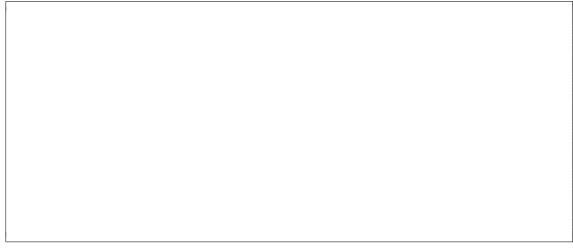
## Our problems multiply

2. (10 points) We will consider a neural network with a slightly unusual structure. Let the input x be  $d \times 1$  and let the weights be represented as  $k \times 1 \times d$  vectors,  $W^{(1)}, \ldots, W^{(k)}$ . Then the final output is

$$\hat{y} = \prod_{i=1}^k \sigma(W^{(i)}x) = \sigma(W^{(1)}x) \times \cdots \times \sigma(W^{(k)}x) .$$

Define  $a^{(j)} = \sigma(W^{(j)}x)$ .

(a) What is  $\partial L(\hat{y}, y)/\partial a^{(j)}$  for some j? Since we have not specified the loss function, you can express your answer in terms of  $\partial L(\hat{y}, y)/\partial \hat{y}$ .



(b) What are the dimensions of  $\partial a^{(j)}/\partial W^{(j)}$ ?

(c) What is  $\partial a^{(j)}/\partial W^{(j)}$ ? (Recall that  $d\sigma(v)/dv = \sigma(v)(1-\sigma(v))$ .)

Name: \_\_\_\_\_