CIS4930 - Math for Machine Learning

Homework 6

Fernando Sckaff

April 11 2023

Question 1

You are given a set $T = \{T_1, T_2, \ldots, T_N\}$ of numbers. Given T, you compute a scalar quantity $w \equiv \frac{1}{\beta}\log\sum_{k=1}^{N}\exp\{\beta T_k\}$ where $\beta > 0$. What is

$$\lim_{\beta\to\infty} w?$$

Use L'Hôpital's rule (or equivalent) to relate this limiting value to a property of the set T. You are allowed to assume that the elements T_k of the set are distinct numbers.

Next, you compute a scalar quantity $z \equiv \sum_{k=1}^{N} T_k \frac{\exp\{\beta T_k\}}{\sum_{i=1}^{N}\exp\{\beta T_i\}}$ where $\beta > 0$. What is

$$\lim_{\beta\to\infty} z?$$

Relation this limiting value to a property of the set T. You are again allowed to assume that the elements T_k of the set are distinct numbers. Show all steps.

Solution:

Using the fact that $\exp\{\beta T_k\}$ grows much faster than 1 as $\beta \longrightarrow \infty$, we have that the only term in the sum that contributes to the limit is the one with the largest value of βT_k, say βT_M. Thus, we can rewrite the sum as

$$\textstyle\sum_{k=1}^{N}\exp\{\beta T_k\} \approx \exp\{\beta T_M\}.$$

Substituting this approximation into the expression for w, we have

$$w \approx \frac{1}{\beta}\log(\exp\{\beta T_M\}) = \frac{1}{\beta}(\beta T_M) = T_M.$$

Therefore, the limiting value of w as $\beta \longrightarrow \infty$ is the maximum element of the set T, i.e.,

$$\lim_{\beta\to\infty} w = \max(T).$$

Next, we consider the second expression,

$$z \equiv \sum_{k=1}^{N} T_k \frac{\exp\{\beta T_k\}}{\Sigma_{i=1}^{N}\exp\{\beta T_i\}}.$$

We can rewrite the denominator as

$$\Sigma_{i=1}^{N}\exp\{\beta T_i\} \approx \exp\{\beta T_M\},$$

where βT_M is again the largest value in the set βT_k. This is because the exponentials with smaller values of βT_k will be negligible compared to the exponential with βT_M as $\beta \longrightarrow \infty$. Substituting this approximation, we have

$$z \approx \frac{\sum_{k=1}^{N} T_k \exp\{\beta T_k\}}{\exp\{\beta T_M\}}.$$

Now we apply L'Hôpital's rule to the numerator and denominator to obtain

$$\begin{aligned}\lim_{\beta\to\infty} z &= \lim_{\beta\to\infty}\frac{\sum_{k=1}^{N} T_k \exp\beta T_k}{\exp\beta T_M} \\ &= \lim_{\beta\to\infty}\frac{T_M\exp\beta T_M}{\exp\beta T_M} \\ &= T_M\end{aligned}$$

Hence, $\lim_{\beta\to\infty}\frac{\sum_{k=1}^{N} T_k\exp\beta T_k}{\exp\beta T_M} = T_M$, where $T_M = \max(T_k)$ and this limit is achieved when β approaches infinity.

Question 2

For a sigmoid activation function $z(u) = \sigma(u) = \frac{1}{1+\exp\{-u\}}$, consider the following binary cross entropy (BCE) loss function:

$$\ell(y, z(u)) = y \log \frac{y}{z} + (1-y) \log \frac{1-y}{1-z}.$$

(a) Show that $\ell(y, z(u))$ can be written as

$$\ell(y, z(u)) = -yu + \log(1 + \exp\{u\}) + \text{terms independent of } u.$$

(b) Demonstrate this equivalence in PyTorch. Take the provided XOR code and show that a linear output with BCEWithLogitsLoss function gives the same results as a sigmoid output with BCELoss.

Solution:
(a) We have

$$\begin{aligned}
\ell(y, z(u)) &= y \log \frac{y}{z} + (1-y) \log \frac{1-y}{1-z} \\
&= -y \log z + (y-1) \log(1-z) + y \log y + (1-y) \log(1-y) \\
&= y \log(\frac{1-z}{z}) - \log(1-z) + y \log y + (1-y) \log(1-y) \\
&= y \log(\frac{\frac{1+\exp(-u)}{1+\exp(-u)} - \frac{1}{1+\exp(-u)}}{\frac{1}{1+\exp(-u)}}) - \log(\frac{1+\exp(-u)}{1+\exp(-u)} - \frac{1}{1+\exp(-u)}) + y \log y + (1-y) \log(1-y) \\
&= y \log(\frac{\exp(-u)(1+\exp(-u))}{1+\exp(-u)}) - \log(\frac{\exp(-u)}{1+\exp(-u)}) + y \log y + (1-y) \log(1-y) \\
&= y \log(\exp(-u)) + y \log(1+\exp(-u)) - y \log(1+\exp(-u)) - \log(\exp(-u)) + \\
&\log(1+\exp(-u)) + y \log y + (1-y) \log(1-y) \\
&= -yu + \log(1+\exp(u)) + y \log y + (1-y) \log(1-y) \quad 😄
\end{aligned}$$

(b) on Google Colab, Link:

https://colab.research.google.com/drive/1tw9S-G2PdBg1cMDcl9RaU9kepfDSB8Um?usp=sharing

Question 3

Show, using the Schwarz inequality,

$$\left(\sum_{k=1}^{N} |x_k y_k|\right)^2 \leqslant \sum_{k=1}^{N} |x_k|^2 \sum_{k=1}^{N} |y_k|^2$$

that for $c_n \geqslant 0, \sum_{n=1}^{N} c_n = 1$ and a vector $x \in \mathbb{R}^N$ with elements $x_n \geqslant 0, \sum_{n=1}^{N} c_n x_n^2 \geqslant (\sum_{n=1}^{N} c_n x_n)^2$. Show all steps.

Solution:
Consider, since $0 \leqslant c_n \leqslant 1$, then $c_n^2 \leqslant c_n$.

$$(\sum_{n=1}^{N} c_n x_n)^2 \leqslant \sum_{n=1}^{N} |c_n|^2 \sum_{n=1}^{N} |x_n|^2$$

However, $\sum_{n=1}^{N} c_n \geqslant \sum_{n=1}^{N} c_n^2$. Therefore,

$$\begin{aligned}\sum_{n=1}^{N} c_n x_n^2 &\geqslant \sum_{n=1}^{N} c_n^2 x_n^2 \\ &\geqslant (\sum_{n=1}^{N} c_n x_n)^2 \quad 😄\end{aligned}$$

$\boxed{\text{F.S.}}$

Note:-

- Question 1: Almost Correct. Incorrect application of L'Hopital.
- Question 2: Correct
- Question 3: 10/25. Incorrect use of the Schwarz inequality.