CentraleSupélec

Deep Learning

Assignment 2

François Meunier

1 Multilingual word embeddings

Let's prove that $\underset{W}{\arg\min}\|WX - Y\|_F = UV^T$

$$\begin{aligned}
\|WX - Y\|_F^2 &= \|WX\|_F^2 - 2\langle WX, Y\rangle_F + \|Y\|_F^2 \\
&= \mathrm{Tr}(X^T W^T W X) - 2\langle WX, Y\rangle_F + \|Y\|_F^2 \\
&= \mathrm{Tr}(X^T X) - 2\langle WX, Y\rangle_F + \|Y\|_F^2 \\
&= \|X\|_F^2 + \|Y\|_F^2 - 2\langle WX, Y\rangle_F
\end{aligned}$$

Therefore :

$$\underset{W}{\arg\min}\|WX - Y\|_F = \underset{W}{\arg\max}\,\mathrm{Tr}(WXY^T)$$

Let's denote the SVD of W as $U_W \Sigma_W V_W^T$. W,

$$\begin{aligned}
W^T W = I_d &\Rightarrow (U_W \Sigma_W V_W^T)^T (U_W \Sigma_W V_W^T) = I_d \\
&\Rightarrow V_W \Sigma_W^2 V_W^T = I_d \\
&\Rightarrow V_W \Sigma_W^2 = I_d V_W \\
&\Rightarrow V_W (\Sigma_W^2 - I_d) = 0 \\
&\Rightarrow \Sigma_W = I_d
\end{aligned}$$

Finaly we have,

$$\begin{aligned}
\mathrm{Tr}(WXY^T) &= \mathrm{Tr}(U_W V_W^T V \Sigma_W U^T) \\
&= \mathrm{Tr}((U^T U_W)(V V_W^T)\Sigma) \\
&\leq \mathrm{Tr}(\Sigma)
\end{aligned}$$

This last inequality is an equality if $U = U_W$ and $V = V_W$, which proves the final result :

$$UV^T = \underset{W}{\arg\min}\|WX - Y\|_F$$

2 Sentence classifiction in BoV

With the weighted average we obtain a 41% on the train set and 32% accurracy respectively on train and dev set.

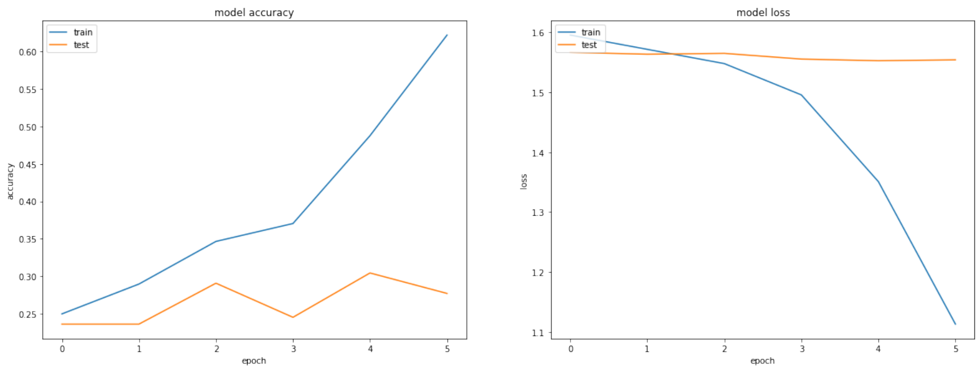
The unweighted version gives a 45% accuracy on the train set and 35% on the dev set.

3 Deep Learning models for classification

Categorical crossentropy loss : let $\{x_1, \ldots, x_n\}$ be the sentences, $\mathbf{y} = \{y_1, \ldots, y_n\}$ the real classes, and $\hat{\mathbf{y}} = \{\hat{y}_1, \ldots, \hat{y}_n\}$ prediction. The categorical cross entropy is :

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{n} \hat{y}_i \log y_i$$

The LSTM overfits the training data really fast. Best reached accuracy on the validation set is 30% after 4 epochs.



The new model uses already trained words encoding, however performance is similar.