Problem 1: Show the multi-class Softmax reduces to two-class Softmax when C = 2 (30 points)

With $C = 2$ the multiclass softmax cost

$$\sum_{c=1}^{C} \sum_{p \in \Omega_c} \log \left(1 + \sum_{\substack{j=1 \\ j \neq c}}^{C} e^{\left(b_j - b_c\right) + \mathbf{x}_p^T \left(\mathbf{w}_j - \mathbf{w}_c\right)} \right),$$

reduces to

$$\sum_{p \in \Omega_1} \log \left(1 + e^{\left(b_2 - b_1\right) + \mathbf{x}_p^T \left(\mathbf{w}_2 - \mathbf{w}_1\right)}\right) + \sum_{p \in \Omega_2} \log \left(1 + e^{\left(b_1 - b_2\right) + \mathbf{x}_p^T \left(\mathbf{w}_1 - \mathbf{w}_2\right)}\right).$$

Now note that because we have that $y_p = \begin{cases} -1 & p \in \Omega_1 \\ +1 & p \in \Omega_2 \end{cases}$, the cost in (7.24) can be written equivalently as

$$\sum_{p \in \Omega_1} \log \left(1 + e^{-y_p\left(\left(b_2 - b_1\right) + \mathbf{x}_p^T \left(\mathbf{w}_2 - \mathbf{w}_1\right)\right)}\right) + \sum_{p \in \Omega_2} \log \left(1 + e^{-y_p\left(\left(b_2 - b_1\right) + \mathbf{x}_p^T \left(\mathbf{w}_2 - \mathbf{w}_1\right)\right)}\right),$$

which can then be written in a more compact form as

$$\sum_{p=1}^{P} \log \left(1 + e^{-y_p\left(\left(b_2 - b_1\right) + \mathbf{x}_p^T \left(\mathbf{w}_2 - \mathbf{w}_1\right)\right)}\right).$$

Finally letting $b = b_2 - b_1$ and $\mathbf{w} = \mathbf{w}_2 - \mathbf{w}_1$, we arrive at the familiar two-class softmax cost function

$$\sum_{p=1}^{P} \log \left(1 + e^{-y_p\left(b + \mathbf{x}_p^T \mathbf{w}\right)}\right).$$

Problem 2: Balanced accuracy in the multi-class setting (20 points)

Suppose we have formed the $(C-1) \times (C-1)$ confusion matrix $\mathbf{Q}$ for a general C-class classification, where the (i, j)th entry $Q_{i,j}$ is the number of datapoints in class i that have been assigned label j by the classifier.

Then the accuracy for class i can be found as

$$\mathcal{A}_i = \frac{Q_{i,i}}{\sum_{j=0}^{C-1} Q_{i,j}}$$

Averaging the accuracies for all C classes we have

$$\text{balanced accuracy} = \frac{\sum_{i=0}^{C-1} \mathcal{A}_i}{C} = \frac{1}{C} \sum_{i=0}^{C-1} \frac{Q_{i,i}}{\sum_{j=0}^{C-1} Q_{i,j}}$$