Problem 3

a)

784 features because
shape = 1000, 28, 28 → collapse

Input layer
784

Feeds back in

↑ learning rate   cost

after ~~the first~~ each pass we go back and use gradient descent to optimize
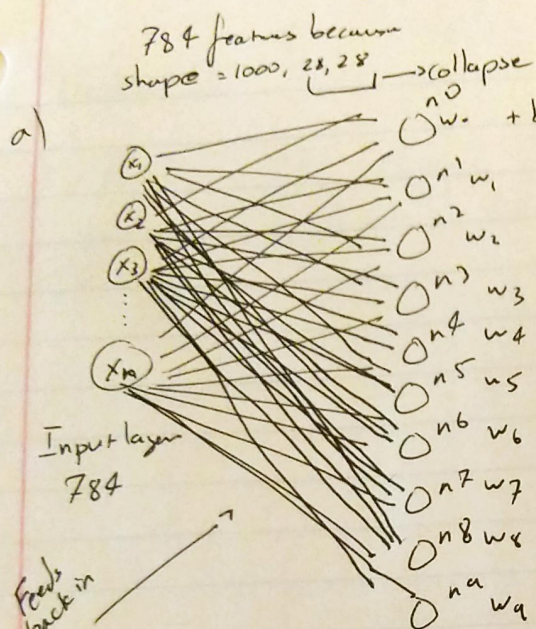
Only 1 layer of neurons, since l_1 rest
→ so its actually the output layer
using sigmoid neurons
activation fxn → $f(z) = \frac{1}{1 + e^{-z}}$

cost → $L(\hat{y}^{(i)}, y^{(i)}) =$
$$-\left(y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)})\right)$$

each neuron is connected to all inputs

take neuron 0 for example

$$x_1 W_0 + x_2 W_0 + \ldots + x_{784} W_0$$

where the w_0 corresponds to n_0, we also add the bias term

b) Unless the cost was minimized, would make it as NaN and attempt to reclassify by ~~adjusting the~~ the adjusted parameters.