Answer to CS224n Assignment 3

Pengfei Gao, Fanny Yang, Hao Yin [*]

Problem 1

(a) *(TBW)*

    i J.P. Morgan could either be a PERor a ORG ; Stanford can either be a ORG or be a PER.

    ii Because sometimes we need contexts. For example, if you say George Stanford is a great man. Then Stanford is a PER. If you say Stanford University, then Stanford is a ORG.

    iii The first letter of that word is capital or not; There is no 'a','the' in front of that word

(b)   i $e \in \mathbb{R}^{1 \times (2w+1)D}$, $W \in \mathbb{R}^{(2w+1)D \times H}$, $U \in \mathbb{R}^{H \times C}$.

    ii For each time step, forming e takes $O((2w+1)D)$, computing h takes $O((2w+1)DH)$, then computing $\hat{y}$ takes $O(HC)$. Therefore, for each time step, the total time for predicting label complexity is $O(H((2w+1)D+C))$. Now for a sentence of length T, the total time complexity is

$$O(HT((2w+1)D+C)).$$

(c) Finished! Need to compare with Junzi.

(d) *(TBW)*

Problem 2

Ask Pengfei.

Problem 3

(a)   i $w_h = 1$, $u_h = 1$, $b_h = 0$.

    ii $w_z = 1$, $u_z = 0$, $u_h = 1$, and then w_h can be any number.

(b)   i We show that the parameter values that realizes this behavior do not exist.

Suppose $h^{(t-1)} = 0$, then $x^{(t)} = 1$ must lead to $h^{(t)} = 1$ while $x^{(t)} = 0$ must lead to $h^{(t)} = 0$, thus

$$
\begin{aligned}
u_h + b_h &> 0, \\
b_h &\leq 0,
\end{aligned}
$$

from which we know that $u_h > 0$. Now suppose $h^{(t-1)} = 1$, then $x^{(t)} = 1$ must lead to $h^{(t)} = 0$ while $x^{(t)} = 0$ must lead to $h^{(t)} = 1$, thus

$$
\begin{aligned}
u_h + w_h + b_h &\leq 0, \\
w_h + b_h &> 0,
\end{aligned}
$$

from which we know that $u_h < 0$. Therefore, we must have $u_h > 0$ and $u_h < 0$, which is a contradiction.

ii $b_r = 1$, $w_z = 1$, $u_z = -1$, $w_h = -1$, $u_h = 1$.