2. Multilingual word embeddings

$$\min \|WX - Y\|_F^2 = \min Tr[(WX-Y)^T(WX-Y)]$$

$$= \min Tr(X^TX)$$

$$= \min [Tr(X^TX) + Tr(Y^TY) - 2Tr(X^TW^TY)]$$

$$= \min [Tr(X^TX) + Tr(Y^TY) - 2Tr(W^TYW^T)]$$

$$= \min [Tr(X^TX) + (Y^TY) - 2Tr(\Sigma U^TWV)]$$

so we need to minimize $-2Tr(\Sigma U^TWV)$

For a positive diagonal matrix Σ,

we always know $Tr(\Sigma) \geq Tr(\Sigma O)$ for $O \in O_d(\mathbb{R})$

Only when $O = Id$ we can get "$=$".

Such being the case, $Tr(\Sigma U^TWV) \leq Tr(\Sigma)$

To minimize $-2Tr(\Sigma U^TWV)$

we can set $W = UV^T$

3. Sentence classification with BoV

the weighted model train mean_squared_error is 1.16783707865

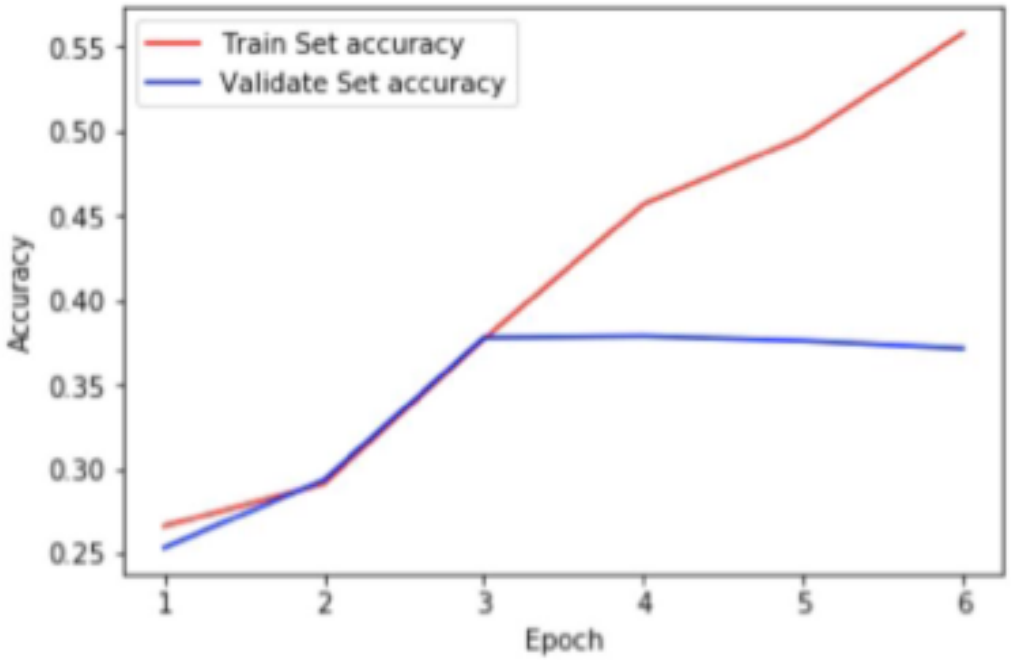
the weighted model dev mean_squared_error is 1.29427792916

4. Deep learning models for classification

Loss function: categorical_crossentropy

$$L = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y_i}) + (1-y_i)\log(1-\hat{y_i})\right]$$

where y_i is the ground truth and $\hat{y_i}$ is the predicted pro for the truth



creative:

Instead of word vectors, I use the sentence vectors and implement it in LSTM model. Compared with traditional embedding method, LSTM or RNN can extract more information. The validate accuracy is higher than previous method.