Machine Learning Exercise Sheet 09

Moritz Schüler
Matrikelnr: 03721689
December 13, 2019

---

1 Deep Learning

Problem 1

In machine learning you often come across problems which contain the following quantity

$$y = \log \sum_{i=1}^{N} e^{x_i} \tag{1}$$

For example if we want to calculate the log-likelihood of neural network with a softmax output we get this quantity due to the normalization constant. If you try to calculate it naively, you will quickly encounter underflows or overflows, depending on the scale of x_i. Despite working in log-space, the limited precision of computers is not enough and the result will be inf or $-\inf$.

To combat this issue we typically use the following identity:

$$y = \log \sum_{i=1}^{N} e^{x_i} = a + \log \sum_{i=1}^{N} e^{x_i - a} \tag{2}$$

for an arbitrary a. This means, you can shift the center of the exponential sum. A typical value is setting a to the maximum ($a = max_i\ x_i$), which forces the greatest value to be zero and even if the other values would underflow, you get a reasonable result.
Your task is to show that the identity holds.

$$
\begin{aligned}
\log \sum_{i=1}^{N} e^{x_i} &= a + \log \sum_{i=1}^{N} e^{x_i - a} \quad \mid e \\
\sum_{i=1}^{N} e^{x_i} &= e^{a} + \sum_{i=1}^{N} e^{x_i - a} \\
\sum_{i=1}^{N} e^{x_i} &= e^{a} + \sum_{i=1}^{N} e^{x_i} \cdot e^{-a} \quad | e^{-a} \text{ doesn't depend on the sum} \\
\sum_{i=1}^{N} e^{x_i} &= e^{a} + e^{-a} \cdot \sum_{i=1}^{N} e^{x_i} \quad \mid \log \\
\log \sum_{i=1}^{N} e^{x_i} &= a \log e - a \log e + \log \sum_{i=1}^{N} e^{x_i} \quad \text{with } \log a \cdot b = \log a + \log b \text{ and } \log a^{b} = b \cdot \log a \\
\log \sum_{i=1}^{N} e^{x_i} &= \log \sum_{i=1}^{N} e^{x_i}
\end{aligned}
$$

Problem 2

Similar to the previous exercise we can compute the output of the softmax function $\pi_i = \frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}}$ in a numerically stable way by shifting by an arbitrary constant a:

$$
\frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}} = \frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}} \tag{3}
$$

often chosen $a = max_i \ x_i$. Show that the above identity holds.

$$
\begin{aligned}
\frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}} &= \frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}} \quad \mid \log \\
\log e^{x_i} - \log \sum_{i=1}^{N} e^{x_i} &= \log e^{x_i - a} - \log \sum_{i=1}^{N} e^{x_i - a} \\
x_i - \log \sum_{i=1}^{N} e^{x_i} &= x_i - a - \log e^{-a} - \log \sum_{i=1}^{N} e^{x_i} \\
x_i - \log \sum_{i=1}^{N} e^{x_i} &= x_i - a + a - \log \sum_{i=1}^{N} e^{x_i} \\
x_i - \log \sum_{i=1}^{N} e^{x_i} &= x_i - \log \sum_{i=1}^{N} e^{x_i}
\end{aligned}
$$

Problem 3

Jupyter Notebook