HW 5: ECE 601 Machine Learning for Engineers

Important Notes:

(a) When a HW question asks for writing a code, you would need to include the entire code as well as the output of the program as well as any other analysis requested in the question.

(b) Combine your solutions in **one** zip file called `homework5_UMassUSERNAME.zip`.

1. Consider a neural network output layer with three neurons, producing raw scores (logits) z_1, z_2, z_3. The softmax activation function is defined as:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}} \tag{1}$$

   (a) Given logits $z_1 = 3$, $z_2 = 1$, and $z_3 = -2$, compute the softmax probabilities $\sigma(z_1), \sigma(z_2), \sigma(z_3)$. Express your answer as fractions and decimals (rounded to 4 decimal places).

   (b) Show that the softmax function is invariant to adding the same constant c to all logits, i.e., prove that $\sigma(z_i + c) = \sigma(z_i)$ for any constant c.

   (c) Compute the partial derivative of $\sigma(z_i)$ with respect to z_j (both when $i = j$ and when $i \neq j$).

   (d) Consider two different sets of logits: $(z_1, z_2, z_3) = (5, 2, -1)$ and $(z_1, z_2, z_3) = (50, 20, -10)$. Compute the softmax values for both cases. What do you observe about numerical stability? How can this issue be addressed in practical implementations?

   (e) Prove that the softmax function amplifies the differences between logits, making larger logits even more dominant in the probability distribution. Hint: Compare the ratio $\frac{\sigma(z_1)}{\sigma(z_2)}$ for different values of $z_1 - z_2$.

2. You have gained experience applying Convolutional Neural Networks (CNN) to the MNIST dataset, which consists of grayscale images. The progression of your learning journey introduces you to the CIFAR-10 dataset. CIFAR-10 comprises 60,000 color images in 10 classes, each class having 6,000 images, making the dataset not only more colorful but also more complex than MNIST. The objective of this exercise is to develop a CNN model that classifies images from the CIFAR-10 dataset efficiently. Follow the provided template, entitled `CNN_CIFAR10.ipynb`. Instructions for each part are provided. Replace the placeholder `'write your code here'` in the template with your code.

NOTE: You are designing four CNN models. You need to train and test each of the four models separately.