AURIAU Vincent
MVA_MP2

Deep Learning for Natural Language Processing

Project

1. Monolingual embeddings
2. Multilingual word embeddings

Question:

Let's find $W_* = argmin(||WX - Y||)\ with\ W^T W = I$.

Let's call $W_* = argmin\left(||WX - Y||^2\right)$

$$W_* = argmin(< WX - Y, WX - Y >) = argmin(||X||^2 + ||Y||^2 - 2 < WX, Y >)$$

We can transform the problem into:

$$W_* = argmax(< W, YX^T >) = argmax(< W, U\Sigma V^{\mathrm{T}} >) = argmax(< U^T WV, \Sigma >)$$

With $U\Sigma V^{\mathrm{T}}$ the SVD decomposition of YX^T.

$$W_* = argmax(S, \Sigma)\ where\ S = U^T WV$$

S is an orthonormal matrix and thus is maximised when S equals the identity matrix (equality case in the Cauchy Schwartz formula). Thus,

$$I = U^T RV$$

$$W_* = UV^T$$

3. Sentence classification with BoV

Question:

Using the mean and idf-weighted average, we got the following results:

-Accuracy on the training set for the mean model: 0.4331

-Accuracy on the dev set fot the mean model: 0.3705

Accuracy on the training set for the idf-weighted average model: 0.4415

-Accuracy on the dev set fot the idf-weighted average model: 0.3551

We tuned the parameters of the logistic regression using the results on the dev set. We can see that the results are better using the mean model (the results on the dev test are more important, because they are closer to what we will get with new data). It's the one we used for the "logreg_bov_y_test_sst.txt" file.

AURIAU Vincent
MVA_MP2

4. Deep Learning models for classification

Question:

I used the categorical cross entropy loss for the 5-class classification, because it's a multiclass classification problem.

$$CE = -\sum_{i=1}^{2} t_i log(s_i)$$

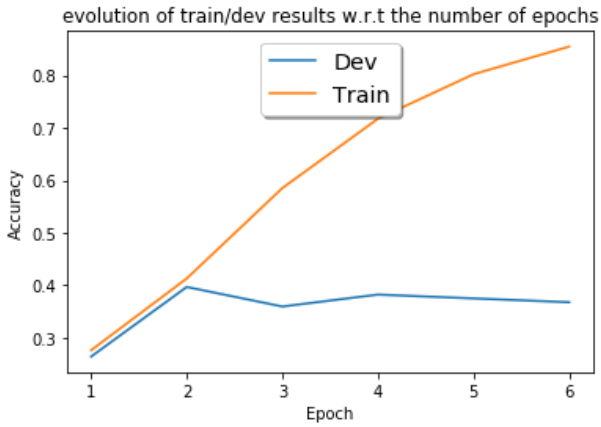
With:

-ti equals 1 if the element is found to be in the ith class

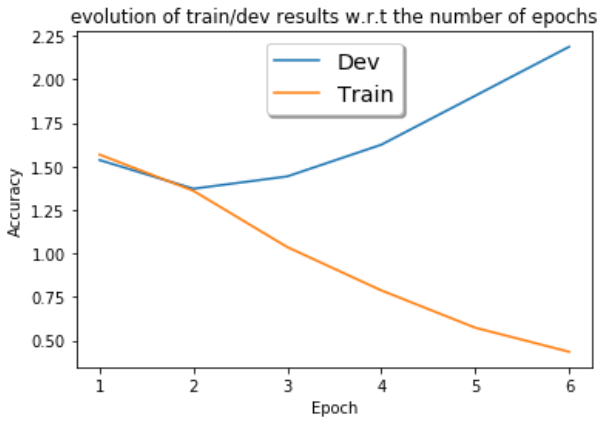
- si is the output of the considered final neurons (the ith neuron on the last layer), output just after the softmax activation here.

In the example of a multiclass classification problem, one $t_i$will be equal to 1 and all the other to 0, leaving only one term.

Question:



Evolution of accuracy



Evolution of loss

We can see the chart of the evolution of the accuracy and the loss of the algorithm for the training and development set w.r.t. the number of epochs.

We can see that from the third epoch, the algorithm starts to overfit: it gets more accurate on the training set and less accurate on the dev test. It's why to compute the prediction, we only kept two epochs.

Question:

We used here a Conv1D network and with an idf-weighted mean from the BoV model. We can work on the features of the algorithm to try not to overfit while getting the best accuracy.

It can change, but the results of the convent are more variable than with the lstm and overfit less.