Neural Networks and Natural Language

Vladimir Feinberg

July 10, 2017

We review techniques for natural language processing (NLP) with DNNs. Content is mostly from Dr. Goodfellow's Deep Learning Book, but also taken from Dr. Hinton's Coursera Class, lecture week 4.

1 Class Prediction

With many word outputs, softmax penalties make extremely sparse gradients if each word is a class. Resolve this by moving the output class into the input, and output a single scalar probability when parameterized. This is the serial architecture (Fig. 1). The serial architecture takes a long time to find candidates which are
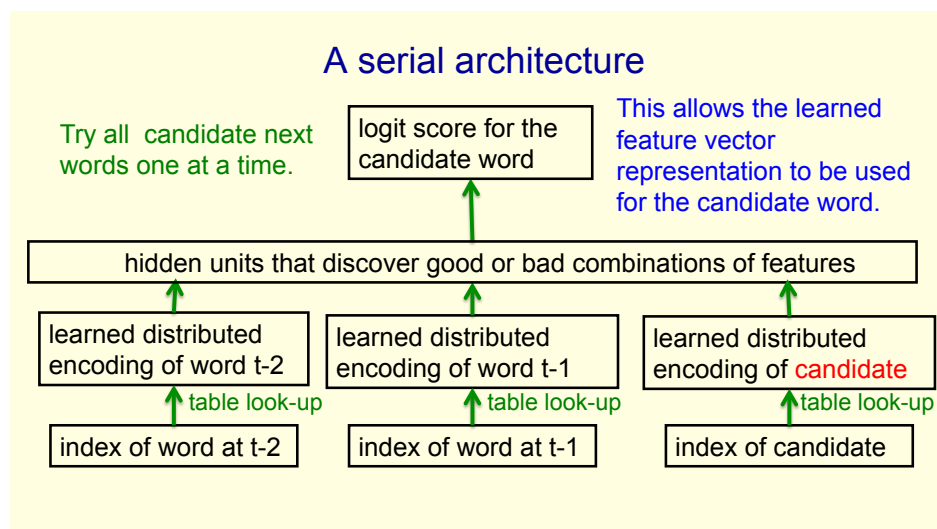

Figure 1: The serial architecture, which folds output complexity into input complexity, from Hinton's Coursera course lecture slides, week 4, slide 25.

assessed by the model as likely, and it can be improved to consider fewer candidates (Mnih and Hinton 2009).