UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering
ECE 417 Multimedia Signal Processing

Lecture 19 Sample Problem Solutions

Problem 19.1

1.

$$
\begin{aligned}
a_i &= u_1 x_i + u_0 = x_i \\
y_i &= \sigma(a_i) = \sigma(x_i) \\
&= \{\sigma(-4.97), \sigma(-1), \sigma(1), \sigma(4.97)\} \\
E_i &= \frac{1}{2}\left(y_i - \zeta_i\right)^2 = \frac{1}{2}\left(\sigma(x_i) - \zeta_i\right)^2 \\
&= \left\{\frac{1}{2}(1-\sigma(-4.97))^2, \frac{1}{2}\sigma(-1)^2, \frac{1}{2}(1-\sigma(1))^2, \frac{1}{2}\sigma(4.97)^2\right\} \\
&= \left\{\frac{1}{2}(-\sigma(4.97))^2, \frac{1}{2}\sigma(-1)^2, \frac{1}{2}(-\sigma(-1))^2, \frac{1}{2}\sigma(4.97)^2\right\} \\
&= \left\{\frac{1}{2}(0.99), \frac{1}{2}(0.07), \frac{1}{2}(0.07), \frac{1}{2}(0.99)\right\}
\end{aligned}
$$

$$E = \frac{1}{4}\sum E_i = \frac{1}{8}(2.12) = 0.265$$

2. Suppose, for example, $u_1 = -10000$, and $u_0 = 20000$. Then

$$
\begin{aligned}
a_i &= \{69700, 30000, 10000, -39700\} \\
y_i &= \sigma(a_i) \approx \{1, 1, 1, 0\} \\
E_i &= \frac{1}{2}(y_i - \zeta_i)^2 = \{0, 0.5, 0, 0\} \\
E &= \frac{1}{4}\sum E_i = \frac{1}{8}
\end{aligned}
$$

3.

$$\begin{aligned}
\delta_i = \frac{\partial E_i}{\partial a_i} &= \frac{\partial E_i}{\partial y_i}\frac{\partial y_i}{\partial a_i} \\
&= (y_i - \zeta_i)\sigma'(a_i) = (\sigma(x_i) - \zeta_i)\sigma'(a_i) \\
&= \{(\sigma(-4.97) - 1)\sigma'(-4.97), (\sigma(-1) - 0)\sigma'(-1), (\sigma(1) - 1)\sigma'(1), (\sigma(4.97) - 0)\sigma'(4.97)\} \\
&= \{-\sigma(4.97)\sigma'(4.97), \sigma(-1)\sigma'(-1), -\sigma(-1)\sigma'(-1), \sigma(4.97)\sigma'(4.97)\} \\
&= \{-0.0134, 0.067, -0.067, 0.0134\} \\
\frac{\partial E_i}{\partial u_0} &= \frac{\partial E_i}{\partial a_i}\frac{\partial a_i}{\partial u_0} \\
&= \delta_i \\
\sum \frac{\partial E_i}{\partial u_0} &= 0 \\
\frac{\partial E_i}{\partial u_1} &= \frac{\partial E_i}{\partial a_i}\frac{\partial a_i}{\partial u_1} \\
&= \delta_i x_i \\
&= \{0.067, -0.067, -0.067, 0.067\} \\
\sum \frac{\partial E_i}{\partial u_1} &= 0
\end{aligned}$$

Since $\sum \frac{\partial E_i}{\partial u_0} = 0$, u_0 does not change. Since $\sum \frac{\partial E_i}{\partial u_1} = 0$, u_1 does not change. So the neural net stays in this sub-optimal configuration.

4. For example, if SGD were to randomly choose to present token #1 over and over again, then u_0 would keep changing by $-\frac{\partial E_1}{\partial u_0} = 0.0134$, and u_1 would keep changing by $-\frac{\partial E_1}{\partial u_1} = -0.067$. After u_1 goes negative (and by this time u_0 would be a big positive number), then we would start to present all of the tokens one after the other.