Philippe Wylezek-Serrano

CSC 487

Lab 2 Analysis

Multiclass Problem Analysis

The data is a set of newswires and their topics, published to the news company Reuters, the year of 1986. In total there are 46 different topics, each with 10 or more instances within the set. Our goal is to take a datapoint and be able to classify it as one of the 46 different topics, hence the multiclass part of this problem.

Data Analysis

To keep our dataset at a manageable level, we restrict the number of words that are present in each sample to 10,000, specifying that as we import. This prevents low frequency words from interfering with our model and its attempts to classify the articles.

Each sample is representative of the original articles, except instead of words, there are integer indices that reference that word in an external dictionary. For example, the sentence "The dog is lazy" would be transformed into [4, 17, 11, 29] (Assuming the external dictionary is as follows 4=the, 17=dog, etc..)

The samples, however, cannot be fed in their current forms. They must be transformed into a more suitable representation for our model (i.e. a tensor). To do so, we perform multi-hot encoding, a stripping the original data of both its frequency and order. Every sample becomes a 1D tensor of size 10,000, with a 1 occurring when a word appears in the original text, and 0's elsewhere.

For a similar reason and in a similar fashion, the labels of our data must also be vectorized. We do so through one-hot encoding, which, in a similar manner, transforms each label into a 1D tensor of size 46, with a 1 at the index matching the proper label, and 0 everywhere else.

Model/ Network Architecture Analysis

The model consists of a Sequence of 3 layers. To preserve the information across each layer, the number of nodes in each layer will not reduce until we reach the output layer. The output layer will naturally have a size of 46, to match the number of classes that exist in the data.

Considering that we are facing a classification problem, it is appropriate to utilize SoftMax as the activation function for the output layer, to produce a probability distribution across the 46 possible classes. Thus a prediction will be a 1D tensor with the shape (46), where each index 'i' contains a probability from the model that the input article is of class 'i'.

For the same reason, categorical cross entropy is an appropriate loss function, since said function measures the distance between the two probability distributions in their entirety.