Assignment 1

Arvid Frydenlund

August 29, 2015

1 Q1

softmax($\mathbf{x}$) = softmax($\mathbf{x} + c$) (Note, remember softmax is defined as a function taking in a vector)

Considering element i

$$\frac{e^{x_i}}{\sum_j e^{x_j}} = \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} \tag{1}$$

since,

$$\frac{e^{x_i} e^c}{\sum_j e^{x_j} e^c} = \frac{e^{x_i} \cancel{e^c}}{\sum_j e^{x_j} \cancel{e^c}} \tag{2}$$

QED

This is useful for the trick where $c = -\max_{\mathrm{i}}^{\mathrm{n}} \mathbf{x}_{\mathrm{i}}$, since the largest number in the exponential possible is now 0. That is it prevents overflow.

2 Q2 a)

The sigmoid function is

$$y = \frac{1}{1+e^{-x}} = \sigma(x) = (1+e^{-x})^{-1} = (z)^{-1} \tag{3}$$

Where $z = 1 + e^{-x} = 1 + e^r$, and $r = -x$

Then by simple chain rule

$$\frac{dy}{dx} = \frac{dy}{dz} * \frac{dz}{dr} * \frac{dr}{dx} \tag{4}$$

$$\frac{dy}{dx} = -(1+e^{-x})^{-2} * e^{-x} * (-1) = e^{-x} * \left(\frac{1}{1+e^{-x}}\right)^2 = e^{-x}\sigma(x)\sigma(x) \tag{5}$$

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

$$\frac{1 - \frac{1}{1+e^{-x}}}{\frac{1}{1+e^{-x}}} = (1 + e^{-x}) - \frac{1 + e^{-x}}{1 + e^{-x}} = (1 + e^{-x}) - 1 = e^{-x} \tag{6}$$

Then

$$e^{-x}\sigma(x)\sigma(x) = \frac{1 - \sigma(x)}{\cancel{\sigma(x)}}\cancel{\sigma(x)}\sigma(x) = (1 - \sigma(x))(\sigma(x)) \tag{7}$$

QED

3 Q2 b)

Asume K classes, $\hat{y} = \text{softmax}(\theta) = [\hat{y}_1, ..., \hat{y}_K]$ =[softmax(θ_1), ...,softmax(θ_1)]