Exercise 09 of Machine Learning [IN 2064]

Name: Yiman Li
Matr-Nr: 03724352
cooperate with Kejia Chen(03729686)

Problem 1

According to reference [1], underflow occurs when numbers near zero are rounded to zero, and overflow occurs when numbers with large magnitude are approaximated as ∞ or $-\infty$. For the function

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

We can consider what happens when all of the x_i are equal to some constant c. If c has very large magnitude, then exp(c) will over flow, which resulting in the expression as a whole being undefined. However, when translating the expression in the following form

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

By subtracting $a = \max x_i$ in every term, even through there are some values that are rounded zero due to underflow, it results in the largest argument to exp being zero, hence avoiding the problem of overflow, we can still get a stable solution.

Problem 2

In the softmax function, we have

$$\mathrm{softmax}(\boldsymbol{x}_i) = \frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}} \tag{3}$$

We can consider what happens when all of the x_i are equal to some constant c. Analytically, we can see that all of the outputs should be equal to $\frac{1}{n}$. Numerically, this may not occur when c has large magnitude. If c is very negative, then exp(c) will underflow. This means the denominator of the softmax will become zero, so the final result is undefined. When c is very large and positive, exp(c) will overflow, again resulting in the expression as a whole being undefined.

$$\mathrm{softmax}(\boldsymbol{x}_i) = \frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}} \tag{4}$$

When we instead evaluating $\mathrm{softmax}(\boldsymbol{z})$ where $\boldsymbol{z} = \boldsymbol{x} - a = \boldsymbol{x} - \max x_i$, just as shown in equation 4, both overflow and underflow problems can be solved. Simple algebra shows that the value of the softmax function is not changed analytically by adding or subtracting a scalar from the input vector. Subtracting $\max x_i$ results in the largest argument to exp being zero, which results out the possibility of overflow. Likewise, at least one term in the denominator has a value of 1, which rules out thr possibility of underflow in the denominator leading to a division by zero.

Problem 3

To be seen in the end.

References

[1] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.