David Schultz
4/19/2021
Lab 6
Implementing Gradients

Forward Propagation

$\vec{u} = W_1 \vec{x} + \vec{b}_1$

$\vec{h} = ReLU(\vec{u})$

$\vec{v} = W_2 \vec{h} + \vec{b}_2$

$\vec{o} = softmax(\vec{v}) = e^{\vec{v} - max(\vec{v})} / \sum e^{\vec{v} - max(\vec{v})}$

$L = \frac{1}{2}\|\vec{y} - \vec{o}\|_F^2 \sum \vec{y}(-\log_e(\vec{o}))$

$s_1 = \frac{1}{2}\|W_1\|_F^2 \quad s_2 = \frac{1}{2}\|W_2\|_F^2$

$s = s_1 + s_2 \quad J = L + s$

Backward Propagation

$\frac{\partial J}{\partial L} = 1 \quad \frac{\partial J}{\partial s} = 1 \quad \frac{\partial J}{\partial s_1} = 1 \quad \frac{\partial J}{\partial s_2} = 1$

$\frac{\partial s_1}{\partial W_1} = W_1 \quad \frac{\partial s_2}{\partial W_2} = W_2 \quad \frac{\partial s}{\partial W_1} = W_1 \quad \frac{\partial s}{\partial W_2} = W_2$

$\frac{\partial J}{\partial \vec{o}} = prod\left(\frac{\partial J}{\partial L}, \frac{\partial L}{\partial \vec{o}}\right) = \frac{\partial L}{\partial \vec{o}}$

$\frac{\partial J}{\partial \vec{v}} = prod\left(\frac{\partial J}{\partial \vec{o}}, \frac{\partial \vec{o}}{\partial \vec{v}}\right) = \frac{\partial J}{\partial \vec{o}} \odot softmax'(\vec{v})$

$\frac{\partial J}{\partial W_2} = prod\left(\frac{\partial J}{\partial \vec{v}}, \frac{\partial \vec{v}}{\partial W_2}\right) + prod\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial W_2}\right) = \frac{\partial J}{\partial \vec{v}} \vec{h}^T + W_2$

$\frac{\partial J}{\partial \vec{h}} = prod\left(\frac{\partial J}{\partial \vec{v}}, \frac{\partial \vec{v}}{\partial \vec{h}}\right) = W_2^T \frac{\partial J}{\partial \vec{v}}$

$$\frac{\partial J}{\partial \vec{u}} = \text{prod}\left(\frac{\partial J}{\partial \vec{h}}, \frac{\partial \vec{h}}{\partial \vec{u}}\right) = \frac{\partial J}{\partial \vec{h}} \odot \text{ReLU}'(\vec{u})$$

$$\frac{\partial J}{\partial W_1} = \text{prod}\left(\frac{\partial J}{\partial \vec{u}}, \frac{\partial \vec{u}}{\partial W_1}\right) + \text{prod}\left(\frac{\partial J}{\partial s}, \frac{\partial s}{\partial W_1}\right) = \frac{\partial J}{\partial \vec{u}} \vec{x}^T + W_1$$

Forward Propagation Examples

$\vec{x} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$ $W_1 = \begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \\ 0.1 & 0.4 \end{bmatrix}$ $\vec{b}_1 = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix}$

$\vec{u} = W_1 \vec{x} + \vec{b}_1$

$$\begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \\ 0.1 & 0.4 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 3.3 \\ 3.3 \\ 1.8 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 3.4 \\ 3.5 \\ 2.1 \end{bmatrix}$$

$\vec{h} = \text{ReLU}(\vec{u})$

$$\text{ReLU}\left(\begin{bmatrix} 3.4 \\ 3.5 \\ 2.1 \end{bmatrix}\right) = \begin{bmatrix} 3.4 \\ 3.5 \\ 2.1 \end{bmatrix}$$

$W_2 = \begin{bmatrix} 0.3 & 0.5 & 0.1 \\ 0.2 & 0.7 & 0.8 \end{bmatrix}$ $\vec{b}_2 = \begin{bmatrix} 0.3 \\ 0.5 \end{bmatrix}$

$\vec{y} = W_2 \vec{h} + \vec{b}_2$

$$\begin{bmatrix} 0.3 & 0.5 & 0.1 \\ 0.2 & 0.7 & 0.8 \end{bmatrix} \begin{bmatrix} 3.4 \\ 3.5 \\ 2.1 \end{bmatrix} = \begin{bmatrix} 2.98 \\ 4.81 \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 3.28 \\ 5.31 \end{bmatrix}$$

