1

a.

$$\frac{\partial y_i}{\partial x_j}$$
$$= \frac{\partial(\sigma(W_{i1}x_1 + \ldots + W_{id}x_d))}{\partial x_j}$$
$$= \frac{\partial(\frac{1}{1+e^{-(W_{i1}x_1+\ldots+W_{id}x_d)}})}{\partial x_j}$$
$$= \frac{e^{-(W_{i1}x_1+\ldots+W_{id}x_d)}W_{ij}}{(1+e^{-(W_{i1}x_1+\ldots+W_{id}x_d)})^2}$$

let

$$q_i = \frac{e^{-(W_{i1}x_1+\ldots+W_{id}x_d)}}{(1+e^{-(W_{i1}x_1+\ldots+W_{id}x_d)})^2}$$

$$\frac{\partial Y}{\partial X} = \begin{bmatrix} q_1W_{11} & \ldots & q_1W_{1d} \\ \ldots & \ldots & \ldots \\ q_nW_{n1} & \ldots & q_nW_{nd} \end{bmatrix}$$
$$= \begin{bmatrix} q_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & q_n \end{bmatrix} * \begin{bmatrix} W_{11} & \ldots & W_{1d} \\ \ldots & \ldots & \ldots \\ W_{n1} & \ldots & W_{nd} \end{bmatrix}$$

calculate

$$let\ z = Wx$$
$$\sigma\prime(z_i) = q_i$$
$$\sigma\prime(z) = \begin{bmatrix} q_1 \\ q_2 \\ \ldots \\ q_n \end{bmatrix}$$

calculate

$$\frac{\partial Y}{\partial X} = diag(\sigma^{'}) * W$$

b.Derive the quantity $\frac{\partial L}{\partial W} = \sum_{t=0}^{T}\sum_{k=1}^{t}\frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W}$

根据连式法则：

$$h_k = f_1(x;W)$$
$$h_t = f_2(y_1, W_2)$$
$$L_t = Loss(h_t, h_{GT})$$
$$\text{于是有：}$$
$$\frac{\partial L_t}{\partial W_1}\frac{\partial L_t}{\partial W_2}$$
$$\frac{\partial L_t}{\partial W_2} = (\frac{\partial L_t}{\partial h_t})(\frac{\partial h_t}{\partial W_2})$$
$$\frac{\partial L_t}{\partial W} = (\frac{\partial L_t}{\partial h_t})(\frac{\partial h_k}{\partial W})$$
$$\frac{L_t}{W} = \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W}$$
$$\text{于是直到}T\text{次有：}$$
$$\frac{\partial L}{\partial W} = \sum_{t=0}^{T}\sum_{k=1}^{t}\frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W}$$

2.

a.

当T=3：

$$\frac{\partial L}{\partial W} = \sum_{t=0}^{3}\sum_{k=1}^{t}\frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W}$$
$$= \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W} + \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W} + \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W}$$
$$+ \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W} + \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W} + \frac{\partial L_t}{\partial h_t}\frac{\partial h_t}{\partial h_k}\frac{\partial h_k}{\partial W}$$

b.

$$
\begin{gathered}
M^n = M^{n-1}M \\
M = QAQ^{-1} \\
M^n = M^{n-1}QAQ^{-1} \\
= M^{n-2}QAQ^{-1}QAQ^{-1} \\
= M^{n-2}QA^2Q^{-1} \\
\dots \\
= QA^nQ^{-1}
\end{gathered}
$$

c.

$$
A^{30} = \begin{bmatrix} 0.9^{30} & 0 \\ 0 & 0.4^{30} \end{bmatrix} w^{30} = \begin{bmatrix} 0.6 * 0.9^{30} & 0.8 * 0.4^{30} \\ 0.8 * 0.9^{30} & 0.6 * 0.4^{30} \end{bmatrix}
$$

分析：通过计算矩阵的 30 次方最后矩阵的值都会趋于 0,如果特征值的绝对值都小于1则在矩阵n次方后，特征值会趋近于0，所以计算结果趋近于0，但是如果一个特征值大于1，那么在指数增长下，对应的列会趋近于无穷。

3.

a.

三个函数都是LSTMs中的门函数，用于保护和控制单元的状态。每一个门函数其基础函数都是sigmoid函数。其中i_t, o_t在LSTMs中起到控制状态信息存储的功能。
f_t:遗忘层，它对每一个C_{t1}生成一个0到1之间的数，1表示完全保留，0表示完全放弃。
i_t:输入层，生成0到1之间的数，从而决定要更新的值，并且在tanh层创建新的候选值，添加到里面，之后将两个结合，创建一个更新。
o_t:输出层，生成-1到1之间的数，决定单元状态中哪一个需要被输出，在经过tanh之后，利用sigmoid函数相乘，即可得到对应需要的输出。

b.

因为 f t , i t , o t 总是非负数, 并且取值范围为 [0,1],由于 f t , i t , o t 都是属于 sigmoid 函数,则对应的值输出区间为 [0,1], 其余函数由于与 tanh 函数有关,取值范围在 [-1,1] 之间。

c.

因为$\frac{\partial C_t}{\partial C_k} = \prod_{i=k+1}^{t} \frac{\partial C_t}{\partial C_{t-1}}$
所以由$f_t = 1, i_t = 0$可以得:

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \overline{C_t}$$
$$C_t = C_{t-1}$$

所以:

$$\frac{\partial C_t}{\partial C_k} = \prod_{i=k+1}^{t} \frac{\partial C_t}{\partial C_{t-1}}$$
$$= \prod_{i=k+1}^{t} 1$$