Duplicate Question

Embedding/Linear/Sigmoid

(600,1) → FC → BCE

not a sequence model

$$L(y, \hat{y}) = \sum_{i=1}^{N} (y_i log(\hat{y}) + (1 - y_i) log(1 - \hat{y}))$$

infersent

our modifications: dont train on SNLI

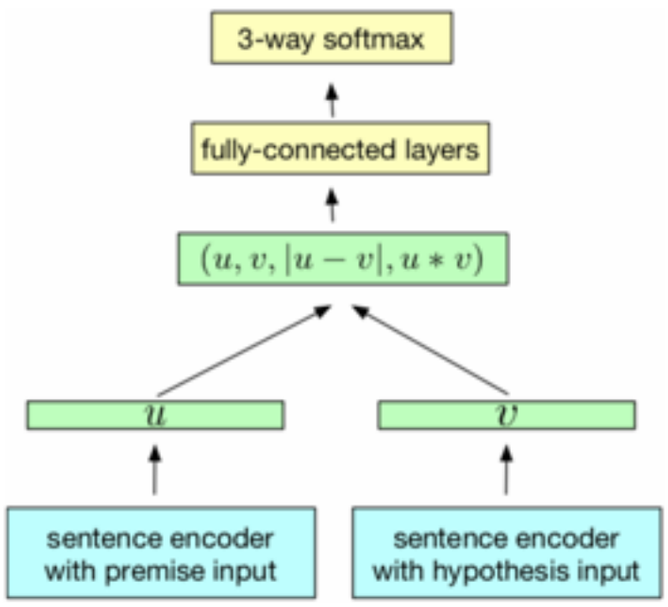


Figure 1: **Generic NLI training scheme.**

$$L(y, \hat{y}) = -\sum_{class=1}^{numclasses} y_i log(\hat{y})$$

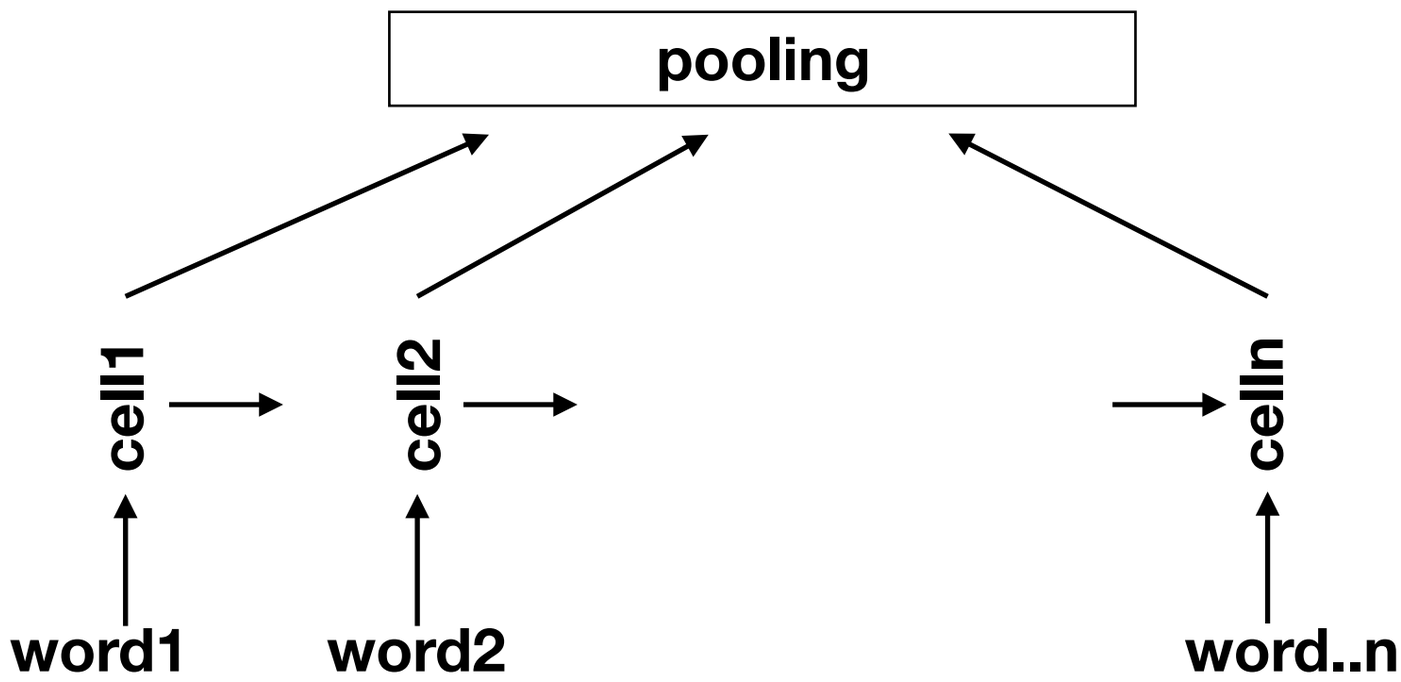
per class label

3 way softmax

sentiment analysis same problem

sequence models make sense?

