$$