CS 152—Notes on Backpropagation (Rev. 1)

Neil Rhodes

October 7, 2019

1 Backpropagation

Once the forward pass is complete, we have computed the loss, $L(\hat{y}, y)$. Along the way, we've computed $z^{[i]}$ and $a^{[i]}$ for each i from 1 to the number of levels.

Our final goal is to compute $\frac{\partial L}{\partial W^{[i]}}$ and $\frac{\partial L}{\partial b^{[i]}}$ for each i from 1 to the number of levels.

We know the partial derivatives of each step along the way:

$$\frac{\partial L}{\partial \hat{y}}$$
$$\frac{\partial a^{[i]}}{\partial z^{[i]}} = (g^{[i]})'(z^{[i]})$$
$$\frac{\partial z^{[i]}}{\partial a^{[i-1]}} = (W^{[i]})^T$$
$$\frac{\partial z_j^{[i]}}{\partial W_{kj}^{[i]}} = a_j^{[i-1]}$$
$$\frac{\partial z^{[i]}}{\partial b^{[i]}} = 1$$

We *could* use the above formulae to compute any desired $\frac{\partial L}{\partial W^{[i]}}$ or $\frac{\partial L}{\partial b^{[i]}}$. However, there would be much recomputation.

To avoid the recomputations, we use dynamic programming to compute the derivatives in a back-to-front manner (thus the term *backpropagation*).

One piece of notation may be new to you: the Hademard product, $M \odot N$ is the point-wise multiplication of matrices M, and N (of the same dimensions). For example, $\begin{bmatrix} 10 & 3 \end{bmatrix} \odot \begin{bmatrix} 1 & 5 \end{bmatrix} = \begin{bmatrix} 10 & 15 \end{bmatrix}$.

Let's say we have K layers.

Remember that we've defined $\hat{y} = a^{[K]}$, the output of the last layer, and $a^{[0]} = x$, the input.

Here's the order that we'll be computing:

$$\begin{aligned}
\frac{\partial L}{\partial a^{[K]}} &= \frac{\partial L}{\partial \hat{y}} \\
\frac{\partial L}{\partial z^{[K]}} &= \frac{\partial L}{\partial a^{[K]}} \odot (g^{[K]})'(z^{[K]}) \\
\frac{\partial L}{\partial W^{[K]}} &= (a^{[K-1]})^T \frac{\partial L}{\partial z^{[K]}} \\
\frac{\partial L}{\partial b^{[K]}} &= \frac{\partial L}{\partial z^{[K]}} \\
\frac{\partial L}{\partial a^{[K-1]}} &= \frac{\partial L}{\partial z^{[K]}} (W^{[K]})^T \\
\frac{\partial L}{\partial z^{[K-1]}} &= \frac{\partial L}{\partial a^{[K-1]}} \odot (g^{[K-1]})'(z^{[K-1]}) \\
\frac{\partial L}{\partial W^{[K-1]}} &= (a^{[K-2]})^T \frac{\partial L}{\partial z^{[K-1]}} \\
\frac{\partial L}{\partial b^{[K-1]}} &= \frac{\partial L}{\partial z^{[K-1]}} \\
&\vdots \\
\frac{\partial L}{\partial a^{[1]}} &= \frac{\partial L}{\partial z^{[2]}} (W^{[2]})^T \\
\frac{\partial L}{\partial z^{[1]}} &= \frac{\partial L}{\partial a^{[1]}} \odot (g^{[1]})'(z^{[1]}) \\
\frac{\partial L}{\partial W^{[1]}} &= (a^{[0]})^T \frac{\partial L}{\partial z^{[1]}} \\
\frac{\partial L}{\partial b^{[1]}} &= \frac{\partial L}{\partial z^{[1]}}
\end{aligned}$$