CMSC 25300: Mathematical Foundations of ML

Problem Set 7

Hung Le Tran

25 Nov 2023

Problem 7.1 (Problem 2)

Solution

(a)

$$p = 2$$

(b) We were given

$$\text{Input } A^{(0)} = \begin{bmatrix} .05 \\ .10 \end{bmatrix}$$

$$W^{(1)} = \begin{bmatrix} 0.15 & 0.20 \\ 0.25 & 0.30 \end{bmatrix}, W^{(2)} = \begin{bmatrix} 0.40 & 0.45 \\ 0.50 & 0.55 \end{bmatrix}$$

$$B^{(1)} = \begin{bmatrix} .35 \\ .35 \end{bmatrix}, B^{(2)} = \begin{bmatrix} .60 \\ .60 \end{bmatrix}$$

Calculating forward pass:

$$\begin{aligned} Z^{(1)} &= W^{(1)}A^{(0)} + B^{(1)} \\ &= \begin{bmatrix} .15 & .20 \\ .25 & .30 \end{bmatrix} \begin{bmatrix} .05 \\ .10 \end{bmatrix} + \begin{bmatrix} .35 \\ .35 \end{bmatrix} \\ &= \begin{bmatrix} 0.3775 \\ 0.3925 \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
A^{(1)} &= sigmoid(Z^{(1)}) = \begin{bmatrix} 0.59326999 \\ 0.59688438 \end{bmatrix} \\
Z^{(2)} &= W^{(2)}A^{(1)} + B^{(2)} \\
&= \begin{bmatrix} .40 & .45 \\ .50 & .55 \end{bmatrix} \begin{bmatrix} 0.59326999 \\ 0.59688438 \end{bmatrix} + \begin{bmatrix} .60 \\ .60 \end{bmatrix} \\
&= \begin{bmatrix} 1.10590597 \\ 1.2249214 \end{bmatrix} \\
\text{Output } A^{(2)} &= \begin{bmatrix} 0.75136507 \\ 0.77292847 \end{bmatrix}
\end{aligned}
$$

Therefore $a_1^{(2)} = o_1 = 0.75136507, a_2^{(2)} = o_2 = 0.77292847$.

(c) Using the mean squared error:

$$
\begin{aligned}
L &= \frac{1}{2}\sum_{k=1}^{2}(o_k - y_k)^2 \\
&= \frac{1}{2}\sum_{k=1}^{2}(a_k^{(2)} - y_k)^2 \\
&= \frac{1}{2}[(0.75136507 - 0.01)^2 + (0.77292847 - 0.99)^2] = 0.29837
\end{aligned}
$$

(d) First note that $\sigma'(z) = \sigma(z)(1 - \sigma(z))$. We want to first calculate $\delta^{(2)}$:

$$
\begin{aligned}
\frac{dL}{do_1} = \frac{dL}{da_1^{(2)}} &= (a_1^{(2)} - y_1) \\
&= (0.75136507 - 0.01) = 0.74136507 \\
\frac{dL}{do_2} = \frac{dL}{da_2^{(2)}} &= (a_2^{(2)} - y_2) \\
&= (0.77292847 - 0.99) = -0.21707153 \\
\Rightarrow \delta^{(2)} &= \begin{bmatrix} 0.74136507 \\ -0.21707153 \end{bmatrix} \odot \begin{bmatrix} a_1^{(2)}(1 - a_1^{(2)}) \\ a_2^{(2)}(1 - a_2^{(2)}) \end{bmatrix} \\
&= \begin{bmatrix} 0.74136507 \\ -0.21707153 \end{bmatrix} \odot \begin{bmatrix} 0.1868156 \\ 0.17551005 \end{bmatrix} \\
&= \begin{bmatrix} 0.13849856 \\ -0.03809824 \end{bmatrix}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\nabla_{W^{(2)}} L &= \delta^{(2)} A^{(1)^T} \\
&= \begin{bmatrix} 0.13849856 \\ -0.03809824 \end{bmatrix} \begin{bmatrix} 0.75136507 & 0.77292847 \end{bmatrix} \\
&= \begin{bmatrix} 0.08216704 & 0.08266763 \\ -0.02260254 & -0.02274024 \end{bmatrix}
\end{aligned}
$$

which means that

$$
\begin{aligned}
\frac{dL}{dw_5} &= \frac{dL}{dw_{11}^{(2)}} = 0.08216704 \\
\frac{dL}{dw_6} &= \frac{dL}{dw_{12}^{(2)}} = 0.08266763 \\
\frac{dL}{dw_7} &= \frac{dL}{dw_{21}^{(2)}} = -0.02260254 \\
\frac{dL}{dw_8} &= \frac{dL}{dw_{22}^{(2)}} = -0.02274024
\end{aligned}
$$

We can continue to calculate $\delta^{(1)}$:

$$
\begin{aligned}
\delta^{(1)} &= [W^{(2)^T} \delta^{(2)}] \odot \sigma'(Z^{(1)}) \\
&= \left[\begin{bmatrix} 0.40 & 0.50 \\ 0.45 & 0.55 \end{bmatrix} \begin{bmatrix} 0.13849856 \\ -0.03809824 \end{bmatrix} \right] \odot A^{(1)} \odot (1 - A^{(1)}) \\
&= \begin{bmatrix} 0.03635031 \\ 0.04137032 \end{bmatrix} \odot \begin{bmatrix} 0.59326999 \\ 0.59688438 \end{bmatrix} \odot \left(1 - \begin{bmatrix} 0.59326999 \\ 0.59688438 \end{bmatrix} \right) \\
&= \begin{bmatrix} 0.00877135 \\ 0.00995425 \end{bmatrix}
\end{aligned}
$$

which implies

$$
\begin{aligned}
\nabla_{W^{(1)}} L &= \delta^{(1)} A^{(0)^T} \\
&= \begin{bmatrix} 0.00877135 \\ 0.00995425 \end{bmatrix} \begin{bmatrix} 0.05 & 0.10 \end{bmatrix} \\
&= \begin{bmatrix} 0.00043857 & 0.00087714 \\ 0.00049771 & 0.00099543 \end{bmatrix}
\end{aligned}
$$

which means that

$$\frac{dL}{dw_1} = \frac{dL}{dw_{11}^{(1)}} = 0.00043857$$
$$\frac{dL}{dw_2} = \frac{dL}{dw_{12}^{(1)}} = 0.00087714$$
$$\frac{dL}{dw_3} = \frac{dL}{dw_{21}^{(1)}} = 0.00049771$$
$$\frac{dL}{dw_4} = \frac{dL}{dw_{22}^{(1)}} = 0.00099543$$

(e)

$$\begin{aligned}
\text{Update } \begin{bmatrix} w_5 & w_6 \\ w_7 & w_8 \end{bmatrix} &= W^{(2)} \leftarrow W^{(2)} - \tau \nabla_{W^{(2)}} L \\
&= \begin{bmatrix} 0.40 & 0.45 \\ 0.50 & 0.55 \end{bmatrix} - \tau \begin{bmatrix} 0.08216704 & 0.08266763 \\ -0.02260254 & -0.02274024 \end{bmatrix} \\
\text{Update } \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \end{bmatrix} &= W^{(1)} \leftarrow W^{(1)} - \tau \nabla_{W^{(1)}} L \\
&= \begin{bmatrix} 0.15 & 0.20 \\ 0.25 & 0.30 \end{bmatrix} - \tau \begin{bmatrix} 0.00043857 & 0.00087714 \\ 0.00049771 & 0.00099543 \end{bmatrix}
\end{aligned}$$

□