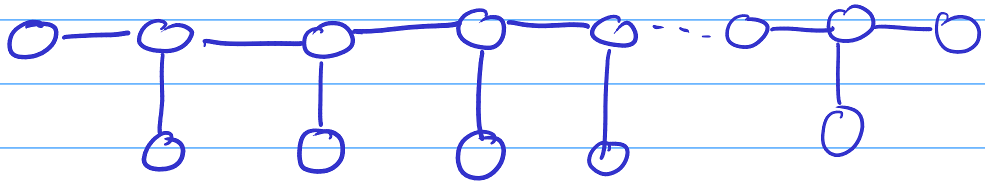
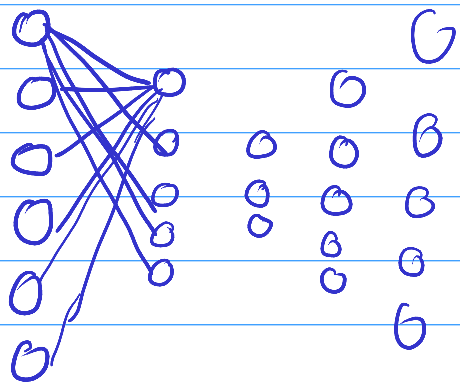
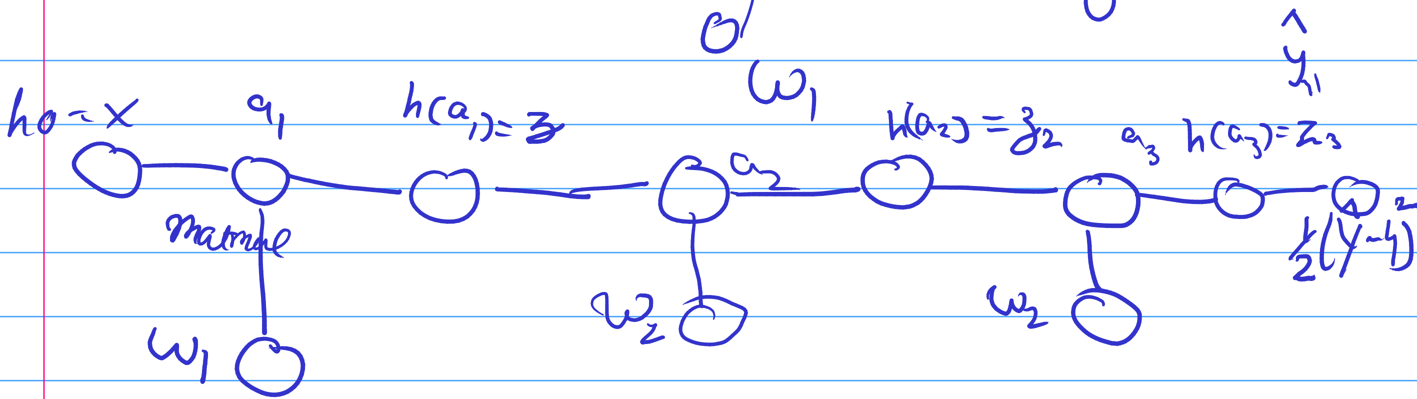
$$E(\omega) = \boxed{\qquad\qquad\qquad}$$







$$\omega^{(k+1)} \leftarrow \omega^{(k)} - \eta \left.\frac{\partial E}{\partial \omega}\right|_{\omega=\omega^k}$$

$$\hookrightarrow \nabla E(\omega)\Big|_{\omega=\omega k}$$

$$\nabla_A(z) = \nabla_C(z)\,\nabla_A(C) = GB^T$$

$$\nabla_C z_{20\times 1} = \Big[\qquad\Big]_{20\times 1}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,10} \\ a_{21} & a_{22} & \cdots & a_{2,10} \\ \cdots & & & a_{20,10} \end{bmatrix}_{20\times 10} \rightarrow \nabla_A (y)$$

So they flatten it $\rightarrow 200 \times 1$

find derivative

and then reshape

treat matrix as a vector (vectorize)

$$\frac{\partial}{\partial \alpha} \sigma(\alpha) = \sigma(\alpha)\,\sigma(1$$