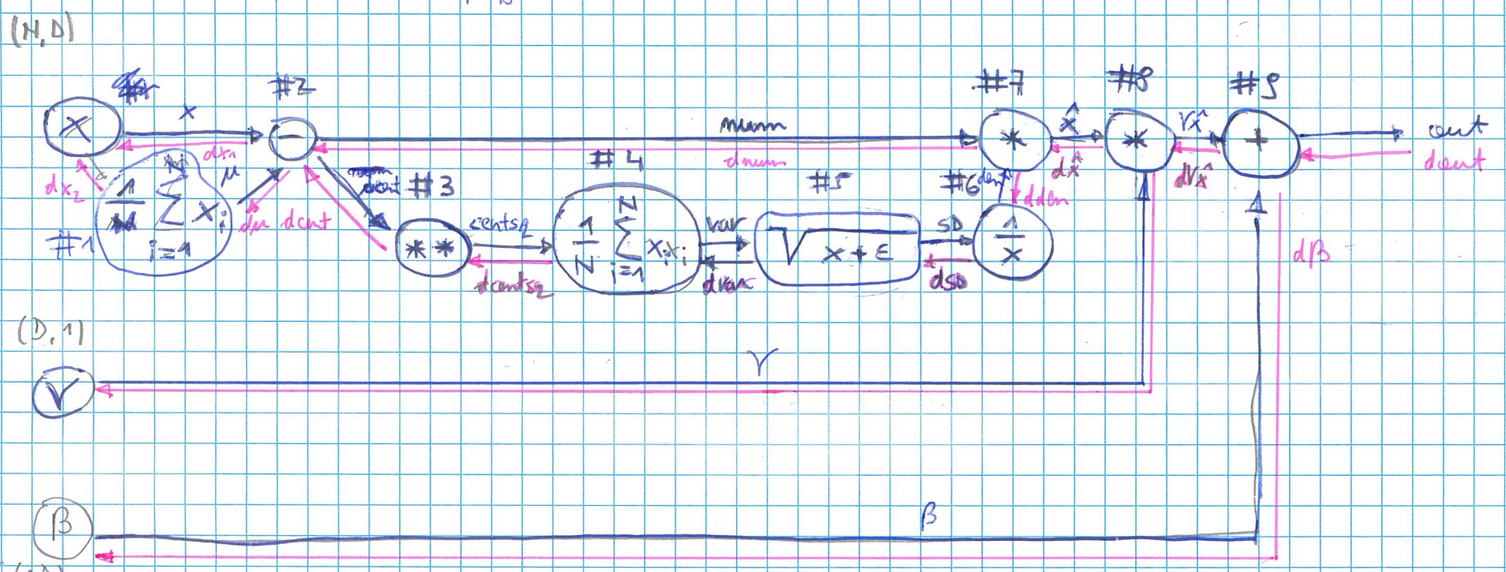
Batch norm forward propagation and back propagation

Scale and shift. mini-batch mean mini-batch variance

Batchnorm: $BN(\gamma,\beta) = \gamma \hat{x}_i + \beta$ $\quad \mu_{\mathcal{B}} = \frac{1}{m}\sum_{i=1}^{m} x_i$ $\quad \sigma_{\mathcal{B}}^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_{\mathcal{B}})^2$

$\hat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$ normalize



DRAFT (1D)

#9: $\partial\beta = 1 \cdot d\text{cent}$

#9 $\partial\gamma\hat{x} = 1 \cdot d\text{cent}$

#8 $\partial\gamma = \hat{x} \cdot \partial\gamma\hat{x}$

#8 $\partial\hat{x} = \gamma \cdot \partial\gamma\hat{x}$

#7 $\partial\text{num} = \text{den} \cdot \partial\hat{x}$

#7 $\partial\text{den} = \text{num} \cdot \partial\hat{x}$

#6 $\partial\text{invr} = -\frac{1}{SD^2} \cdot \partial\text{den}$

#5 $\partial\text{var} = \frac{1}{2\sqrt{\text{var}}} \cdot \partial SD$

#4 $\partial\text{centsq} = \frac{1}{N} \cdot (1) \cdot \partial\text{var}$

#3 $\partial\text{cent} = 2\,\text{cent} \cdot \partial\text{centsq}$

#2 $\partial x = 1 \cdot \partial\text{cent}$

#2 $\partial\mu = -1 \cdot (\partial\text{cent} + \partial\text{num})$

#1 $\partial x = \frac{1}{N} \cdot N \cdot \partial\text{cent}$

Vector case

$\frac{5}{3} \to \begin{bmatrix}5&4&3\\2&2&2\end{bmatrix} - \begin{bmatrix}3&4&5\\1&2&1\end{bmatrix} \Rightarrow (N,D)$

$\circ$ Hadamard product

$d\text{cent} = \underline{1} \quad (N,D)$

#9 $\partial\beta = \underline{1} \cdot \sum_{i=1}^{N} d\text{cent}^T \quad (D,1)$ (N,D)

#9 $\partial\gamma\hat{x} = 1 \circ d\text{cent} \quad (N,D)$

#8 $\partial\hat{x} = \underset{(D,1)^T}{\gamma^T} * \underset{(N,D)}{\partial\gamma\hat{x}} \quad (N,D)$

#8 $\partial\gamma = \left(\sum_{i=1}^{N} \hat{x}_i * \partial\gamma\hat{x}_i\right)^T \quad (D,1)$

↳ Average gradient over all examples.

#7 $\partial\text{num} = \underset{(1,D)}{\text{den}} * \underset{(N,D)}{\partial\hat{x}} \quad (N,D)$

#7 $\partial\text{den} = \sum_{i=1}^{N} \underset{(N,D)}{\text{num}_i} \circ \underset{(N,D)}{\partial\hat{x}_i} \quad (1,D)$

#6 $\partial SD = -\frac{1}{\underset{(1,D)}{SD^2}} \circ \underset{(1,D)}{\partial\text{den}} \quad (1,D)$

#5 $\partial\text{var} = \frac{1}{2 * \text{sqrt}(\text{var})} * \underset{(1,D)}{\partial SD} \quad (1,D)$

#4 $\partial\text{centsq} = \frac{1}{N} \cdot \underset{(N,D)}{[1]} * \underset{(1,D)}{\partial\text{var}}$

#3 $\partial\text{cent} = 2 \cdot \underset{(N,D)}{\text{cent}} \cdot \underset{(N,D)}{d\text{cent}} \quad (N,D)$

#2 $\partial x = 1 \cdot (\partial\text{cent} + \partial\text{num}) \quad (N,D)$

#2 $\partial\mu = -1 \sum_{i=1}^{M} \underset{(1,D)}{(\partial\text{cent} + \partial\text{num})_i} \quad (1,D)$

↳ Need to take row sum.

+

#$\partial x = \frac{1}{N} \underset{(N,D)}{[1]} * \underset{(1,D)}{\partial\mu}$