Manual Backpropagation through BatchNorm

Dienstag, 31. Oktober 2023 10:33

Algorithm:

$y_i = \gamma \hat{x}_i + \beta$

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

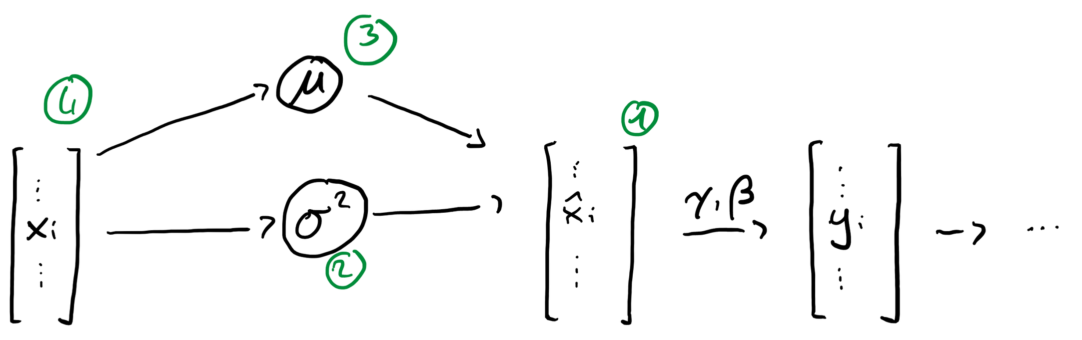
$\mu = \frac{1}{m} \sum x_i$

$\sigma^2 = \frac{1}{m-1} \sum (x_i - \mu)^2$

Given: $\frac{dL}{dy_i}$

Need: $\frac{dL}{dx_i}$

How do the variables depend on each other?



(1): $\frac{dL}{d\hat{x}_i} = \frac{dL}{dy_i} \cdot \frac{dy_i}{d\hat{x}_i}$

$= \frac{dL}{dy_i} \cdot \gamma$

(2): $\frac{dL}{d\sigma^2} = \sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{d\hat{x}_i}{d\sigma^2}$

$= \sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{d}{d\sigma^2} \left(\frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \right)$

$= -\sum_i \frac{dL}{d\hat{x}_i} \cdot (x_i - \mu) \cdot \frac{d}{d\sigma^2} \left(\sqrt{\sigma^2 + \epsilon} \right) \cdot \left(- \frac{1}{\sigma^2 + \epsilon} \right)$

$= -\sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{x_i - \mu}{\sigma^2 + \epsilon} \cdot \frac{d}{d\sigma^2} (\sigma^2 + \epsilon) \cdot \frac{1}{2\sqrt{\sigma^2 + \epsilon}}$

$= -\sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{x_i - \mu}{\sigma^2 + \epsilon} \cdot \frac{1}{2 \cdot \sqrt{\sigma^2 + \epsilon}}$

$= -\sum_i \frac{dL}{d\hat{x}_i} \frac{x_i - \mu}{2} \cdot (\sigma^2 + \epsilon)^{-1} \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}}$

$= -\sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{1}{2} \cdot (x_i - \mu) \cdot (\sigma^2 + \epsilon)^{-\frac{3}{2}}$

$= -\frac{\gamma}{2} \sum_i \frac{dL}{dy_i} \cdot (x_i - \mu) \cdot (\sigma^2 + \epsilon)^{-\frac{3}{2}}$

(3) $\frac{dL}{d\mu} = \sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{d\hat{x}_i}{d\mu} + \frac{dL}{d\sigma^2} \cdot \frac{d\sigma^2}{d\mu}$

= see below

$\frac{d\hat{x}_i}{d\mu} = \frac{d}{d\mu} \left(\frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \right)$

$= \frac{-1}{\sqrt{\sigma^2 + \epsilon}}$

$\frac{d\sigma^2}{d\mu} = \frac{d}{d\mu} \left(\frac{1}{m-1} \sum_i (x_i - \mu)^2 \right)$

$= \frac{2}{m-1} \cdot \sum_i (\mu - x_i)$

$= \frac{2}{m-1} \cdot \left[-m \cdot \mu + \sum_i x_i \right]$

$= \frac{2}{m-1} \left(-m \cdot \frac{1}{m} \sum_i x_i + \sum_i x_i \right)$

$= \frac{2}{m-1} \cdot 0$

$= 0$

$\frac{dL}{d\mu} = \sum_i \frac{dL}{d\hat{x}_i} \cdot \frac{d\hat{x}_i}{d\mu} + \frac{dL}{d\sigma^2} \cdot \frac{d\sigma^2}{d\mu}$

