Forward and Backward Propagation Example

Feb. 10, 2020

Suppose we are doing a regression problem (so our loss function is mean squared error) using a network with the structure:

- Input layer has one feature, X_1
- First hidden layer has two units and a relu activation
- Second hidden layer has two units and a relu activation
- Output layer has one unit and a linear activation

For simplicity, suppose I have just one observation with $X_1^{(1)} = 2$ and $y^{(1)} = 1$.

Also suppose my current estimates of the model parameters are as follows:

- Layer 1:
  - $b_1^{[1]} = 0$, $\left(w_1^{[1]}\right)^T = \begin{bmatrix} 1 \end{bmatrix}$
  - $b_2^{[1]} = 1$, $\left(w_2^{[1]}\right)^T = \begin{bmatrix} -2 \end{bmatrix}$
- Layer 2:
  - $b_1^{[2]} = 0$, $\left(w_1^{[2]}\right)^T = \begin{bmatrix} 0.5 & 10 \end{bmatrix}$
  - $b_2^{[2]} = 1$, $\left(w_2^{[2]}\right)^T = \begin{bmatrix} 1 & 1 \end{bmatrix}$
- Layer 3:
  - $b_1^{[3]} = 0$, $\left(w_1^{[3]}\right)^T = \begin{bmatrix} 1 & 1 \end{bmatrix}$

1. Draw a diagram of this neural network model

2. Forward propagation

Find $a_1^{(1)[1]}$

Find $a_2^{(1)[1]}$

Find $a_1^{(1)[2]}$

Find $a_2^{(1)[2]}$

Find $a_1^{(1)[3]}$

Find the contribution to the MSE from this observation.

If I make a small change to the weight vector for the first unit in layer 2 from $\left(w_1^{[2]}\right)^T = \begin{bmatrix}0.5 & 10\end{bmatrix}$ to $\left(w_1^{[2]}\right)^T = \begin{bmatrix}0.5 & 10.1\end{bmatrix}$, does this affect the model's final prediction for this observation? What can you say about $\frac{\partial J^{(1)}(b,w)}{\partial w_{12}^{[2]}}$?

If I make a small change to the weight vector for the second unit in layer 1 from $\left(w_2^{[1]}\right)^T = \begin{bmatrix}-2\end{bmatrix}$ to $\left(w_2^{[1]}\right)^T = \begin{bmatrix}-2.1\end{bmatrix}$, does this affect the model's final prediction for this observation? What can you say about $\frac{\partial J^{(1)}(b,w)}{\partial w_{21}^{[1]}}$?

If I make a small change to the weight vector for the first unit in layer 2 from $\left(w_1^{[2]}\right)^T = \begin{bmatrix}0.5 & 10\end{bmatrix}$ to $\left(w_1^{[2]}\right)^T = \begin{bmatrix}0.6 & 10\end{bmatrix}$, does this affect the model's final prediction for this observation? What can you say about $\frac{\partial J^{(1)}(b,w)}{\partial w_{11}^{[1]}}$?

If I make a small change to the weight vector for the second unit in layer 2 from $\left(w_2^{[2]}\right)^T = \begin{bmatrix}1 & 1\end{bmatrix}$ to $\left(w_2^{[2]}\right)^T = \begin{bmatrix}1.1 & 1\end{bmatrix}$, does this affect the model's final prediction for this observation? What can you say about $\frac{\partial J^{(1)}(b,w)}{\partial w_{21}^{[2]}}$?

If I make a small change to the weight vector for the first unit in layer 1 from $\left(w_1^{[1]}\right)^T = \begin{bmatrix}1\end{bmatrix}$ to $\left(w_1^{[1]}\right)^T = \begin{bmatrix}1.1\end{bmatrix}$, does this affect the model's final prediction for this observation? What can you say about $\frac{\partial J^{(1)}(b,w)}{\partial w_{11}^{[1]}}$?