Batch Normalization $\longrightarrow$

samples

$X \longrightarrow (m, n_x)$

$Y \longrightarrow (m, n_y)$

$$\mu_l = \sum_{p}^{M} x_{pl} \cdot \frac{1}{N} \quad ; \quad \sigma^2 = \frac{1}{N} \sum_{p}^{M} (x_{pl} - \mu_l)^2$$

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} \longrightarrow \hat{x}_{kl} = \frac{x_{kl} - \mu_l}{\sqrt{\sigma_l^2 + \varepsilon}}$$

$$y = \gamma \hat{x} + \beta \longrightarrow \begin{pmatrix} y_{11} & y_{12} & \cdots \end{pmatrix} = \begin{pmatrix} \gamma_1 \hat{x}_{11} + \beta_1 & \gamma_2 \hat{x}_{12} + \beta_2 & \cdots \end{pmatrix}$$

$$y_{kl} = \gamma_l \hat{x}_{kl} + \beta_l \longrightarrow \frac{\partial y_{kl}}{\partial x_{ij}} = \gamma_l (\delta_{ik} \delta_{jk}) \longrightarrow \text{Element wise.}$$

$$\frac{\partial L}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{l=1}^{n_y} \frac{\partial L}{\partial y_{kl}} \cdot \frac{\partial y_{kl}}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{l=1}^{n_y} \frac{\partial L}{\partial y_{kl}} \cdot \frac{\partial y_{kl}}{\partial \hat{x}_{kl}} \cdot \frac{\partial \hat{x}_{kl}}{\partial x_{ij}}$$

* Just translation $\longrightarrow$

$$\hat{x}_{kl} = x_{kl} - \mu_l = x_{kl} - \frac{1}{N} \sum_{p}^{M} x_{pl}$$

$$\frac{\partial \hat{x}_{kl}}{\partial x_{ij}} = \begin{vmatrix} 1 - \frac{1}{N} & \text{if } k = i,\ l \neq j \\ -1/N & \text{if } k \neq i,\ l = j \\ 0 & \text{Otherwise} \end{vmatrix} = \delta_{lj} \left(\delta_{ik} - \frac{1}{N} \right)$$

* Variance $\longrightarrow$

$$\sigma_\ell^2 = \sum_p^M (x_{p\ell} - \mu_\ell)^2 \cdot \frac{1}{N}$$

$$\frac{\partial \sigma_\ell^2}{\partial x_{ij}} = \frac{1}{N} \cdot 2 \sum_p^M \frac{\partial (x_{p\ell} - \mu_\ell)}{\partial x_{ij}} \cdot (x_{p\ell} - \mu_\ell) = \frac{2\, \delta_{\ell j}}{N} \sum_p^M (x_{p\ell} - \mu_\ell)(\delta_{p,i} - \frac{1}{N})$$

$$= \frac{2\, \delta_{\ell j}}{N} \Big[(x_{i\ell} - \mu_\ell) + \underbrace{\frac{1}{N} \sum_p^M \mu_\ell - \frac{1}{N} \sum_p^M x_{p\ell}}_{\frac{N}{N}\mu_\ell \,-\, \mu_\ell} \Big] =$$

$$= \frac{2}{N} \delta_{\ell,j} (x_{i\ell} - \mu_\ell)$$

* Norm $\longrightarrow$

$$\frac{\partial \hat{x}_{k\ell}}{\partial x_{ij}} = \frac{\partial (x_{k\ell} - \mu_\ell)}{\partial x_{ij}} (\sigma^2 + \varepsilon)^{-1/2} - \frac{1}{2} \frac{\partial \sigma^2}{\partial x_{ij}} (x_{k\ell} - \mu_\ell)(\sigma_\ell^2 + \varepsilon)^{-3/2}$$

$$= \delta_{\ell j} (\delta_{k,i} - \frac{1}{N})(\sigma_\ell^2 + \varepsilon)^{-1/2} - \frac{1}{N} \delta_{\ell,j} (x_{i\ell} - \mu_\ell)(x_{k\ell} - \mu_\ell)(\sigma_\ell^2 + \varepsilon)^{-3/2}$$

* Cost function $\longrightarrow$

$$\frac{\partial h}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{\ell=1}^{n_y} \frac{\partial h}{\partial y_{k\ell}} \cdot \frac{\partial y_{k\ell}}{\partial \hat{x}_{k\ell}} \cdot \frac{\partial \hat{x}_{k\ell}}{\partial x_{ij}} =$$

$$= \sum_{k=1}^{M} \sum_{\ell=1}^{n_y} \frac{\partial h}{\partial y_{k\ell}} \cdot \gamma_\ell \cdot \delta_{\ell j} (\sigma^2 + \varepsilon)^{-1/2} \left[\delta_{k,i} - \frac{1}{N} - \frac{1}{N} (x_{i\ell} - \mu_\ell)(x_{k\ell} - \mu_\ell)(\sigma_\ell^2 + \varepsilon)^{-1} \right]$$

$$= \frac{\gamma_j}{N}(\sigma_j^2+\varepsilon)^{-1/2}\left[\frac{\partial L}{\partial y_{ij}} - \sum_{k=1}^{M}\frac{\partial L}{\partial y_{kj}} - \sum_{k=1}^{M}\frac{\partial L}{\partial y_{kj}}(x_{ij}-\mu_j)(x_{kj}-\mu_j)(\sigma_j^2+\varepsilon)^{-1/2}\right]$$

$$= \frac{\gamma_j}{N}(\sigma_j^2+\varepsilon)^{-1/2}\left[\frac{\partial L}{\partial y_{ij}} - \sum_{k=1}^{M}\frac{\partial L}{\partial y_{kj}} - (x_{ij}-\mu_j)(\sigma_j^2+\varepsilon)^{-1/2}\sum_{k=1}^{M}\frac{\partial L}{\partial y_{kj}}(x_{kj}-\mu_j)\right]$$

$$\frac{\partial L}{\partial \gamma_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{n_y}\frac{\partial L}{\partial y_{kl}}\cdot\frac{\partial y_{kl}}{\partial \gamma_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{n_y}\frac{\partial L}{\partial y_{kl}}\cdot\delta_{lj}\cdot\hat{x}_{kl} =$$

$$= \sum_{k=1}^{M}\frac{\partial L}{\partial y_{kj}}\cdot\hat{x}_{kj}$$

$$\frac{\partial L}{\partial \beta_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{n_y}\frac{\partial L}{\partial y_{kl}}\frac{\partial y_{kl}}{\partial \beta_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{n_y}\frac{\partial L}{\partial y_{kl}}\cdot\delta_{lj} = \sum_{k=1}^{M}\frac{\partial L}{\partial y_{kj}}$$