4. Word Embeddings

Homework due Mar 29, 2023 08:59 -03 Completed

Word Embeddings

0/1 point (graded)

Training a neural network using back-propagation and SGD moves the network weights minimizes the loss function. If the network contains a *bottleneck*, a layer in which many only a few outputs, training will adjust the weights to maximize the useful information c output. In this way, a sparse input representation can be embedded in a lower-dimensi dense, *distributed* representation. Embeddings often have interesting properties like tra visual similarity into geometric proximity.

For example, imagine you have the words "cat", "lion", "car", "bridge". You could have a representation like: "cat" : [0.9, 0.2], "lion" : [0.9, 0.5], "car" : [0.01, 0.8], "bridge" : [0.01, representation, the first component give (not exactly) a measure of "animalness" and th addition, the vectors for similar or related words may be close together in space. In this examine the utility of (the highly popular) word embeddings.

Consider two neural networks for classifying sequences of words that differ only in the

- The first of which uses a sparse *one-hot* encoding of each word in which word i is that contains a 1 in position i and 0s elsewhere. For instance, a dictionary containin *embedding* might be represented as $[0\ 1]$ and $[1\ 0]$, respectively. You may assume contains all words in both the training and testing sets.

- The second neural network, instead, uses a pre-trained embedding of the dictionar represents every word in the dictionary.

Assuming that both networks use $\tanh$ activations and have randomly initialized weigh using the same training set, but different input representation.

Now, at test time, each network is presented with a sequence of words not seen during following statement(s) is/are true about the output of the network for this sequence?

- [ ] The first network cannot classify the sequence correctly
- [ ] The second network cannot classify the sequence correctly
- [x] The second network has a fighting chance at classifying the sequence
- [ ] The first network has a fighting chance at classifying the sequence

Previous Next