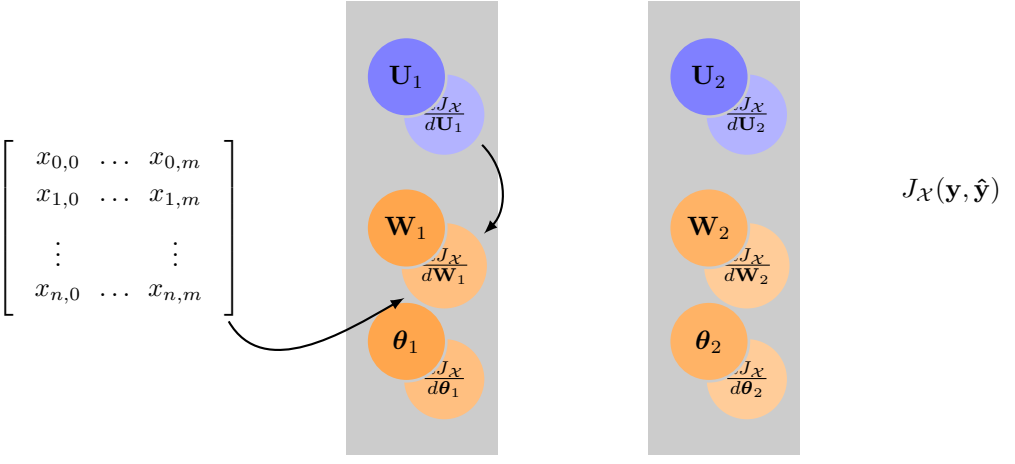
$$\mathbf{U}_1 = f_1\left(\mathbf{X}\mathbf{W}_1^T + \boldsymbol{\theta}_1\right) \qquad \mathbf{U}_2 = f_2\left(\mathbf{U_1}\mathbf{W}_2^T + \boldsymbol{\theta}_2\right)$$



$$\frac{dJ_\mathcal{X}}{d\mathbf{W}_1} = \left(\mathbf{f}_1' \odot \frac{dJ_\mathcal{X}}{d\mathbf{U}_1}\right)^T \mathbf{X}$$