ECE 417 Multimedia Signal Processing
Homework 3

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Tuesday, 9/19/2023; Due: Tuesday, 10/3/2023

Problem 3.1

Suppose $y = \|\mathbf{A}\mathbf{X}\|_F$, the Frobenius norm of $\mathbf{A}$ times $\mathbf{X}$ (note: NOT the squared Frobenius norm!) Find $\partial y/\partial \mathbf{X}$.

Problem 3.2

You are trying to represent a mapping from a real-valued scalar input, $x \in \Re$, to a real-valued scalar output, $y \in \Re$. Suppose that you have a series of training examples, $\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$. You want to represent the mapping from y to x using the following two-layer network:

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{w}_1 x + \mathbf{b}_1 \\ \mathbf{a}_1 &= \text{ReLU}(\mathbf{z}_1) \\ g(x) &= \mathbf{w}_2^T \mathbf{a}_1 + b_2 \end{aligned}$$

Your goal is to find $\mathbf{w}_1$, $\mathbf{b}_1$, $\mathbf{w}_2$ and b_2 in order to minimize $\mathcal{L}$ which is defined as

$$\mathcal{L} = \frac{1}{2n} \sum_{i=1}^{n} (y_i - g(x_i))^2$$

How many hidden nodes (what dimension of $\mathbf{w}_1$) do you need to guarantee that $\mathcal{L} = 0$? In terms of the training tokens x_i and y_i, choose some values of $\mathbf{w}_1$, $\mathbf{b}_1$, $\mathbf{w}_2$ and b_2 that will result in $\mathcal{L} = 0$. Hint: you may find it convenient to assume that the data are sorted in order of increasing x, i.e., $x_i < x_{i+1}$ for all i.

Problem 3.3

Suppose you have a neural net that is configured as a binary classifier. Given a vector input $\mathbf{x}$, it computes a scalar output $g(\mathbf{x}) \in (0, 1)$ according to

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{W}_1 \mathbf{x} + \mathbf{b}_1 \\ \mathbf{a}_1 &= \text{ReLU}(\mathbf{z}_1) \\ z_2 &= \mathbf{w}_2^T \mathbf{a}_1 + b_2 \\ g(\mathbf{x}) &= \sigma(z_2) \end{aligned}$$

where $\sigma(\cdot)$ is the logistic sigmoid, and $\text{ReLU}(\mathbf{z}_1)$ applies rectified linear nonlinearities to each element of $\mathbf{z}_1$. Suppose that the loss function is binary cross entropy, i.e.,

$$\mathcal{L} = -y \ln g(\mathbf{x}) - (1 - y) \ln(1 - g(\mathbf{x}))$$

What is $\partial\mathcal{L}/\partial\mathbf{W}_1$? Express your answer in terms of any of the vectors or matrices specified in the problem, but if you have to define any additional matrices (e.g., if you have to define $\text{ReLU}'(\mathbf{z}_1)$), be sure to specify the values of every element of every such matrix in terms of the elements of the matrices and vectors provided.

Problem 3.4

The filters in convolutional neural nets tend to learn common features of the images on which they are trained. To see how this occurs, consider the following CNN, composed of a one-channel convolutional layer followed by one global pooling layer, with a grayscale input image:

$$z_1[m,n] = w_1[m,n] * x[m,n] \tag{3.4-1}$$

$$g(x) = \max_{m=0}^{M-1} \max_{n=0}^{N-1} z_1[m,n] \tag{3.4-2}$$

Suppose that, initially, $w_1[m,n] = \delta[m,n]$, i.e., $w[0,0] = 1$ and $w[m,n] = 0$ for all other (m,n). The maximization in Eq. (3.4-2) therefore picks out the maximum sample in the input signal, which is at position (m^*, n^*):

$$(m^*, n^*) = \operatorname*{argmax}_{m=0}^{M-1} \operatorname*{argmax}_{n=0}^{N-1} z_1[m,n]$$

Suppose that the loss function is MSE, $\mathcal{L} = \frac{1}{2}(y - g(x))^2$, and suppose that the CNN is trained for one step of gradient descent with a learning rate of $\eta = 1$:

$$w_1[m,n] \leftarrow w_1[m,n] - \frac{\partial\mathcal{L}}{\partial w[m,n]}$$

After the one step of training shown above, what is $w_1[m,n]$? Write your answer in terms of y, x, $g(x)$, $\delta[m,n]$, m^*, n^*, and any of the other variables that have been introduced in the problem statement.