3/19/2019 Untitled2

## **Batch Normalization Derivation**

## Given

$$egin{aligned} \hat{S_i} &= rac{S_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \ & \mu = rac{1}{m} \sum_{i=1}^n S_i \ & \sigma^2 &= rac{1}{m} \sum_{i=0}^n \left( S_i - \mu 
ight)^2 \ & rac{\partial L}{\partial S_i} &= \delta \gamma * rac{\partial \hat{S_i}}{\partial S_i} + rac{\partial \hat{S_i}}{\partial \mu} * rac{\partial \mu}{\partial S_i} + rac{\partial \hat{S_i}}{\partial \sigma^2} * rac{\partial \sigma^2}{\partial S_i} \end{aligned}$$

## analyze each of the three components seperately

1.

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

9

$$rac{\partial \hat{S_i}}{\partial \mu} * rac{\partial \mu}{\partial S_i} =$$

$$a)rac{\partial \hat{S_i}}{\partial \mu} = rac{-1}{\sqrt{\sigma^2 + \epsilon}}$$

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

$$\left[rac{\partial \hat{S_i}}{\partial \mu}*rac{\partial \mu}{\partial S_i} = rac{-1}{N\sqrt{\sigma^2+\epsilon}}
ight]$$

3/19/2019 Untitled2

3.

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

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

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

$$\left|rac{\partial \hat{S_i}}{\partial \sigma^2}*rac{\partial \sigma^2}{\partial S_i} = rac{-S_i^2}{N\sqrt{\sigma^2+\epsilon}}
ight|$$

## **Putting together**

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

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

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

$$= \overline{\left[rac{\delta \gamma}{N\sqrt{\overline{\sigma^2}+\epsilon}}*\left(N-1+-S_i^2
ight)
ight]}$$