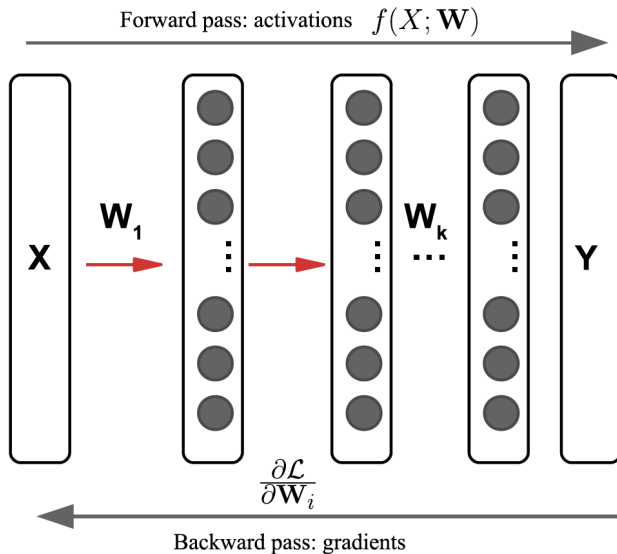


Update steps

$$\mathbf{W}_{t+1} \leftarrow \mathbf{W}_t - \eta \nabla_{\mathbf{W}} \mathcal{L}_B$$

$$\mathcal{L}_{\mathbf{W}_{t+1}} \leq \mathcal{L}_{\mathbf{W}_t} \leq \dots$$



Total Nr. Parameters: **60M**

PREDICTION

832M : Total Nr. FLOPS

