Batch Normalization Derivation

Given

$$\hat{S}_i = \frac{S_i - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

$$\mu = \frac{1}{m}\sum_{i=1}^{n} S_i$$

$$\sigma^2 = \frac{1}{m}\sum_{i=0}^{n} (S_i - \mu)^2$$

$$\frac{\partial L}{\partial S_i} = \delta\gamma * \frac{\partial \hat{S}_i}{\partial S_i} + \frac{\partial \hat{S}_i}{\partial \mu} * \frac{\partial \mu}{\partial S_i} + \frac{\partial \hat{S}_i}{\partial \sigma^2} * \frac{\partial \sigma^2}{\partial S_i}$$

analyze each of the three components seperately

$1.$

$$\boxed{\frac{\partial \hat{S}_i}{\partial S_i} = \frac{1}{\sqrt{\sigma^2 + \epsilon}}}$$

$2.$

$$\frac{\partial \hat{S}_i}{\partial \mu} * \frac{\partial \mu}{\partial S_i} =$$

$$a)\frac{\partial \hat{S}_i}{\partial \mu} = \frac{-1}{\sqrt{\sigma^2 + \epsilon}}$$

$$b)\frac{\partial \mu}{\partial S_i} = \frac{1}{N}$$

$$\boxed{\frac{\partial \hat{S}_i}{\partial \mu} * \frac{\partial \mu}{\partial S_i} = \frac{-1}{N\sqrt{\sigma^2 + \epsilon}}}$$

3.

$$\frac{\partial \hat{S}_i}{\partial \sigma^2} * \frac{\partial \sigma^2}{\partial S_i}$$

$$a)\frac{\partial \hat{S}_i}{\partial \sigma^2} = \frac{-1}{2}(S_i - \mu)(\sigma^2 + \epsilon)^{\frac{-3}{2}}$$

$$b)\frac{\partial \sigma^2}{\partial S_i} = \frac{2}{N}$$

$$\boxed{\frac{\partial \hat{S}_i}{\partial \sigma^2} * \frac{\partial \sigma^2}{\partial S_i} = \frac{-S_i^2}{N\sqrt{\sigma^2 + \epsilon}}}$$

Putting together

$$\frac{\partial L}{\partial S_i} = \delta\gamma * \frac{\partial \hat{S}_i}{\partial S_i} + \frac{\partial \hat{S}_i}{\partial \mu} * \frac{\partial \mu}{\partial S_i} + \frac{\partial \hat{S}_i}{\partial \sigma^2} * \frac{\partial \sigma^2}{\partial S_i}$$

$$= \delta\gamma * \left(\frac{1}{\sqrt{\sigma^2 + \epsilon}} + \frac{-1}{N\sqrt{\sigma^2 + \epsilon}} + \frac{-S_i^2}{N\sqrt{\sigma^2 + \epsilon}}\right)$$

$$= \frac{\delta\gamma}{\sqrt{\sigma^2 + \epsilon}} * \left(1 + \frac{-1}{N} + \frac{-S_i^2}{N}\right)$$

$$= \boxed{\frac{\delta\gamma}{N\sqrt{\sigma^2 + \epsilon}} * \left(N - 1 + -S_i^2\right)}$$