代码再:torch/lib/THNN/generic/BatchNormalization.c

下面写BatchNormalization.c 中bp推导.

gradient_output记做go

论文中6个梯度可以写作:

$$\frac{\partial L}{\partial \hat{x}} = g_0 \cdot W$$

$$\frac{\partial L}{\partial \sigma^2} = g_0 \cdot \hat{x} W \cdot -\frac{1}{2}\sigma^{-3}$$

$$\frac{\partial L}{\partial \mu} = -\frac{g_0 W}{\sigma}$$

$$A = \frac{\partial L}{\partial x} = \frac{g_0 W}{\sigma} + g_0 \hat{x}^2 W \cdot [-1] \cdot \frac{1}{\sigma^3 m} + \frac{-g_0 W}{\sigma m}$$

代码115行为

$$B = \frac{W}{\sigma}\left(g_0 - \frac{g_0}{n} - \frac{\hat{x}(x) g_0}{\sigma^2 n}\right)$$

n=m 那么A=B证毕.

证明中利用了这个式子 $\frac{\partial \ell}{\partial \sigma_{\mathcal{B}}^2} \cdot \frac{\sum_{i=1}^{m} -2(x_i - \mu_{\mathcal{B}})}{m} = 0$