1. Read Deep Learning: An Introduction for Applied Mathematicians. Consider a network as defined in (3.1) and (3.2). Assume that $n_L = 1$, find an algorithm to calculate $\nabla a^{[L]}(x)$.

We follow the notations in the above paper.

First, for $2 \le l \le L$, we have

$$\frac{\partial a^{[l]}}{\partial a^{[l-1]}} = \frac{\partial a^{[l]}}{\partial z^{[l]}} \times \frac{\partial z^{[l]}}{\partial a^{[l-1]}} = D^{[l]}_{n_l \times n_l} W^{[l]}_{n_l \times n_{l-1}}$$

Since $\nabla a^{[L]} = \frac{\partial a^{[L]}}{\partial a^{[1]}} = \frac{\partial a^{[L]}}{\partial a^{[L-1]}} \times \frac{\partial a^{[L-1]}}{\partial a^{[L-2]}} \times \cdots \times \frac{\partial a^{[2]}}{\partial a^{[1]}}$, we have $\nabla a^{[L]} = D^{[L]}_{n_L \times n_L} W^{[L]}_{n_L \times n_{L-1}} D^{[L-1]}_{n_{L-1} \times n_{L-1}} W^{[L-1]}_{n_{L-1} \times n_{L-2}} \cdots D^{[2]}_{n_2 \times n_2} W^{[2]}_{n_2 \times n_1}$

To calculate $\nabla a^{[L]}$, we can follow the steps below.

1. Compute $z^{[l]} = W^{[l]} a^{[l-1]} + b^{[l]}$ for $2 \le l \le L$.
2. Compute $\sigma'(z^{[l]}) = \sigma(z^{[l]})(1 - \sigma(z^{[l]}))$ for $2 \le l \le L$.
3. Compute $\prod_{k=2}^{L} D^{[k]} W^{[k]}$

$$= \prod_{k=2}^{L} \begin{pmatrix} \sigma'(z_1^{[k]}) & & 0 \\ & \ddots & \\ 0 & & \sigma'(z_{n_k}^{[k]}) \end{pmatrix} \begin{pmatrix} w_{11}^{[k]} & \cdots & w_{1,n_{k-1}}^{[k]} \\ \vdots & & \vdots \\ w_{n_k,1}^{[k]} & \cdots & w_{n_k,n_{k-1}}^{[k]} \end{pmatrix}$$

$$= \nabla a^{[L]}.$$

2. There are unanswered questions during the lecture, and there are likely more questions we haven't covered. Take a moment to think about them and write them down here.

In the locally weighted linear regression (LWLR), how to choose the parameter τ ?