Chapter 12: Variations on Backpropagation

Brandon Morgan

1/22/2021

E12.2

From Figure 12.1, we have a single layer network with a *logsig* transfer function. Given the following initial weight and bias, $w(0) = 0,\ b(0) = 0.5$, along with the input target pairs: $p_1 = -2,\ t_1 = 0.8$ and $p_2 = 2,\ t_1 = 1$; compute the SDBP with and without batching.

With Batching

Input p_1

First, we need to calculate our error:

$$a = logsig(wp_1 + b) = logsig(0 + 0.5) = \frac{1}{1 + e^{-0.5}} = 0.62246$$

Then, $e = t_1 - a = 0.8 - 0.62246 = 0.17754$.

Next, we calculate the sensitivity by $s = -2f'(n)e$, where $f' = \frac{e^{-x}}{(1+e^{-x})^2} = a(1-a)$:

$$s = -2(0.62246)(1 - 0.62246)(0.17754) = -0.08344$$

From Eq. (12.7), except without the learning rate, our new weight will be $-sp = 0.08344(-2) = -0.16689$. From Eq. (12.8), except without the learning rate, our new bias will be $-s = 0.08344$. Therefore, the direction of the initial step in the (w, b) plane would be:

$$\begin{bmatrix} -0.16689 \\ 0.08344 \end{bmatrix}$$

Input p_2

First, we need to calculate our error:

$$a = logsig(wp_1 + b) = logsig(0 + 0.5) = \frac{1}{1 + e^{-0.5}} = 0.62246$$

Then, $e = t_1 - a = 1 - 0.62246 = 0.37754$.

Next, we calculate the sensitivity by $s = -2f'(n)e$, where $f' = \frac{e^{-x}}{(1+e^{-x})^2} = a(1-a)$:

$$s = -2(0.62246)(1 - 0.62246)(0.37754) = 0.15864$$

From Eq. (12.7), except without the learning rate, our new weight will be $-sp = 0.15864(2) = 0.3173$. From Eq. (12.8), except without the learning rate, our new bias will be $-s = -0.15864$. Therefore, the direction of the initial step in the (w, b) plane would be:

$$\begin{bmatrix} 0.3173 \\ -0.15864 \end{bmatrix}$$

Average

For batching, we add all the partial gradients for each test input target pair and average them:

$$\frac{1}{2}(\begin{bmatrix} -0.16689 \\ 0.08344 \end{bmatrix} + \begin{bmatrix} 0.3173 \\ -0.15864 \end{bmatrix}) = \begin{bmatrix} 0.07520 \\ -0.0376 \end{bmatrix}$$

This would be the first step of the batch mode SDBP.

Without Batching

Input p_1

First, we need to calculate our error:

$$a = logsig(wp_1 + b) = logsig(0 + 0.5) = \frac{1}{1 + e^{-0.5}} = 0.62246$$

Then, $e = t_1 - a = 0.8 - 0.62246 = 0.17754$.

Next, we calculate the sensitivity by $s = -2f'(n)e$, where $f' = \frac{e^{-x}}{(1+e^{-x})^2} = a(1 - a)$:

$$s = -2(0.62246)(1 - 0.62246)(0.17754) = -0.08344$$

From Eq. (12.7), except without the learning rate, our new weight will be $-sp = 0.08344(-2) = -0.16689$. From Eq. (12.8), except without the learning rate, our new bias will be $-s = 0.08344$. Therefore, the direction of the initial step in the (w, b) plane would be:

$$\begin{bmatrix} -0.16689 \\ 0.08344 \end{bmatrix}$$

Input p_2

First, we need to calculate our error with the new updated weight and bias:

$$a = logsig(wp_1 + b) = logsig(-0.16689(2) + 0.08344) = \frac{1}{1 + e^{-(-0.25034)}} = 0.433774$$

Then, $e = t_1 - a = 1 - 0.433774 = 0.56226$.

Next, we calculate the sensitivity by $s = -2f'(n)e$, where $f' = \frac{e^{-x}}{(1+e^{-x})^2} = a(1 - a)$:

$$s = -2(0.433774)(1 - 0.433774)(0.56226) = -0.27677$$

From Eq. (12.7), except without the learning rate, our new weight will be $-sp = -0.27677(2) = -0.5535$. From Eq. (12.8), except without the learning rate, our new bias will be $-s = 0.27677$. Therefore, the direction of the second step in the (w, b) plane would be:

$$\begin{bmatrix} -0.5535 \\ 0.27677 \end{bmatrix}$$

Compare this final step with the batched averaged step above on the contour plot calculated in E12.1