Assignemnt Set #1

Amir Sadovnik
COSC 525 (Spring 2021)

☐ **Problem 1 (15).**
Consider a simple perceptron with a sigmoid activation function and two inputs. Assume the following values for the weights: $w_1 = 0.2, w_2 = 0.3, b = 0.1$. Given a labeled data point with $x_1 = 2, x_2 = 1, y = 1$ answer the following:

(a) What is the output when feeding this data point to the network?

(b) If using binary cross-entropy, what is the loss for this data sample?

(c) In general, what is the role of the bias in the perceptron?

☐ **Problem 2 (18).**
Given the following 4×4 image on the left $\mathbf{M}$ and a convolutional layer with a single 2×2 kernel $\mathbf{K}$ with stride one and no padding.

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

(a) Calculate the output of the convolutional layer

(b) Answer (a) if instead the input has two channels (both equal to $\mathbf{M}$), and the kernel had two channels (both equal to $\mathbf{K}$)

(c) Answer (a) if the input is unchanged, but we now have two different kernels both equal to $\mathbf{K}$.

☐ **Problem 3 (25).**
Consider the following code creating a simple convolutional neural network:

```
model=Sequential()
model.add(layers.Conv2D(filters=6,kernel_size=5, strides=2,padding='valid'
                        ,input_shape=(21,21,3)))
model.add(layers.MaxPooling2D(pool_size=3, strides=2, padding='valid'))
model.add(layers.Conv2D(filters=10,kernel_size=3, strides=1,padding='same'))
model.add(layers.Flatten())
model.add(layers.Dense(10))
```

For each of the 5 layers calculate the following:

- The dimensions of the output
- The number of trainable parameters (weights and biases)

☐ Problem 4 (20).
Assume you are training a deep neural network using stochastic gradient descent

(a) Sketch a typical curve which shows the training loss as a function of training steps, for the following cases: too small of a learning rate, to large of a learning rate, optimal learning rate.

(b) Explain two reasons the training cost might go up when performing stochastic gradient descent, and what can be done to partly alleviate these problems.

(c) Explain the difference between batch, mini-batch and stochastic gradient descent. Why do we then to usually use mini-batch?

(d) One of the problems with backpropagation is the vanishing gradient when networks get very deep. Explain how GoogLeNet and Resnet attempt to alleviate this problem.

☐ Problem 5 (22).
Given the following neural network:

```
model3=Sequential()
model3.add(layers.Conv1D(filters=1,kernel_size=2, strides=1,padding='same',
                  input_shape=(785,1),,activation='sigmoid'))
model3.add(layers.Flatten())
model3.add(layers.Dense(1,activation='sigmoid'))
```

Let the outputs of the layers in the network be denoted using $X_i, c_i(X), o_i(X)$ for the INPUT, CONV1 and OUTPUT layers respectively where i is the index of the neuron in the layer. You may use those without explicitly calculating them. In addition the weights can be denoted w_i^c, w_i^f for the CONV1 and OUTPUT layers respectively. The learning rate is set to α. Given a labeled datapoint (x, y), show how you would calculate the following:

(a) The update to the weight i of the output node layer.

(b) The update to the bias of the output node layer.

(c) The update to the weights w_0^c, w_1^c of the convolutional layer.

(d) The update to the bias of the convolutional layer.