DS 598 DEEP LEARNING

Sungjoon Park (BUID: U38522578)

Feb. 8th, 2024

Knowledge Check 5

1. (Problem 5.2)
   $<$ (i) when the training label $y = 0$ $>$

$$L = -\log[1 - sig[f[x, \phi]]]$$

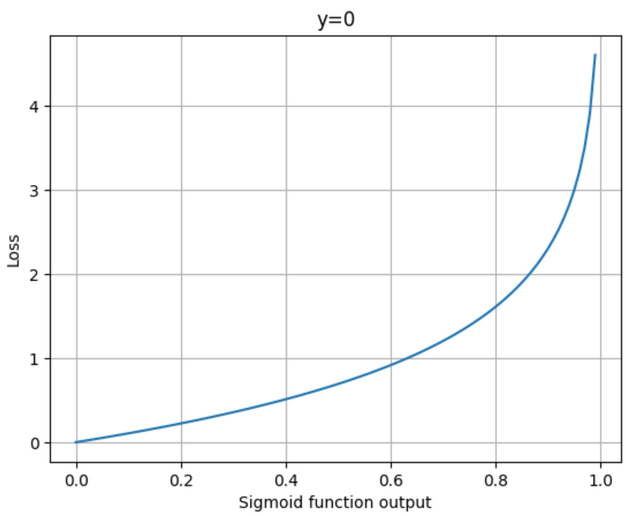

Figure 1: Loss as a function of the transformed network output when $y = 0$

$<$ (ii) when the training label $y = 1$ $>$

$$L = -\log[sig[f[x, \phi]]]$$

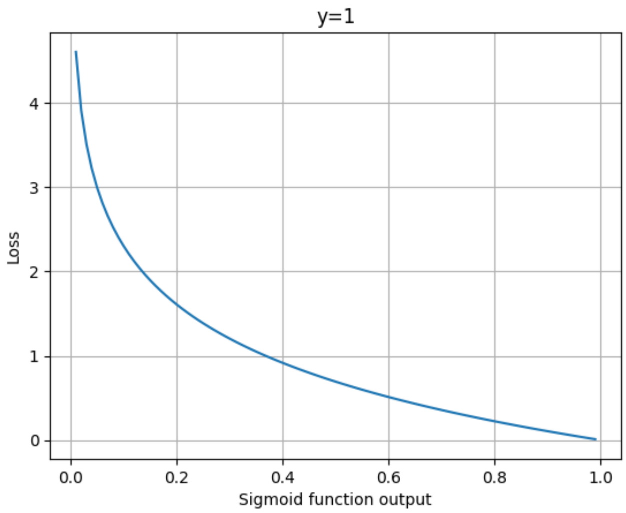

Figure 2: Loss as a function of the transformed network output when $y = 1$

2. (Problem 5.3)
   Probability Distribution Function for the target y: $Pr(y|\mu, \kappa) = \frac{\exp[\kappa \cos[y-\mu]]}{2\pi \cdot \text{Bessel}_0[\kappa]}$
   Distribution Parameter: μ
   Constant (Exogenous variable): κ
   Machine learning model: $\mu = f[\mathbf{x}, \phi] \rightarrow Pr(y|\mu, \kappa) = Pr(y|f[\mathbf{x}, \phi], \kappa)$
   By entering the probability distribution of the target y, we can construct a negative log-likelihood loss function like below over the training dataset pairs $\{x_i, y_i\}$.
   Loss function:

$$\begin{aligned} L(\phi) &= -\sum_{i=1}^{I} \log[Pr(y_i|f[\mathbf{x}_i, \phi], \kappa)] \\ &= -\sum_{i=1}^{I} \log[\frac{\exp[\kappa \cos[y_i - \mu_i]]}{2\pi \cdot \text{Bessel}_0[\kappa]}] \\ &= -\sum_{i=1}^{I} [\kappa \cos[y_i - \mu_i]] + \sum_{i=1}^{I} \log[2\pi \cdot \text{Bessel}_0[\kappa]] \end{aligned}$$

   Since the second term in the right-hand side is constant, the loss function can be simplified as follows.

$$
\begin{aligned}
L(\phi) &= -\sum_{i=1}^{I}[\kappa\cos[y_i - \mu_i]] \\
&= -\sum_{i=1}^{I}[\kappa\cos[y_i - f[\mathbf{x}_i, \phi]]]
\end{aligned}
$$

3. (Problem 5.6)
   Probability Distribution Function for the target y: $Pr(y = k|\lambda) = \frac{\lambda^k \exp(-\lambda)}{k!}$
   Distribution Parameter: λ
   Machine learning model: $\lambda = f[\mathbf{x}, \phi] \to Pr(y|\lambda) = Pr(y|f[\mathbf{x}, \phi])$
   Since λ must be non-negative, the output layer of the network f should consider it. By entering the probability distribution of the target y, we can construct a negative log-likelihood loss function like below over the training dataset pairs $\{x_i, y_i\}$.
   Loss function:

$$
\begin{aligned}
L(\phi) &= -\sum_{i=1}^{I}\log[Pr(y_i|f[\mathbf{x}_i, \phi])] \\
&= -\sum_{i=1}^{I}\log[\frac{f[\mathbf{x}_i, \phi]^{y_i}\exp(-f[\mathbf{x}_i, \phi])}{y_i!}] \\
&= -\sum_{i=1}^{I} y_i \log[f[\mathbf{x}_i, \phi]] + \sum_{i=1}^{I} f[\mathbf{x}_i, \phi] + \sum_{i=1}^{I}\log[y_i!]
\end{aligned}
$$

4. (Problem 5.7)
   Considering the independence of $\mathbf{y}$,
   (i stands for the indices of observations and j stands for indices of elements of $\mathbf{y}_i$.)
   Likelihood: $\prod_{i=1}^{N}\prod_{j=1}^{10}\frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{(y_{ij}-f[\mathbf{x}_{ij},\phi])^2}{2\sigma^2}]$
   Since constant terms, $\sum_{i=1}^{N}\sum_{j=1}^{10}\log[\frac{1}{\sqrt{2\pi\sigma^2}}]$ and $\sum_{i=1}^{N}\sum_{j=1}^{10} 2\sigma^2$, can be ignored,
   Negative Log-likelihood: $-\sum_{i=1}^{N}\sum_{j=1}^{10}[-(y_{ij}-f[\mathbf{x}_{ij}, \phi])^2] = \sum_{i=1}^{N}\sum_{j=1}^{10}(y_{ij}-f[\mathbf{x}_{ij}, \phi])^2$
   It is the same as the objective function of least squares, which is mean squared error (MSE) like below.
   MSE: $\sum_{i=1}^{N}\sum_{j=1}^{10}\epsilon_{ij}^2 = \sum_{i=1}^{N}\sum_{j=1}^{10}(y_{ij}-f[\mathbf{x}_{ij}, \phi])^2$, where $y_{ij} = f[\mathbf{x}_{ij}, \phi] + \epsilon_{ij}$