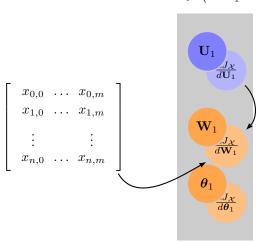
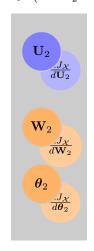
$$\mathbf{U}_1 = f_1 \left(\mathbf{X} \mathbf{W}_1^T + \boldsymbol{\theta}_1 \right) \quad \mathbf{U}_2 = f_2 \left(\mathbf{U}_1 \mathbf{W}_2^T + \boldsymbol{\theta}_2 \right)$$





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

 $\frac{dJ_{\mathcal{X}}}{d\mathbf{W}_{1}} = \left(\mathbf{f}_{1}' \odot \frac{dJ_{\mathcal{X}}}{d\mathbf{U}_{1}}\right)^{T} \mathbf{X}$