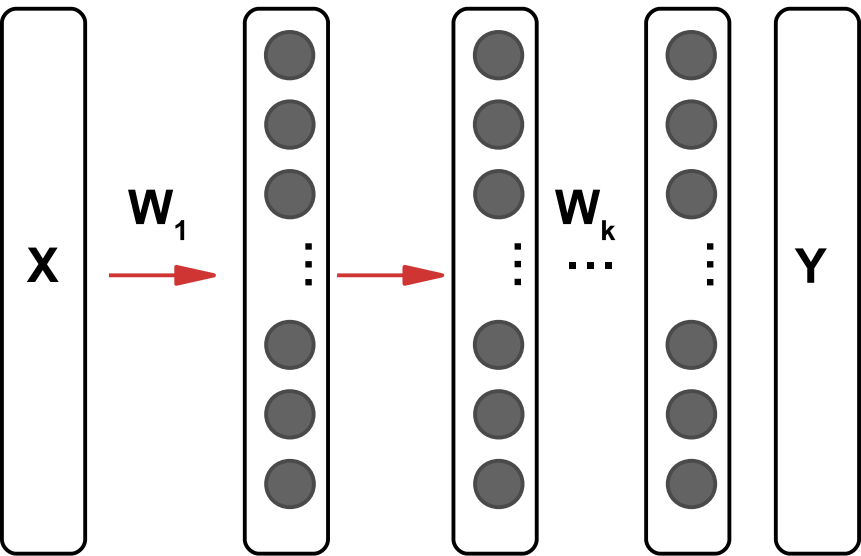


Forward pass: activations $f(X; \mathbf{W})$



$\nabla_{\mathbf{W}} \mathcal{L}_j$

Backward pass: gradients