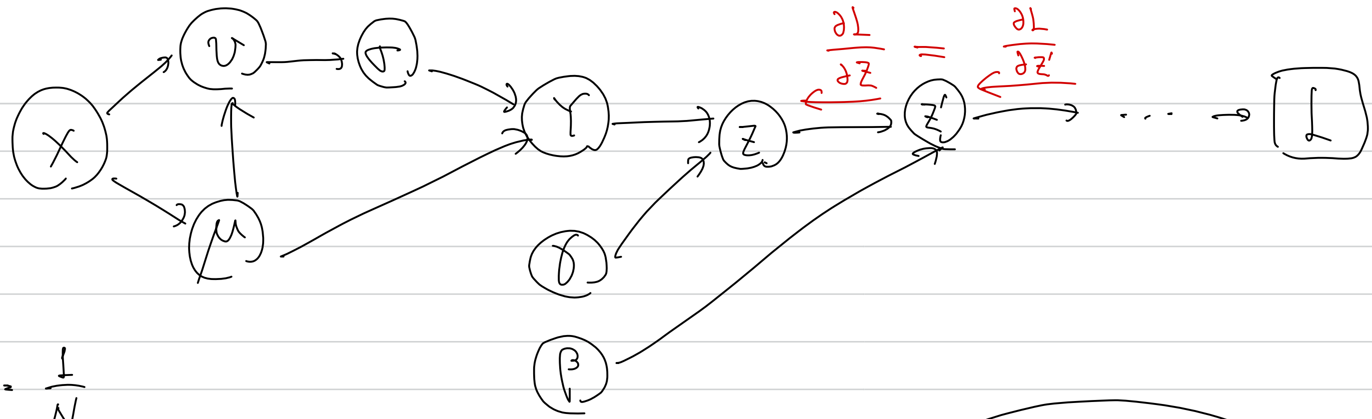

$$\frac{\partial \mu}{\partial X_i} = \frac{1}{N}$$

$$\frac{\partial v}{\partial X_i} = \frac{2}{N}\sum_{k=1}^{N}(x_k-\mu)\left(\delta_{ik}-\frac{1}{N}\right)$$

$$= \frac{2}{N}(x_i-\mu) - \frac{2}{N^2}\underbrace{\sum_{k=1}^{N}(x_k-\mu)}_{= N\mu - N\mu = 0}$$

$$= \frac{2}{N}(x_i-\mu)$$

μ is a function that takes X_i as input

$$\frac{\partial \sigma}{\partial X_i} = \frac{\partial \sigma}{\partial v}\cdot\frac{\partial v}{\partial X_i}$$

$$= \frac{1}{2\sqrt{v+\epsilon}}\cdot\frac{2}{N}(x_i-\mu)$$

$$= \frac{x_i-\mu}{N\sigma}$$

$$\frac{\partial Y_j}{\partial X_i} = \frac{1}{\sigma^2}\cdot\left\{\frac{\partial (X_j-\mu)}{\partial X_i}\cdot\sigma - (X_j-\mu)\frac{\partial \sigma}{\partial X_i}\right\}$$

$$= \frac{1}{\sigma^2}\left\{\left(\delta_{ij}-\frac{1}{N}\right)\cdot\sigma - \frac{1}{N\sigma}\cdot(X_j-\mu)(X_i-\mu)\right\}$$

$$= \frac{1}{\sigma}\left\{\delta_{ij} - \frac{1}{N} - \frac{1}{N}\cdot\frac{(X_j-\mu)}{\sigma}\cdot\frac{(X_i-\mu)}{\sigma}\right\}$$

$$= \frac{1}{\sigma}\left\{\delta_{ij} - \frac{1}{N}(1+y_i y_j)\right\}$$

$$\frac{\partial L}{\partial Y_j} = \gamma\cdot\frac{\partial L}{\partial Z_j}$$

$$\frac{\partial L}{\partial X_i} = \sum_{j=1}^{N}\frac{\partial L}{\partial Y_j}\cdot\frac{\partial Y_j}{\partial X_i}$$

$$= \frac{\gamma}{\sigma}\sum_{j=1}^{N}\frac{\partial L}{\partial Z_j}\cdot\left\{\delta_{ij}-\frac{1}{N}(1+y_i y_j)\right\}$$

$$= \frac{\gamma}{\sigma}\left\{\frac{\partial L}{\partial Z_i} - \frac{1}{N}\sum_{j=1}^{N}(1+y_i y_j)\frac{\partial L}{\partial Z_j}\right\} = \frac{\gamma}{\sigma}\left(\frac{\partial L}{\partial Z_i} - \frac{1}{N}\sum_{j=1}^{N}\frac{\partial L}{\partial Z_j} - \frac{y_i}{N}\underbrace{\sum_{j=1}^{N} y_j\frac{\partial L}{\partial Z_j}}_{= \frac{dL}{d\gamma}}\right)$$