$\vec{o} = \text{softmax}(\vec{y})$

$$\begin{bmatrix} e^{3.28}/\sum e^{\vec{y}} \\ e^{5.31}/\sum e^{\vec{y}} \end{bmatrix} = \begin{bmatrix} 26.576/228.926 \\ 202.35/228.926 \end{bmatrix} = \begin{bmatrix} 0.116 \\ 0.884 \end{bmatrix}$$

$L = -\sum \vec{y} \log_e(\vec{o})$ $\vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$$-\sum \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \odot \begin{bmatrix} -2.154 \\ -0.123 \end{bmatrix}\right) = 0.123$$

$$s_1 = \frac{1}{2}\|W_1\|_F^2 = \frac{1}{2}\sum\left(\begin{bmatrix}0.3 & 0.5\\ 0.2 & 0.7\\ 0.1 & 0.4\end{bmatrix}^2\right) = 0.52$$

$$s_2 = \frac{1}{2}\|W_2\|_F^2 = \frac{1}{2}\sum\left(\begin{bmatrix}0.3 & 0.5 & 0.1\\ 0.2 & 0.7 & 0.8\end{bmatrix}^2\right) = 0.76$$

$$s = s_1 + s_2 = 0.52 + 0.76 = 1.28$$
$$J = L + s = 0.123 \quad +1.28 = 1.403$$

Backward Propagation Examples

$$\frac{\partial J}{\partial out} = 1.403 \quad \frac{\partial J}{\partial s} = 1.403 \quad \frac{\partial J}{\partial s_1} = 1.403 \quad \frac{\partial J}{\partial s_2} = 1.403$$

$$\frac{\partial J}{\partial L} = 1.403 \quad \frac{\partial s}{\partial W_1} = 1.403\, W_1 \quad \frac{\partial s}{\partial W_2} = 1.403\, W_2$$

~~$\frac{\partial J}{\partial \vec{o}} = \frac{\partial L}{\partial o} 1.403 =$~~ $\frac{\partial J}{\partial \vec{v}} = 1.403(\vec{o} - \vec{y}) = \begin{bmatrix}0.163\\ -0.163\end{bmatrix}$

$$\frac{\partial J}{\partial W_2} = \begin{bmatrix}0.163\\ -0.163\end{bmatrix}\begin{bmatrix}3.4 & 3.5 & 2.1\end{bmatrix} + 1.403\begin{bmatrix}0.3 & 0.5 & 0.1\\ 0.2 & 0.7 & 0.8\end{bmatrix}$$

$$\frac{\partial J}{\partial W_2} = \begin{bmatrix}0.554 & 0.571 & 0.342\\ -0.554 & -0.571 & -0.342\end{bmatrix} + \begin{bmatrix}0.421 & 0.702 & 0.140\\ 0.281 & 0.982 & 1.122\end{bmatrix}$$

$$\frac{\partial J}{\partial W_2} = \begin{bmatrix}0.975 & 1.273 & 0.482\\ -0.273 & 0.411 & 0.78\end{bmatrix} \qquad \frac{\partial J}{\partial \vec{b}_2} = \frac{\partial J}{\partial \vec{v}} = \begin{bmatrix}0.163\\ -0.163\end{bmatrix}$$

$$\frac{\partial J}{\partial \vec{h}} = \begin{bmatrix}0.3 & 0.2\\ 0.5 & 0.7\\ 0.1 & 0.8\end{bmatrix}\begin{bmatrix}0.163\\ -0.163\end{bmatrix} = \begin{bmatrix}0.016\\ -0.033\\ -0.114\end{bmatrix}$$

$$\frac{\partial J}{\partial \vec{u}} = \begin{bmatrix}1\\ 1\\ 1\end{bmatrix} \odot \begin{bmatrix}0.016\\ -0.033\\ -0.114\end{bmatrix} = \begin{bmatrix}0.016\\ -0.033\\ -0.114\end{bmatrix}$$

1.403 W_1

$$\frac{\partial J}{\partial W_1} = \begin{bmatrix} 0.016 \\ -0.033 \\ -0.114 \end{bmatrix} \begin{bmatrix} 6 & 3 \end{bmatrix} + \begin{bmatrix} 1.703 & 1.103 & 1.503 \\ 1.603 & 2.103 & 2.203 \end{bmatrix}$$

$$\frac{\partial J}{\partial W_1} = \begin{bmatrix} 0.096 & 0.048 \\ -0.198 & -0.099 \\ -0.684 & -0.342 \end{bmatrix} + \begin{bmatrix} 0.421 & 0.702 \\ 0.281 & 0.982 \\ 0.140 & 0.561 \end{bmatrix}$$

$$\frac{\partial J}{\partial W_1} = \begin{bmatrix} 0.517 & 0.75 \\ 0.083 & 0.883 \\ -0.544 & 0.219 \end{bmatrix}$$