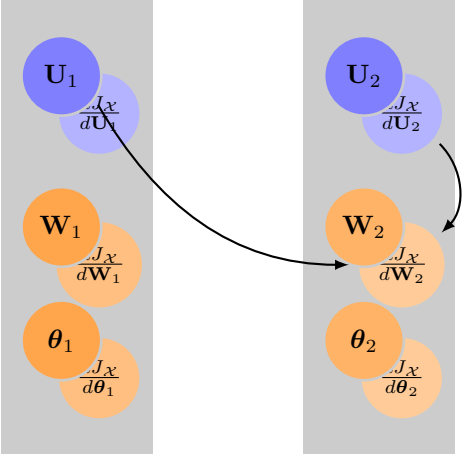
$$\mathbf{U}_1 = f_1\left(\mathbf{X}\mathbf{W}_1^T + \boldsymbol{\theta}_1\right) \qquad \mathbf{U}_2 = f_2\left(\mathbf{U_1}\mathbf{W}_2^T + \boldsymbol{\theta}_2\right)$$

$$\begin{bmatrix} x_{0,0} & \dots & x_{0,m} \\ x_{1,0} & \dots & x_{1,m} \\ \vdots & & \vdots \\ x_{n,0} & \dots & x_{n,m} \end{bmatrix}$$



$$J_\mathcal{X}(\mathbf{y}, \hat{\mathbf{y}})$$

$$\frac{dJ_\mathcal{X}}{d\mathbf{W}_2} = \left(\mathbf{f}_2' \odot \frac{dJ_\mathcal{X}}{d\mathbf{U}_2}\right)^T \mathbf{U}_2$$