

# Assignment 1: CS 224d (Jerrick Hoang)

Due: Wednesday

1. Softmax is translational invariant  $\rho(\mathbf{x}+c)_j = \frac{\exp(x_j+c)}{\sum_{k=1}^K \exp(x_k+c)} = \frac{\exp(x_j)\exp(c)}{\exp(c)\sum_{k=1}^K \exp(x_k)} = \frac{\exp(x_j)}{\sum_{k=1}^K \exp(x_k)} = \rho(\mathbf{x})$
2. (a) Let  $\rho(\mathbf{x}) = \frac{1}{1+e^{-\mathbf{x}}}$ . Then  $\frac{\partial \rho}{\partial \mathbf{x}} = \frac{-1}{(1+e^{-\mathbf{x}})^2} \frac{d(1+e^{-\mathbf{x}})}{d\mathbf{x}} = \frac{e^{-\mathbf{x}}}{(1+e^{-\mathbf{x}})^2} = (1 - \rho(\mathbf{x}))\rho(\mathbf{x})$   
 (b) Let  $h(z)_j = \frac{e^{z_j}}{\sum e^{z_i}}$ . Let's derive  $\frac{\partial h(z)_j}{\partial z_i}$ . When  $i = j$ , must find  $\frac{\partial h(z)_j}{\partial z_j} = \frac{\partial}{\partial z_j} \frac{e^{z_j}}{\sum e^{z_i}} = \frac{e^{z_j}(\sum e^{z_i}) - e^{2z_j}}{(\sum e^{z_i})^2} = h(z)_j(1 - h(z)_j)$ . When  $i \neq j$ ,  $\frac{\partial h(z)_j}{\partial z_i} = \frac{\partial}{\partial z_i} \frac{e^{z_j}}{\sum e^{z_i}} = \frac{-e^{z_j}e^{z_i}}{(\sum e^{z_i})^2} = -h(z)_i h(z)_j$ .  
 Now,  $\frac{\partial CE(y, \hat{y})}{\partial \theta_j} = -\sum_i y_i \frac{\partial}{\partial \theta_j} \log h(\theta)_i = -\sum_i \frac{y_i}{h(\theta)_i} \frac{\partial}{\partial \theta_j} h(\theta)_i = -\sum_{i \neq j} \frac{y_i}{h(\theta)_i} h(\theta)_i h(\theta)_j - (h(\theta)_j(1 - h(\theta)_j)) = -h(\theta)_j(\sum_{i \neq j} y_i + 1 - h(\theta)_j)$
- (c)  $\frac{\partial \hat{y}}{\partial x} = \frac{\partial \hat{y}}{\partial h} \frac{dh}{dx} =$