Numerical Stability

Computing Log-Sum-Exp

This technical is well explained in this article under the context of computing log-likelihood:

$$LL = \log \sum_i \exp(x_i)$$

It is not difficult for $\exp(x_i)$ to overshoot the upper bound of `float` and returns `NaN`.

We can make x_i smaller by substracting a from it. Because

$$\exp(x_i - a) = \frac{\exp(x_i)}{\exp(a)}$$

we have

$$LL = \log \left(\exp(a) \cdot \sum_i \exp(x_i - a) \right) = a + \log \sum_i \exp(x_i - a)$$

It is straightforward to choose $a = \max(x_i)$.

Computing Softmax

In the softmax function/cost we need to computing a sum of exponentials on the denominator:

$$p(j) = \frac{\exp(x_j)}{\sum_i \exp(x_i)}$$

Again, it is easy for $\exp(x_i)$ to overshoot the upper bound of `float`, and we want a form $\exp(x_i - a)$. This time, we have

$$\frac{\exp(x_j - a)}{\sum_i \exp(x_i - a)} = \frac{\exp(a) \exp(x_j)}{\exp(a) \sum_i \exp(x_i - a)} = p(j)$$

Again, we can choose $a = \max(x_i)$.