Review Deep Averaging Networks

The goal of a sentiment analysis task is to tell whether the sentiment behind a text is positive or negative. In this case, what you are trying to learn is a classification model that classifies whether a sequence of words is positive or negative.

One way to approach this task is to use a deep averaging network. A deep averaging network consists of two components: a word embedding and a traditional neural network (sometimes even a linear classifier, which is a neural network without hidden layers).

A deep averaging network makes predictions, following these four steps:

1. Input a sequence of words $s_1, \ldots, s_n$
2. Convert these words to word embeddings $\mathbf{w}_1, \ldots, \mathbf{w}_n$ based on the network's internal word embedding
3. Average all the word embeddings to get a vector representation of the sequence $\mathbf{v} = \frac{1}{n}(\mathbf{w}_1 + \cdots + \mathbf{w}_n)$
4. Pass the vector $\mathbf{v}$ to the neural network to determine whether the sequence is positive or negative

A deep averaging network (DAN) can be trained using SGD. Training a DAN is almost identical to training a standard neural network. We use a standard cross entropy loss to predict if a sentence is, for example, positive or negative. When we do back-propagation, we update the weights of the neural network. However, we do one additional step: We also take a gradient update with respect to the word embeddings. This gradient can easily be computed with the chain rule using the back-propagation algorithm.

Deep averaging networks are one of the simplest neural sequence models. In this case, the averaging operation is the encoder (it produces a vector representation of the sequence) and the neural network is the decoder (the output is a simple scalar).

☆ **Key Points**

A deep averaging network consists of two components: a traditional neural network and a word embedding.

Use stochastic gradient descent (SGD) to train deep averaging networks.