1. Preprocessed the sentiment analysis dataset <https://www.kaggle.com/snap/amazon-fine-food-reviews>.
2. the Glove word vectors

<http://nlp.stanford.edu/data/glove.6B.zip>

extracted the 100 dimensional file (glove.6B.100d.txt) from the zipped folder.

1. Preprocessed the review dataset by considering the column “review score” >3 as positive reviews and other negative review. For training one local machine consider 5000 positive and negative reviews each for training dataset.

2000 reviews for test dataset.

1. Truncate 200 most common and least common words from reviews on training and test dataset.
2. Represent each word to corresponding embedding from Glove 100 dimensional vector. Use non-trainable embedding layer in keras or tensorflow to represent the same.

Note: Unknown word which are not present in glove.6B.100d.txt, replace with random 100 dimensional vector ranging between (-.0.5 to +0.5).

1. Train a Convolutional neural network and a fully connected layer at the top, to classify the reviews. Now run the network by changing the following hyper-parameters:

|  |  |  |  |
| --- | --- | --- | --- |
| Hidden Layers | Convolution Window | Convolution size | Regularization |
| 1 | (5\*5, 4\*4, 3\*3) | [ 16,32, 64] | Dropout of 0.8 after each layer |
| 2 | (5\*5, 4\*4, 3\*3) | [ 16,32, 64] | Batch normalization after each layer (except the first) |

1. Hand written review fot testing.