# TP 2: Deep learning for NLP

## Multilingual word embeddings:

We have:

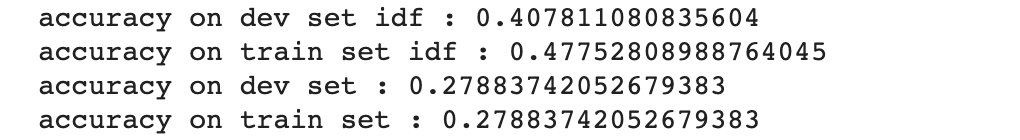
Using the singular value decomposition and the proprieties of the scalar product we have:

Since are orthogonal matrices, the maximum is reached when which gives .

## Sentences classification with BOW:

We have trained a logistic regression over bag-of-words embedding using two embedding strategies: using a normal average or a weighted average.

We have the following results for the accuracy:



## Deep learning model for classification:

Question: For this classification task, we used the categorical cross-entropy. For Five classes this loss is defined by the following expression:

This loss can be view as the distance/divergence between two distribution. Its definition comes from the Kullback-Leibler divergence.

We first tried to plot the accuracy and the loss for the train and dev sets over 10 epochs.

To avoid overfitting, we used dropouts. Given the plot of the loss on the dev set we considered stopping the training earlier and used instead 5 epochs. We obtain the following plots:

Une image contenant carte, texte

Description générée automatiquementUne image contenant carte, texte

Description générée automatiquement