Assignment 3 for CS224d

Lifu Huang

April 2016

1 RNN's(Recursive Neural Network)

(a)

$$
\begin{aligned}
\delta^{(s)} &= \hat{y} - y \\
\delta^{(1)} &= f'(h^{(1)}) \circ (U^T \delta^{(s)} + \delta_{above}) \\
\delta_{below} &= (W^{(1)})^T \delta^{(1)} \\
\nabla_U J &= \delta^{(s)} (h^{(1)})^T \\
\nabla_{b^{(s)}} J &= \delta^{(s)} \\
\nabla_{W^{(1)}} J &= \delta^{(1)} \left[(h^{(1)}_{left})^T \ (h^{(1)}_{right})^T \right] \\
\nabla_{b^{(1)}} J &= \delta^{(1)} \\
\nabla_{\left[L^T_{left} \ L^T_{right}\right]^T} J &= \delta_{below}
\end{aligned}
$$

(b)

Please see code files for details.

(c)

(a)

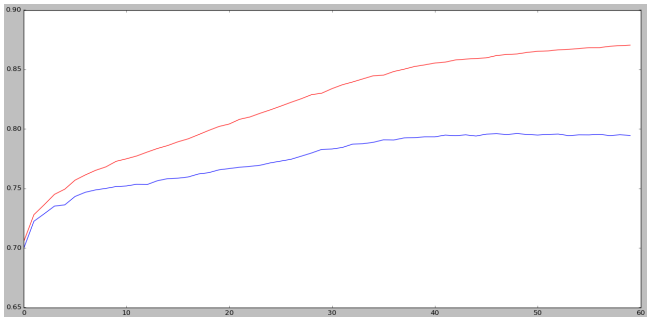

Figure 1: Accuracy on Training and Dev Set over Epochs

(b)

Beacause training for too many epochs may lead to the problem of over fitting.

(c)

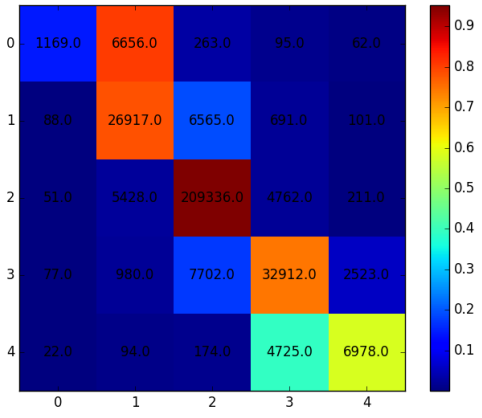

Figure 2: Confusion Matrix on Training Set

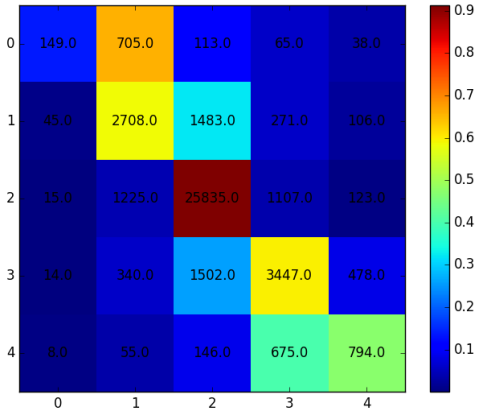

Figure 3: Confusion Matrix on Dev Set

(d)

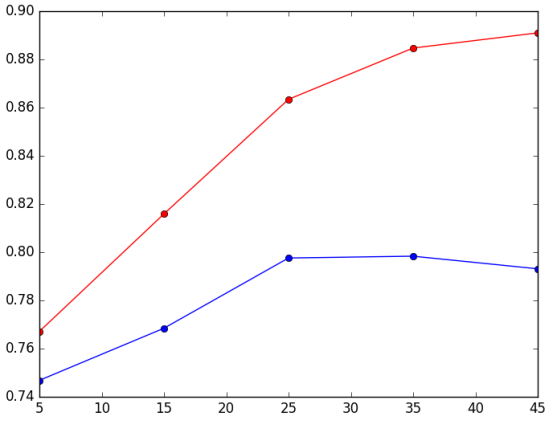

Figure 4: Accuracy on Training and Dev Set over wvecDims

2 2-Layer Deep RNN's

(a)

(b)

(c)

(d)

(e)

(f)

3 Extra Credit: Recursive Neural Tensor Networks