ADVANCED METHODS IN NLP
EXERCISE #1 SOLUTION

Uri Avron [uriavron@gmail.com] [308046994]
Ofri Kleinfeld [] []
Ido Calman [] []

March 24, 2018

1 Basics

(a) Indeed, for every dimension i:

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

which implies that $softmax(\mathbf{x}+c) = softmax(\mathbf{x})$. $\square$

(c) Let us compute the gradient:

$$\begin{aligned}\frac{\partial \sigma}{\partial x} &= \frac{\partial (1+e^{-x})^{-1}}{\partial x} \\ &= -(1+e^{-x})^{-2}(-e^{-x}) = e^{-x}(1+e^{-x})^{-2} \\ &= (\frac{1}{\sigma(x)} - 1)\sigma^2(x) \\ &= \sigma(x)(1-\sigma(x))\end{aligned}$$

2 Word2vec

(a) Let us write $CE(y, \hat{y}) = -y \cdot \log(\hat{y})$. Therefore the gradient with respect to