

Implement a network for window classification which will learn Named Entity classes such as Person, Location, Property, Facility, Organization. You will implement the word embedding layer, the feed forward neural network and corresponding backpropagation algorithm. Each window's vectors are given as input to a hidden layer which has dimensionality H . This layer is used as features for a classifier which will return the probabilities that the center word belongs to the above mentioned classes. You should implement backpropagation and gradient check to make sure your neural network code is bug free. You should learn parameters with Stochastic Gradient Decent (SGD) for a mini batch at a time. You should visualize model variants and tune various model parameters.

You should vary the following:

1. Activation functions
2. Single layer vs multiple layers
3. Initialization of model parameters (W) (random vs uniform random number between $[-r, r]$ where r is inversely proportional to fan_in and fan_out)
4. Cost function: maximize log likelihood vs max margin
5. Un-regularised vs regularised
6. Single training item vs mini-batch
7. classifier

This assignment will help you to learn how to implement a network and what is the role of each item listed above.

Dataset:

Select a suitable dataset e.g. Reuters Corpus of news articles from the text corpora of nltk (details are given <https://www.nltk.org/book/>). If the dataset does not contain all the classes, then you can mix two or more datasets.

Network Analysis:

It is possible to get higher training accuracies by trying out different values of:

1. the regularization constant λ ,
2. the learning rate α (good try is 0.001)
3. the window size C ,
4. the hidden layer size H ,
5. the number of iterations or epochs,
6. training the word vectors or keeping them fixed
7. the size of the word vectors

Innovation:

Bringing out simple and novel mechanism to improve the performance of the network will fetch extra credit up to 20%.

Deliverables:

Running Code with readme file and input sample file

Prepare a report which should contain the following:

1. Short note on data preparation
2. Ground truth
3. Network details
4. Results
 - a. Gradient check results
 - b. Precision, recall and F1 measures with variations in hyper-parameters
 - c. Choose the prominent or interesting variations for showing the performance of the network (as suggested in (a)) with varying parameters. Showing only basic results may not attract credits.

Note: use of libraries is limited only to download databases.