Batch Normalization Derivatives

Mayank Agarwal

November 4, 2017

Batch Normalization equations are as follows:

$$\mu_B \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \tag{1}$$

$$\sigma_B^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{2}$$

$$\hat{x_i} \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{3}$$

$$y_i \leftarrow \gamma\hat{x_i} + \beta \tag{4}$$

As per equation 4,

$$\frac{\partial l}{\partial \gamma} = \frac{\partial l}{\partial y}\hat{x} \tag{5}$$

and

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{N}\frac{\partial l}{\partial y} \tag{6}$$

Also,

$$\frac{\partial l}{\partial \hat{x}_i} = \frac{\partial l}{\partial y}\gamma \tag{7}$$

$$\frac{\partial l}{\partial \sigma_B^2} = \sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}(x_i - \mu_b)\frac{-1}{2}(\sigma_B^2 + \epsilon)^2 \tag{8}$$

$$\frac{\partial l}{\partial \mu_B} = \sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}\frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} \tag{9}$$

$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i}\frac{1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial l}{\partial \sigma_B^2}\frac{2(x_i - \mu_B)}{m} + \frac{\partial l}{\partial \mu_B}\frac{1}{m} \tag{10}$$

Substituting intermediate derivates in equation 10 with their derivatives.

$$\frac{\partial l}{\partial x_i} = \frac{\gamma}{\sqrt{\sigma_B^2+\epsilon}}\frac{\partial l}{\partial y_i} + [\sum_{i=1}^{m}\frac{\partial l}{\hat{x}_i}(x_i-\mu_B)\frac{-1}{2}(\sigma_B^2+\epsilon)^{-3/2}]\frac{2(x_i-\mu_B)}{m} + \frac{1}{m}\sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}\frac{-1}{\sqrt{\sigma_B^2+\epsilon}} \tag{11}$$

$$\frac{\partial l}{\partial x_i} = \frac{\gamma}{\sqrt{\sigma_B^2+\epsilon}}\frac{\partial l}{\partial y_i} - \frac{(x_i-\mu_B)}{m}\sum_{i=1}^{m}\frac{\gamma\frac{\partial l}{\partial y_i}}{(\sigma_B^2+\epsilon)}\frac{(x_i-\mu_B)}{\sqrt{\sigma_B^2+\epsilon}} - \frac{1}{m}\sum_{i=1}^{m}\gamma\frac{\partial l}{\partial y_i}\frac{-1}{\sqrt{\sigma_B^2+\epsilon}} \tag{12}$$

Let,

$$\delta = \frac{\gamma\frac{\partial l}{\partial y_i}}{\sqrt{\sigma_B^2+\epsilon}}$$

Then,

$$\frac{\partial l}{\partial x_i} = \delta - \frac{(x_i-\mu_B)}{m}\sum_{i=1}^{m}\hat{x}_i\delta - \frac{1}{m}\sum_{i=1}^{m}\delta \tag{13}$$