2023/02/07

Exercise 3

Suppose that the output $\hat{y}_k$ of a given unit in a neural network is given by the softmax function *i.e.*:

$$\hat{y}_k = \frac{\exp(a_k)}{\sum_j \exp(a_j)}. \tag{2}$$

- Show that the output of the softmax function does not change if you shift, in all components, the activations a_j by some constant c.
- Explain why the shift $c = -\max_j(a_j)$ can be useful.

- Invariance to shift

  suppose shift all activations a_j by constant c : $a'_j = a_j + c$

  $$\hat{y}'_k = \frac{e^{a_k+c}}{\sum_j e^{a_j+c}} = \frac{e^{a_k}e^{c}}{\sum_j e^{a_j}e^{c}} = \frac{e^{a_k}\cancel{e^{c}}}{\cancel{e^{c}}\sum_j e^{a_j}} = \frac{e^{a_k}}{\sum_j e^{a_j}} = \hat{y}_k$$

- Choosing $a'_j = a_j - \max_j(a_j)$

  obs: $a'_j \leq 0$ it helps numerical stability!

  - when some a_j are large e^{a_j} can overflow
  - if maximum is 0 we ensure all components to be ≤ 1 and avoid overflo!