Vocab size : 250 words

Embedding dim : $d = 10$

WordReps : 250×10

$$x_1 \quad x_2 \quad \ldots \quad x_{10}$$

word1
word2
$\vdots$
word250

Train i^{th} 4-gram: word 1-4 (e.g., I will be there)

1-of-k encoding of word4

True label: $\bigcirc \quad \bigcirc \quad \ldots \quad \bigcirc$
$t_{i,1} \quad t_{i,2} \quad \quad\quad t_{2,250}$ Compare

Softmax

Output Layer
$(h_k^{(2)})$ $\quad \bigcirc \quad \bigcirc \quad \ldots \quad \bigcirc$
$h_1^{(2)} \quad h_2^{(2)} \quad\quad h_{250}^{(2)}$

$W_{50 \times 250}^{(2)}$ (hidToOut)

Hidden Layer
$(h_j^{(1)})$ $\quad \bigcirc \quad \bigcirc \quad \ldots \quad \bigcirc$
$h_1^{(1)} \quad h_2^{(1)} \quad\quad h_{30}^{(1)}$

$W_{30 \times 50}^{(1)}$ (repToHid)

Input Layer $\bigcirc \quad \bigcirc \ldots \bigcirc \quad \bigcirc \ldots \bigcirc \quad \bigcirc \ldots \bigcirc$
$x_1 \quad x_2 \quad\quad x_{10} \quad x_{11} \quad\quad x_{20} \quad x_{21} \quad\quad x_{30}$

$\underbrace{\qquad\qquad}_{word1} \qquad \underbrace{\qquad\qquad}_{word2} \qquad \underbrace{\qquad\qquad}_{word3}$