Lecture 25: Deep neural networks continued

Professor Ilias Bilionis

Loss functions for classification networks

PREDICTIVE
SCIENCE LABORATORY

Training a binary classifier network

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

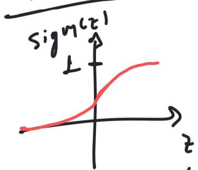
$y_{1:n} = (y_1, \ldots, y_n)$, $y_i \in \{0,1\}$

dense neural network

DNN w/ param. θ.

Model :

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



Train : Loss=? Maximum likelihood estimate => Loss

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

$$\Rightarrow p(y_i \mid x_i, \theta) = \left[\text{sigm}(f(x_i;\theta))\right]^{y_i \in \{0,1\}} \left[1 - \text{sigm}(f(x_i;\theta))\right]^{1-y_i}$$

PREDICTIVE
SCIENCE LABORATORY

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

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

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

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

log rules

$$= -\sum_{i=1}^{n} \underbrace{\left\{ y_i \,\text{sigm}(f(x_i,\theta)) + (1-y_i)\left[1 - \text{sigm}(f(x_i,\theta))\right] \right\}}_{\text{cross-entropy}}$$

$$= -\frac{1}{n} \sum_{i=1}^{n} n \cdot (\ \ldots \ )$$ — for stochastic optimization

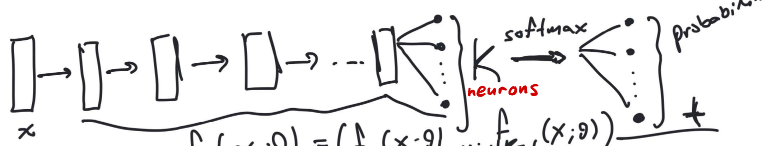
required when using batches (subsampling)

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:



$\underline{f}(x;\theta) = (f_0(x;\theta), \dots, f_{K-1}(x;\theta))$

vector function, K outputs

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

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

PREDICTIVE SCIENCE LABORATORY

?

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

likelihood of single observation

$$p(y_i \mid x_i, \theta) = \text{softmax}_{y_i}(\underline{f}(x_i;\theta))$$

likelihood of all observations

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

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

$$= -\sum_{i=1}^{n} \log \text{softmax}_{y_i}(\underline{f}(x_i,\theta))$$

?

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