Problem Solving 11

Stochastic Gradient Descent

1. Given is the following dataset, where each row is a sample represented by a 4-pixel image and the measured output is the last column.

D_{train} = [[255, 128, 128, 0, 0], [55, 128, 128, 128, 1], [192, 128, 128, 0, 0],

[100, 128, 128, 100, 1], [30, 64, 128, 30, 2], [20, 64, 128, 0, 2]]

Consider a neural network (NN) composed of 3 linear activations (a 3-class fully-connected layer), followed by softmax, parameterized by weights W_{4x3} and biases B_{1x3}, initialized with the values 0. Using the learning rate $\alpha = 0.1$, calculate the updates ΔW and Δb for 2 epochs using the cross-entropy loss L. Recall that:

$$\Delta w_{i,j} = -\alpha \frac{\partial L}{\partial w_{i,j}} = \alpha\, x_i \begin{cases} (1 - p_j), & \text{if answer } a = j; \\ -p_j, & \text{otherwise.} \end{cases}$$

$$\Delta b_j = -\alpha \frac{\partial L}{\partial b_j} = \alpha \begin{cases} (1 - p_j), & \text{if } a = j; \\ -p_j, & \text{otherwise.} \end{cases}$$

where a is the ground-truth class, p_j is the softmax of $l_j = b_j + x \cdot w_j$.

2. Calculate the training accuracy.

3. Consider now the following validation set data:

D_{val} = [[155, 64, 64, 32, 0], [0, 128, 128, 192, 1], [20, 40, 80, 20, 2]]

What is the accuracy of the previous model for this data? What are the outputs of the neural network for each validation sample?

4. Now coming from the same initialization (all parameters zeroed), but now using the batch size of 3, calculate the gradients, the update of $\Delta w_{i,j}$ and Δb_j and the new parameter values, for 1 epoch with learning rate $\alpha = 0.1$.

5. Calculate the training accuracy of the new model.

6. Considering again the same validation set, what is the accuracy for the model trained with batch size of 3? What are the outputs of the NN for each validation sample?