深度学习 hw1

尹嘉骏 201300063

第一问

$W^{[2]}$ 与 $b^{[2]}$ 的维度分别为 $K \times D_a, K \times 1$
最后一层输出结果的维度为 $K \times m$

第二问

$\hat{y}_k = \frac{exp(\mathbf{z}_k^{[2]})}{Z}$
$\frac{\partial \hat{y}_k}{\partial \mathbf{z}_k^{[2]}} = \frac{Z \cdot exp(\mathbf{z}_k^{[2]}) - [exp(\mathbf{z}_k^{[2]})]^2}{Z^2} = \hat{y}_k(1-\hat{y}_k)$

第三问

当 $i \neq k$ 时，有
$\frac{\partial \hat{y}_k}{\partial \mathbf{z}_i^{[2]}} = \frac{0 \cdot Z - exp(\mathbf{z}_k^{[2]}) \cdot exp(\mathbf{z}_i^{[2]})}{Z^2} = -\hat{y}_k \hat{y}_i$

第四问

$L = -\sum_{i=1}^{K} y_i log(\hat{y}_i) = -log(\hat{y}_k)$
而 $\hat{y}_k = \frac{exp(\mathbf{z}_k^{[2]})}{Z}$
故 $L = -logexp(\mathbf{z}_k^{[2]}) + logZ = -\mathbf{z}_k^{[2]} + logZ = -\mathbf{z}_k^{[2]} + log(\sum_{j=1}^{K} exp(\mathbf{z}_j^{[2]}))$
当 i=k 时，有 $\frac{\partial L}{\partial \mathbf{z}_i^{[2]}} = -1 + \frac{exp(\mathbf{z}_k^{[2]})}{\sum_{j=1}^{K} exp(\mathbf{z}_j^{[2]})} = -1 + \hat{y}_k$
当 $i \neq k$ 时，有 $\frac{\partial L}{\partial \mathbf{z}_i^{[2]}} = 0 + \frac{exp(\mathbf{z}_i^{[2]})}{\sum_{j=1}^{K} exp(\mathbf{z}_j^{[2]})} = -1 + \hat{y}_k$