Jefferson Li
lizhe20
jeffersonli.li@mail.utoronto.ca

Assignment 3
Intro to Machine Learning

2020-11-06

1 Question 1

- Case 1 : $k \neq j$

$$\frac{\partial \sigma(z_{1:K}, k)}{\partial z_j} = \frac{-e^{z_k}}{(\Sigma_{l=1}^{K} e^{z_l})^2} \cdot e^{z_j} \qquad \text{Apply chain rule on } z_j \tag{1}$$

- Case 1 : $k = j$

$$\frac{\partial \sigma(z_{1:K}, j)}{\partial z_j} = \frac{e^{z_j}(\Sigma_{l=1}^{K} e^{z_l}) - e^{2z_j}}{(\Sigma_{l=1}^{K} e^{z_l})^2} \qquad \text{Apply division rule on } z_j \tag{2}$$

2 Question 2

$$\frac{\partial \Sigma_{k=1}^{K} y_k log(\sigma(z_{1:K}, k))}{\partial \mathbf{w}_j} = \tag{3}$$

$$= \Sigma_{k=1}^{K} \frac{y_k}{\sigma(z_{1:K}, k)} \frac{\partial \sigma(z_{1:K}, k)}{\partial z_j} \frac{\partial z_j}{\mathbf{w}_j} \qquad \text{Apply derivative rule for logs} \tag{4}$$

$$= \Sigma_{k=1}^{K} \frac{y_k}{\sigma(z_{1:K}, k)} \frac{\partial \sigma(z_{1:K}, k)}{\partial z_j} \mathbf{x} \qquad \text{Since } z_j = \mathbf{w}_j^T \mathbf{x} \tag{5}$$

$$= \mathbf{x}(\Sigma_{k \in [1,K] \setminus j} \frac{y_k}{\sigma(z_{1:K}, k)} \frac{\partial \sigma(z_{1:K}, k)}{\partial z_j} + \frac{y_j}{\sigma(z_{1:K}, j)} \frac{\partial \sigma(z_{1:K}, j)}{\partial z_j}) \qquad \text{split up special case of } k = j \tag{6}$$

$$= \mathbf{x}(\Sigma_{k \in [1,K] \setminus j} \frac{y_k \Sigma_{l=1}^{K} e^{z_l}}{e^{z_k}} \frac{-e^{z_k}}{(\Sigma_{l=1}^{K} e^{z_l})^2} \cdot e^{z_j} + \frac{y_j \Sigma_{l=1}^{K} e^{z_l}}{e^{z_j}} \frac{e^{z_j}(\Sigma_{l=1}^{K} e^{z_l}) - e^{2z_j}}{(\Sigma_{l=1}^{K} e^{z_l})^2}) \tag{7}$$

$$= \mathbf{x}[(\Sigma_{k \in [1,K] \setminus j} y_k \sigma(z_{1:K,j})) + y_j(1 - \sigma(z_{1:K,j}))] \qquad \text{Cancellations} \tag{8}$$

$$\tag{9}$$

3 Question 3

$$\frac{\partial E(\mathbf{w}_{1:K})}{\partial \mathbf{w}_j} = \tag{10}$$

$$= (-\Sigma_{i=1}^{N} \Sigma_{k=1}^{K} y_{i,k} log(\sigma(z_{i,1:K}, k))' \tag{11}$$

$$= -\Sigma_{i=1}^{N} [(\Sigma_{k \in [1,K] \setminus j} y_k \sigma(z_{1:K,j})) + y_j(1 - \sigma(z_{1:K,j}))] \mathbf{x}_i \qquad \text{From Question 2} \tag{12}$$

$$\tag{13}$$

4 Question 4

$$\hat{E}(\mathbf{w}_{1:K}) = -log(p(\mathbf{w}_{1:K})\Pi_{i=1}^{N} p(y = y_i|\mathbf{x}_i, \mathbf{w}_{1:K})) \tag{14}$$

$$= \underbrace{-log(p(\mathbf{w}_{1:K}))}_{\text{regularization term}} \underbrace{-log(\Pi_{i=1}^{N} p(y = y_i|\mathbf{x}_i, \mathbf{w}_{1:K}))}_{E(\mathbf{w}_{1:K})} \tag{15}$$

$$\tag{16}$$

$$\text{Let } D(\mathbf{w}_{1:K}) \text{ be } \frac{\partial E(\mathbf{w}_{1:K})}{\partial \mathbf{w}_j} \text{ as defined in question 3} \tag{17}$$

$$\frac{\partial \hat{E}(\mathbf{w}_{1:K})}{\partial \mathbf{w}_k} = -(log(\frac{1}{(2\pi\beta\alpha^D)^{(D+1)/2}})\Sigma_{k=1}^{K} log(exp(\frac{-\mathbf{w}_k^T C^{-1}\mathbf{w}_k}{2})))' + D(\mathbf{w}_{1:K}) \tag{18}$$

$$= -(log(\frac{1}{(2\pi\beta\alpha^D)^{(D+1)/2}})\Sigma_{k=1}^{K} \frac{-\mathbf{w}_k^T C^{-1}\mathbf{w}_k}{2})' + D(\mathbf{w}_{1:K}) \tag{19}$$

$$= -\frac{log(\frac{1}{(2\pi\beta\alpha^D)^{(D+1)/2}})}{2}\Sigma_{k=1}^{K}(C^{-1} + C^{-1T})\mathbf{w}_k + D(\mathbf{w}_{1:K}) \qquad \text{from Common matrix identities pdf} \tag{20}$$