Assignment 1 Notes

mn

Contents

1 Assigment 1

Org latex export is not that great yet, and one still has to work on the alignment.

1.1 1.a) Softmax function $\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$

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

1.2 2.a) Derivative of the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$

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

1.3 2.b) Derivative of the Cross Entropy Loss $CE(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\sum_i y_i \log(\hat{y}_i)$

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

$$\frac{\partial}{\partial\theta_j} CE(\boldsymbol{y}, \hat{\boldsymbol{y}})_j = -\frac{\partial}{\partial\theta_j} \sum_i y_i \log \hat{y}_i = -\frac{\partial}{\partial\theta_j} \sum_i y_i \log(e^{\theta_i} / \sum_l e^{\theta_l}) = -\frac{\partial}{\partial\theta_j} \sum_i y_i (\theta_i - \log \sum_l e^{\theta_l})$$

$\boldsymbol{y}$ is a $one-hot$ label vector with non-zero value at index k.
If $j \neq k$:

$$\frac{\partial}{\partial\theta_j} CE(\boldsymbol{y}, \hat{\boldsymbol{y}})_j = \frac{\partial}{\partial\theta_j} \log \sum_l e^{\theta_l} = \hat{y}_j$$

If $j = k$:

$$\frac{\partial}{\partial\theta_j} CE(\boldsymbol{y}, \hat{\boldsymbol{y}})_j = \hat{y}_j - 1$$

given that

$$\frac{\partial}{\partial\theta_j} \log \sum_l e^{\theta_l} = \frac{e^{\theta}_j}{\sum_l e^{\theta_l}} = \hat{y}_j$$