NLP - Home Assignment 1

Ido Calman, Ofri Kleinfeld, Uri Avron

March 23, 2018

1 Basics

(a) We prove that $softmax(x+c) = softmax(x)$:

$$softmax(x+c)_i = \frac{e^{(x_i+c)}}{\sum_j e^{(x_j+c)}} = \frac{e^c e^{x_i}}{\sum_j e^c e^{x_j}} = \frac{e^c e^{x_i}}{e^c \sum_j e^{x_j}} = \frac{e^{x_i}}{\sum_j e^{x_j}} = softmax(x)_i$$

So for each coordinate i it holds that $softmax(x)_i = softmax(x+c)_i$ then softmax is invariant to addition by constant.

(c) Assume that x is a scalar, and define the sigmoid function as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Also define:

$$\bar{\sigma}(x) = \frac{1}{\sigma(x)} = 1 + e^{-x}$$

So by chain rule we get:

$$\frac{\partial}{\partial x}[\bar{\sigma}(x)] = \frac{\partial}{\partial x}[\frac{1}{\sigma(x)}] = \frac{-1}{\sigma(x)^2} \cdot \sigma^{'}(x)$$

$$\implies \sigma^{'}(x) = -\frac{\partial}{\partial x}[\bar{\sigma}(x)] \cdot \sigma(x)^2$$

On the other hand, by simply derivating $\bar{\sigma}(x)$ we get:

$$\frac{\partial}{\partial x}[\bar{\sigma}(x)] = \frac{\partial}{\partial x}[1+e^{-x}] = -e^{-x} = 1 - 1 - e^{-x} =$$
$$= 1 - (1+e^{-x}) = 1 - \frac{1}{\sigma(x)}$$

Then substituting the latter equation into the former we arrive at:

$$\sigma^{'}(x) = -((1 - \frac{1}{\sigma(x)}) \cdot \sigma(x)^2) = -(\frac{\sigma(x)-1}{\sigma(x)} \cdot \sigma(x)^2) =$$
$$= -((\sigma(x)-1)\sigma(x)) = \sigma(x) \cdot (1-\sigma(x))$$