treat words as a sequence . embeddings for short term dependencies



word=embedding

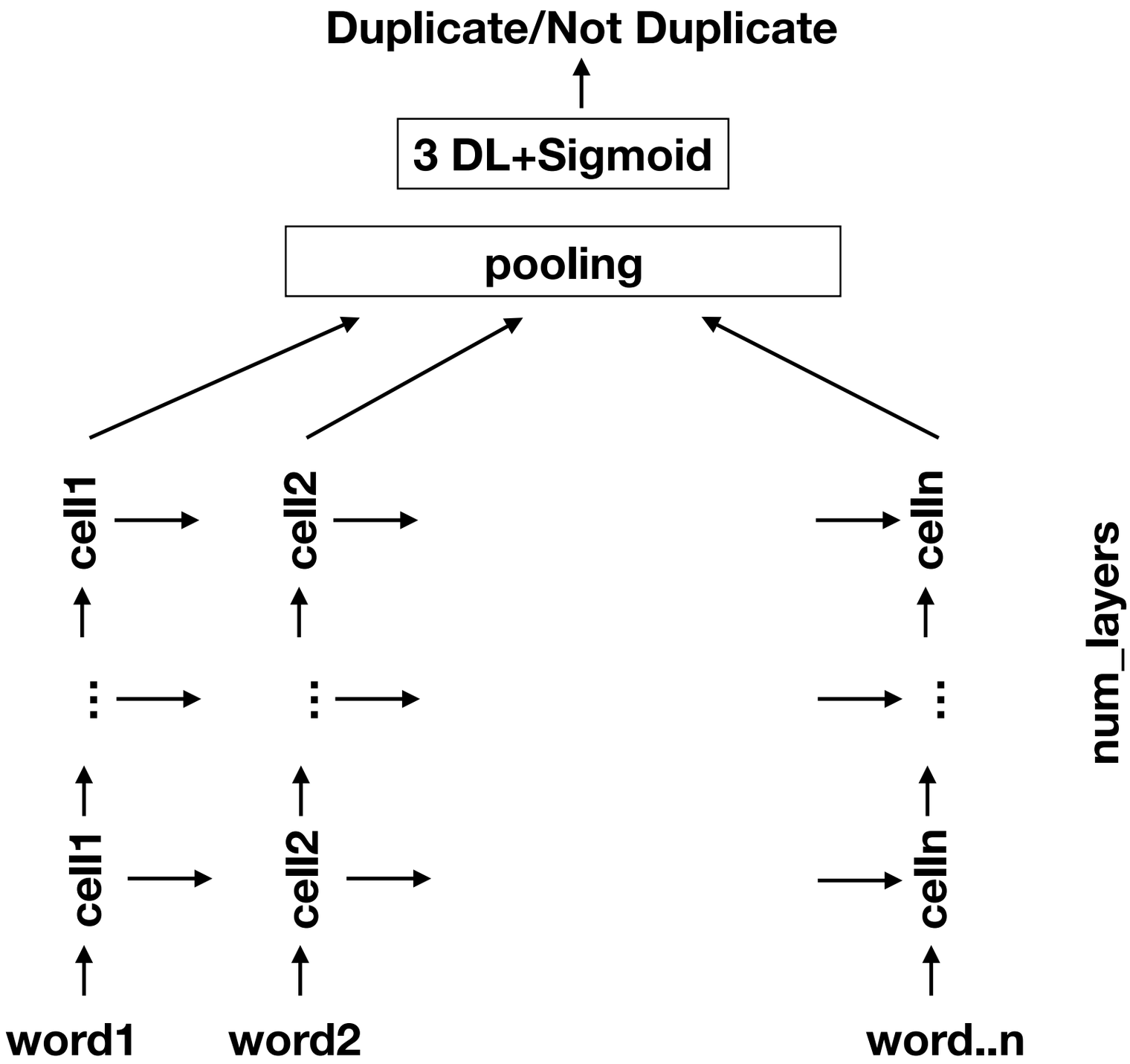
https://github.com/mchablani/deep-learning/blob/master/sentiment-rnn/Sentiment_RNN.ipynb

sentiment analysis w/RNN

only care about last stage output

$$L(y, \hat{y}) = |y - \hat{y}|^2$$

propagate loss through RNN


Duplicate/Not Duplicate
3 DL+Sigmoid
pooling
cell1
cell2
celln
cell1
cell2
celln
num_layers
word1
word2
word..n