Neural Networks Project

BREAKING THE SOFTMAX BOTTLENECK: A HIGH-RANK RNN LANGUAGE MODEL

About

The paper shows that learning a Softmax-based RNN network can be reduced to matrix factorization problem and that the matrix can be high-rank due to the fact that natural language is highly context-dependent.

The name of the proposed model is Mixture of Softmaxes. The main idea behind it is extending softmax expressiveness without loosing generalization. This is done by adding some parameters to Softmax layer.

Resources and plan

The proposed model was tested on two language modeling datasets: Peen Treebank and WikiText-2 and on dialog dataset: Switchboard. Experiments show that the model gives better results than RNN with basic Softmax in all datasets.

I'll try to reproduce and test this model using Pytorch and numpy. To do this I'll download and preprocess required datasets. They are all available for free on the web. Next I'll implement the MoS layer. The paper states that the formula is easy to implement and it also provides all hyperparameters.

The MoS model increases number of parameters and computations but paper's authors didn't check its performance, so I'll try to make some timing efficiency tests to compare this to basic softmax layer model.h

Marcin Gruza