Some Terms and Background
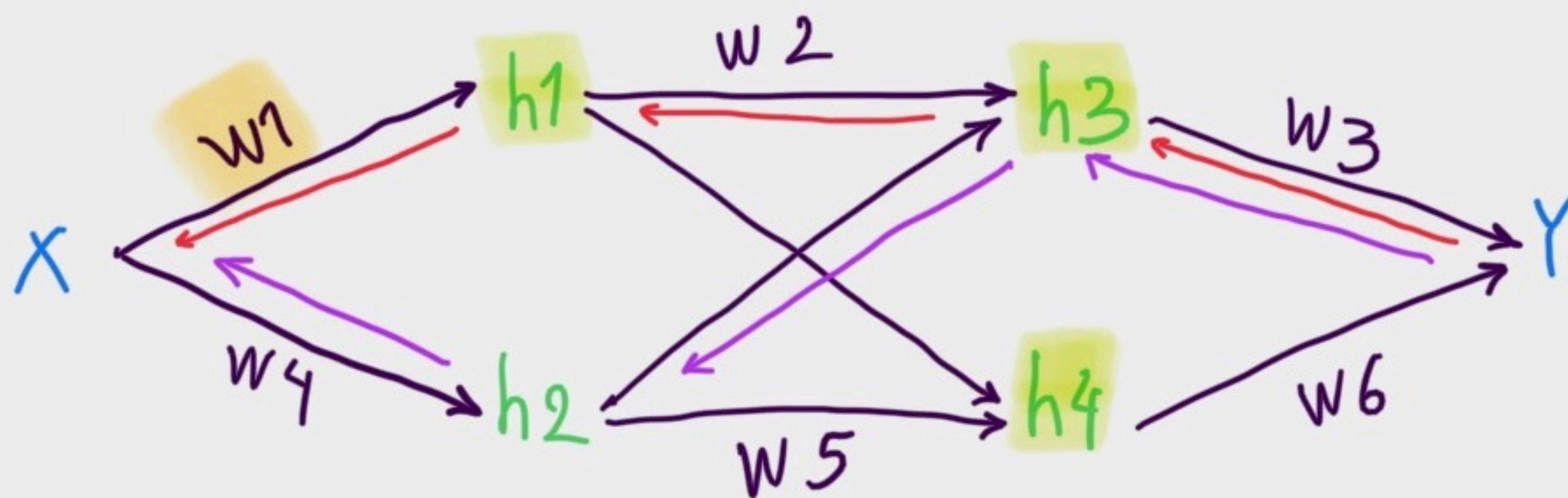
- Why is mini-batch training good?

  1) reduces the complexity of training process

  2) reduces noise

- What is the update rule for weight matrix W1?

$$\frac{\partial y}{\partial w_1} = \frac{\partial y}{\partial h_3} \frac{\partial h_3}{\partial h_1} \frac{\partial h_1}{\partial w_1} + \frac{\partial y}{\partial h_4} \frac{\partial h_4}{\partial h_1} \frac{\partial h_1}{\partial w_1}$$

red path   purple path