CS 224N Assignment #1.

1. Softmax

(a) $\text{softmax}(x)_i = \dfrac{e^{x_i}}{\sum_j e^{x_j}}$

$$\text{softmax}(x_i+c) = \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} = \frac{e^c \cdot e^{x_i}}{e^c \cdot \sum_j e^{x_j}} = \frac{e^{x_i}}{\sum_j e^{x_j}} = \text{softmax}(x_i).$$

(b). CODE.

2. Neural Network Basics.

(a). $\sigma(x) = \dfrac{1}{1+e^{-x}} \Rightarrow \dfrac{d}{dx}\sigma(x) = \dfrac{-(-e^{-x})}{(1+e^{-x})^2} = (1-\sigma x)\,\sigma x.$

(b)

$\hat{y} = \text{softmax}(\theta)$

$$\frac{\partial CE(y,\hat{y})}{\partial \theta} = \frac{\partial CE(y,\hat{y})}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta}.$$

The first step is to calculate the derivative of softmax() function.

$$\hat{y_i} = \frac{e^{\theta_i}}{\sum_k e^{\theta_k}}$$

$$\frac{\partial \hat{y_i}}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \frac{1}{1+Ae^{-\theta_i}} \quad (\text{where } A = \sum_k e^{\theta_k} - e^{\theta_i}).$$

$$= \frac{1}{(1+Ae^{-\theta_i})}\left(1-\frac{1}{1+Ae^{-\theta_i}}\right) = \hat{y_i}(1-\hat{y_i}).$$

$$\frac{\partial \hat{y_i}}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{e^{\theta_i}}{A+e^{\theta_j}} \quad (\text{where } A = \sum_k e^{\theta_k} - e^{\theta_j}).$$

$$= \frac{e^{\theta_i}}{A+e^{\theta_j}} \cdot \frac{-e^{\theta_j}}{A+e^{\theta_j}} = -\hat{y_i} \cdot \hat{y_j}$$

$$\frac{\partial CE(y,\hat{y})}{\partial \theta_i} = -\sum_k y_k \cdot \frac{\partial \log \hat{y_k}}{\partial \theta_i} = -\sum_k y_k \cdot \frac{1}{\hat{y_k}} \frac{\partial \hat{y_k}}{\partial \theta_i}$$

$$= -(y_i \cdot \hat{y_i} \cdot \frac{1}{\hat{y_i}}(1-\hat{y_i}) + \sum_{k\neq i} y_k \frac{1}{\hat{y_k}}(-\hat{y_i}\hat{y_k}))$$

$$= -y_i + \sum_k \hat{y_i} y_k = \hat{y_i} - y_i$$

$$\Rightarrow \frac{\partial CE(y,\hat{y})}{\partial \theta} = \hat{y} - y$$