Excercise 2 – Softmax and Gradients

Yoav Goldberg

Q1

Consider the softmax function:

$$\text{softmax}(\mathbf{x})_{[i]} = \frac{\exp(\mathbf{x}_{[i]})}{\sum_j \exp(\mathbf{x}_{[j]})}$$

Show that $\text{softmax}(\mathbf{x}) = \text{softmax}(\mathbf{x}+c)$ for every constat c. ($\mathbf{x}+c$ means adding c to every element in $\mathbf{x}$).

Q2

Consider a multi-class log-linear classifier (that is, multi-class linear classifier followed by a softmax) with a cross-entropy loss.

$$\hat{y} = \arg\max_i \hat{\mathbf{y}}_{[i]}$$
$$\hat{\mathbf{y}} = f(\mathbf{x}) = \text{softmax}(\mathbf{x}\mathbf{W} + \mathbf{b})$$

$$\ell(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_i \mathbf{y}_{[i]} \log(\hat{\mathbf{y}}_{[i]})$$

Compute the gradients of the parameters $\mathbf{W}, \mathbf{b}$ with respect to the loss:

$$\frac{\partial \ell(\mathbf{y}, f(\mathbf{x}))}{\partial \mathbf{W}_{[i,j]}}$$

$$\frac{\partial \ell(\mathbf{y}, f(\mathbf{x}))}{\partial \mathbf{b}_{[i]}}$$

(Example of grandients computation for a multi-class linear classifier with the hinge loss is given in chapter 2 of the neural-networks for NLP book.)