$$V = \frac{1}{N}\sum_{k=1}^{N}(x_k-\mu)^2, \quad \mu = \frac{1}{N}\sum_{k=1}^{N}x_k, \qquad \text{for } \frac{\partial \mu}{\partial X} = \frac{1}{N}\begin{bmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{bmatrix}^{(N,D)}$$

$$\sigma = \sqrt{V+\varepsilon}, \quad Y = \frac{X-\mu}{\sigma}$$

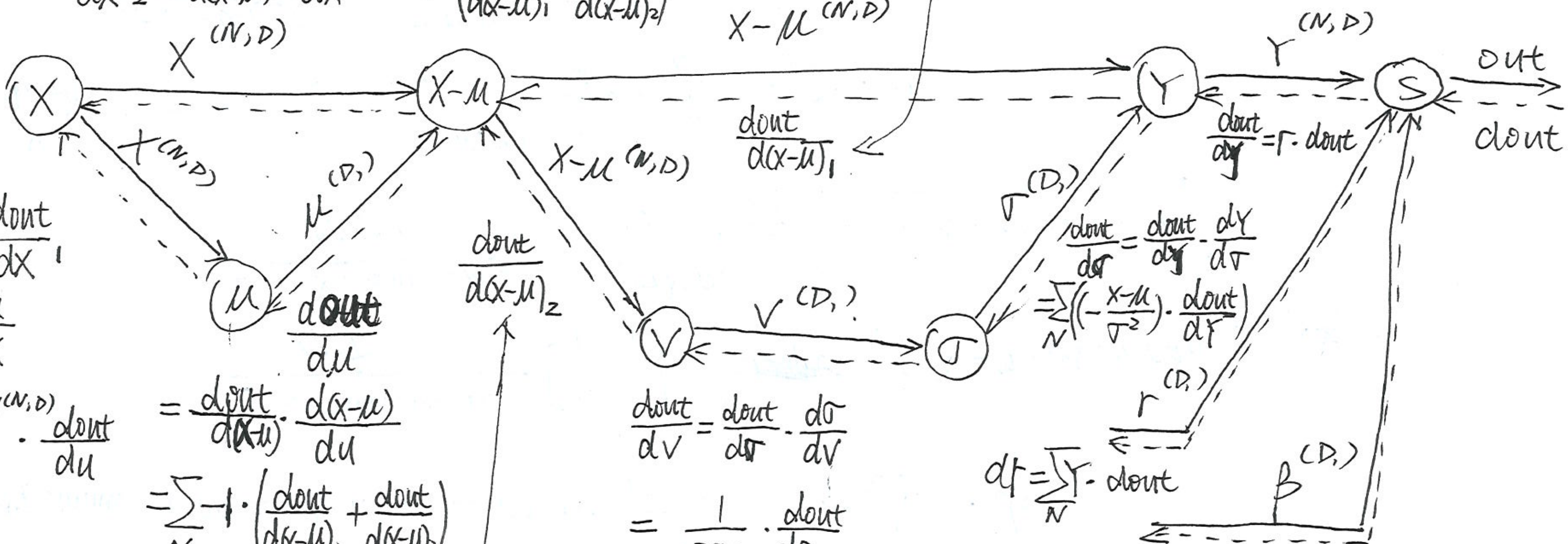
S: Scale function, $out = \Gamma \cdot Y + \beta$

$$\frac{d\sigma}{dV} = \frac{1}{2}\frac{1}{\sqrt{V+\varepsilon}} = \frac{1}{2\sigma}, \quad \frac{dY}{d\sigma} = -\frac{X-\mu}{\sigma^2}$$

$$\frac{dout}{d(X-\mu)_1} = \frac{dout}{dY}\cdot\frac{dY}{d(X-\mu)} = \frac{1}{\sigma}\begin{bmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{bmatrix}^{(N,D)}\cdot\frac{dout}{dY}$$

$$\frac{dout}{dX_2} = \frac{dout}{d(X-\mu)}\cdot\frac{d(X-\mu)}{dX} = 1\cdot\left(\frac{dout}{d(X-\mu)_1}+\frac{dout}{d(X-\mu)_2}\right)$$

$$dX = \frac{dout}{dX_1} + \frac{dout}{dX_2}$$



$$\frac{dout}{dX_1} = \frac{dout}{d\mu}\cdot\frac{d\mu}{dX} = \frac{1}{N}\begin{bmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{bmatrix}^{(N,D)}\cdot\frac{dout}{d\mu}$$

$$\frac{dout}{d\mu} = \frac{dout}{d(X-\mu)}\cdot\frac{d(X-\mu)}{d\mu} = \sum_N -1\cdot\left(\frac{dout}{d(X-\mu)_1}+\frac{dout}{d(X-\mu)_2}\right)$$

$$\frac{dout}{dV} = \frac{dout}{d\sigma}\cdot\frac{d\sigma}{dV} = \frac{1}{2\sigma}\cdot\frac{dout}{d\sigma}$$

$$\frac{dout}{d\sigma} = \frac{dout}{dY}\cdot\frac{dY}{d\sigma} = \sum_N\left(\left(-\frac{X-\mu}{\sigma^2}\right)\cdot\frac{dout}{dY}\right)$$

$$\frac{dout}{dY} = \Gamma\cdot dout$$

$$d\Gamma = \sum_N Y\cdot dout$$

$$d\beta = \sum_N Y\cdot dout$$

$$\frac{dout}{d(X-\mu)_2} = \frac{dout}{dV}\cdot\frac{dV}{d(X-\mu)} = \frac{2}{N}(X-\mu)\begin{bmatrix} 1 & \cdots & 1 \\ 1 & \cdots & 1 \end{bmatrix}^{(N,D)}\cdot\frac{dout}{dV}$$

Batch Normalization Alternative V