Simon Fraser University
Department of Computer Science

Deep Learning - Assignment 1

Name: Niloufar Saeidi
ID: 301590708

A.1

Least-squares error:

$E(w) = \frac{1}{2}((x.w) - y)^2$

$\nabla E(w) = [\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, ...] = \frac{\partial E}{\partial w_i} = \frac{1}{2}(2)(x.w - y)(x)$

Cross Entropy:

$E(w) = yln(p) + (1-y)ln(1-p)$

chain rule: $\frac{\partial E}{\partial x} = \frac{\partial E}{\partial p} \times \frac{\partial p}{\partial x}$

$\frac{\partial E}{\partial p} = \frac{y}{p} - \frac{1-y}{1-p}$

$\frac{\partial p}{\partial x} = \frac{\partial \sigma(x)}{\partial x} = \sigma(x) \times (1 - \sigma(x))$

$\frac{\partial E}{\partial x} = (\frac{y}{p} - \frac{1-y}{1-p}) \times \sigma(x) \times (1 - \sigma(x))$

A.2

$$\begin{pmatrix} 0.5 & 1.5 \\ 3.5 & 2 \end{pmatrix} \begin{pmatrix} 0.5 & 1 \\ 3 & 2 \end{pmatrix} + \begin{pmatrix} 1.25 & 3 \end{pmatrix} = \begin{pmatrix} 0.25 + 4.5 & 0.5 + 3 \\ 1.75 + 6 & 3.5 + 4 \end{pmatrix} + \begin{pmatrix} 1.25 & 3 \\ 1.25 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 6.5 \\ 9 & 10.5 \end{pmatrix}$$

In broadcasting in w*x+b, b is usually give in one row and we copy the same row for each data point.

A.3

1

$a_c = Sigmoid(a_a.W_{ac} + a_b.W_{bc} + W_{0c})$

$a_d = Sigmoid(a_c.W_{cd} + W_{0d})$

$\Rightarrow a_x = Sigmoid(a_a.W_{ax} + a_b.W_{bx} + W_{0x})$

$\Delta[x] = -sigmoid'(a_x)\frac{\partial E(a_x, y_x)}{\partial a_x}$

$W_{xy} \leftarrow W_{xy} + \alpha \times a_x \times \Delta[y]$

2

I perform the first iteration on data point 1 and the second iteration on data point 2:

$a_c = sigmoid(w_{0c} + w_{ac}.a + w_{bc}.b) = sigmoid(0.2 + 0.2 \times 1 + 0.1 \times 0) = 0.598$

$a_d = sigmoid(w_{0d} + w_{cd}.c) = sigmoid(0.2 + 0.1 \times 0.598) = 0.612$

$\Delta[d] = (a_d - 1) * sigmoid'(w_{cd} \times a_c + w_{0d}) = (0.612 - 1) \times sigmoid'(0.460) = -0.388 \times 0.247 = -0.0958$

$\Delta[c] = sigmoid'(w_{ac} \times a_x + w_{0c}) \times (\Delta[d] \times w_{cd}) = sigmoid'(0.2 \times 1 + 0.2) \times (-0.0960 * 0.1) = 0.247 \times (-0.00960) = -0.00237$

$w_{cd} \leftarrow w_{cd} - lr \times \Delta[d] \times a_c = 0.1 - 0.1 \times (-0.0958) \times 0.599 = 0.1 + 0.00572 = 0.1057$

$w_{0d} \leftarrow w_{0d} - lr \times \Delta[d] = 0.2 - 0.1 \times (-0.0958) = 0.2 + 0.00958 = 0.2095$

$w_{ac} \leftarrow w_{ac} - lr \times \Delta[c] \times a_x = 0.2 - 0.1 \times (-0.00237) \times 1 = 0.2 + 0.00024 = 0.2002$

$w_{0c} \leftarrow w_{0c} - lr \times \Delta[c] = 0.2 - 0.1 \times (-0.00237) = 0.2 + 0.000237 = 0.2002$

Second iteration on data point 2:

$a_c = sigmoid(w_{ac} \times a_x + w_{0c}) = sigmoid(0.2 \times 0 + 0.2) = sigmoid(0.2) = 0.545$

$a_d = sigmoid(w_{cd} \times a_c + w_{0d}) = sigmoid(0.1 \times 0.55 + 0.21) = sigmoid(0.055 + 0.21) = sigmoid(0.264) = 0.565 - \Delta[d] = (a_d - 0) \times sigmoid'(w_{cd} \times a_c + w_{0d}) = (0.565 - 0) \times sigmoid'(0.2645) = 0.565 \times 0.2467 = 0.1396$

$\Delta[c] = sigmoid'(w_{ac} \times a_x + w_{0c}) \times (\Delta[d] \times w_{cd}) = sigmoid'(0.2 \times 0 + 0.2) \times (0.1396 \times 0.1) = 0.2493 \times 0.01396 = 0.003475$

$w_{cd} \leftarrow w_{cd} - lr \times \Delta[d] \times a_c = 0.1057 - 0.1 \times 0.1396 \times 0.550 = 0.1057 - 0.0076 = 0.098$

$w_{0d} \leftarrow w_{0d} - lr \times \Delta[d] = 0.21 - 0.1 \times 0.1396 = 0.21 - 0.01396 = 0.1956$

$w_{ac} \leftarrow w_{ac} - lr \times \Delta[c] \times a_x = 0.2002 - 0.1 \times 0.00347 \times 0 = 0.2002 - 0 = 0.2002$

$w_{0c} \leftarrow w_{0c} - lr \times \Delta[c] = 0.200237073162 - 0.1 \times 0.00347 = 0.2002 - 0.000347 = 0.1999$

Data	a_c	$\Delta[c]$	a_d	$\Delta[d]$	W_{0c}	W_{ac}	W_{bc}	W_{cd}	W_{0d}
x_1	0.598	-0.00237	0.613	-0.0958	0.2002	0.2002	0.1	0.1057	0.21
x_2	0.5498	0.00347	0.565	0.1396	0.1999	0.2002	0.1	0.098	0.1956