NLP project

Raphael Achddou

January 2019

1 Monolingual embeddings

2 Multilingual word embeddings

Question 1 - We want to solve the following problem :

$$argmin_{\mathcal{O}_d(\mathbf{R})}(||WX - Y||_F).$$

For this, we use the properties of the Froebenius scalar product and properties of orthogonality.

$$\begin{aligned} ||WX - Y||_F &= tr((WX - Y)^T(WX - Y)), \\ &= tr(X^T W^T WX + Y^T Y - X^T W^T Y - Y^T WX), \\ &= tr(X^T X) + tr(Y^T Y) - tr(X^T W^T Y) - tr(Y^T WX), \\ &= ||X||_F + ||Y||_F - 2 \times tr(X^T W^T Y). \end{aligned} \tag{1}$$

Therefore the argmin becomes :

$$\begin{aligned} argmin(||WX - Y||_F) &= argmax(tr(X^T W^T Y)), \\ &= argmax(< WX, Y >), \\ &= argmax(< W, YX^T >), \\ &= argmax(< W, U\Sigma V^T >), \\ &= argmax(< U^T WV, \Sigma >). \end{aligned} \tag{2}$$

$S = U^T WV$ is an orthogonal matrix as it is a product of orthogonal matrices, therefore the scalar product is maximal when $S = I$,

$$S = I \iff W = UV^T.$$

3 Sentence classification with BoV

Question 1 - After fine-tuning the coefficient of L2-regularization, we obtained the best result for C = 0.4 for the logistic regression without weighting and C = 0.75 for the weighted average.
In the first case, we got a training error of 0.5423 and a validation error of 0.5813. In the second case, we got a training error of 0.5559 and a validation error of 0.6122. We notice that, surprisingly, the idf encoding leads to a larger error on the validation process.

4 Deep Learning models for classification

Question 1 - For this problem we chose to optimize the categorical cross-entropy loss which is very efficient for classification problems. Its expression is the following :

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=0}^{4} P(y_i \in C_c) \times log(P(y_i \in C_c))$$

, where y_i is the possible output of the prediction. Indeed, the network computes a probability vector for each class.

Question 2

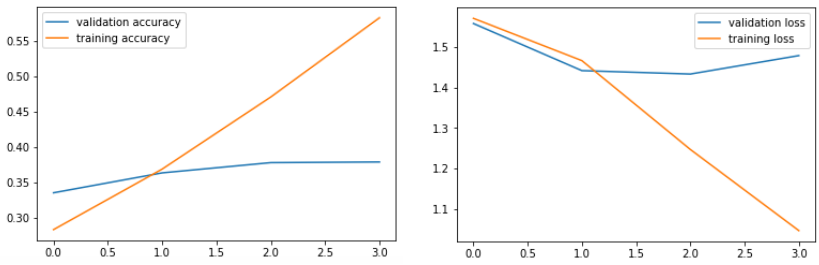

Figure 1: Plots of the accuracy and the loss during the training of the LSTM network