$= \frac{-1}{\sqrt{\sigma^2 + \epsilon}} \cdot \sum_i \frac{dL}{d\hat{x}_i} + 0$ | see (1)

$= \frac{-\gamma}{\sqrt{\sigma^2 + \epsilon}} \cdot \sum_i \frac{dL}{dy_i}$

$= -\gamma \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \cdot \sum_i \frac{dL}{dy_i}$

(4) $\frac{dL}{dx_i} = \frac{dL}{d\mu} \cdot \frac{d\mu}{dx_i} + \frac{dL}{d\sigma^2} \cdot \frac{d\sigma^2}{dx_i} + \frac{dL}{d\hat{x}_i} \cdot \frac{d\hat{x}_i}{dx_i}$

= see below

$\frac{d\mu}{dx_i} = \frac{d}{dx_i} \left(\frac{1}{m} \sum_{j=1}^{m} x_j \right)$

$= \frac{1}{m}$

$\frac{d\sigma^2}{dx_i} = \frac{d}{dx_i} \left(\frac{1}{m-1} \sum (x_i - \mu)^2 \right)$

$= \frac{1}{m-1} \cdot \frac{d}{dx_i} \left((x_i - \mu)^2 \right)$

$= \frac{1}{m-1} \cdot (2x_i - 2\mu)$

$= \frac{2}{m-1} \cdot (x_i - \mu)$

$\frac{d\hat{x}_i}{dx_i} = \frac{d}{dx_i} \left(\frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \right)$

$= \frac{1}{\sqrt{\sigma^2 + \epsilon}}$

$\frac{dL}{dx_i} = \frac{dL}{d\mu} \cdot \frac{d\mu}{dx_i} + \frac{dL}{d\sigma^2} \cdot \frac{d\sigma^2}{dx_i} + \frac{dL}{d\hat{x}_i} \cdot \frac{d\hat{x}_i}{dx_i}$

$= -\gamma \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \cdot \sum_j \frac{dL}{dy_j} \cdot \frac{1}{m} - \frac{\gamma}{2} \sum_j \frac{dL}{dy_j} \cdot (x_j - \mu) \cdot (\sigma^2 + \epsilon)^{-\frac{3}{2}} \cdot \frac{2}{m-1} \cdot (x_i - \mu) + \frac{dL}{dy_i} \cdot \gamma \cdot \frac{1}{\sqrt{\sigma^2 + \epsilon}}$

$= (\sigma^2 + \epsilon)^{-\frac{1}{2}} \left[-\frac{\gamma}{m} \cdot \sum_{j=1}^{m} \frac{dL}{dy_j} - \frac{\gamma}{2} \cdot (\sigma^2 + \epsilon)^{-1} \cdot \frac{2}{m-1} \cdot (x_i - \mu) \sum_{j=1}^{m} \frac{dL}{dy_j} \cdot (x_j - \mu) + \frac{dL}{dy_i} \cdot \gamma \right.$

$= \gamma \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \left[-\frac{1}{m} \cdot \sum_{j=1}^{m} \frac{dL}{dy_j} - \frac{1}{m-1} \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \cdot (x_i - \mu) \sum_{j=1}^{m} \frac{dL}{dy_j} \cdot (x_j - \mu) + \frac{dL}{dy_i} \right]$

$= \gamma \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \left[-\frac{1}{m} \sum_{j=1}^{m} \frac{dL}{dy_j} - \frac{1}{m-1} \cdot \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \cdot \sum_{j=1}^{m} \frac{dL}{dy_j} \cdot \frac{x_j - \mu}{\sqrt{\sigma^2 + \epsilon}} + \frac{dL}{dy_i} \right]$

$= \gamma \cdot (\sigma^2 + \epsilon)^{-\frac{1}{2}} \left[\frac{dL}{dy_i} - \frac{1}{m} \sum_{j=1}^{m} \frac{dL}{dy_j} - \frac{\hat{x}_i}{m-1} \cdot \sum_{j=1}^{m} \frac{dL}{dy_j} \cdot \hat{x}_j \right]$