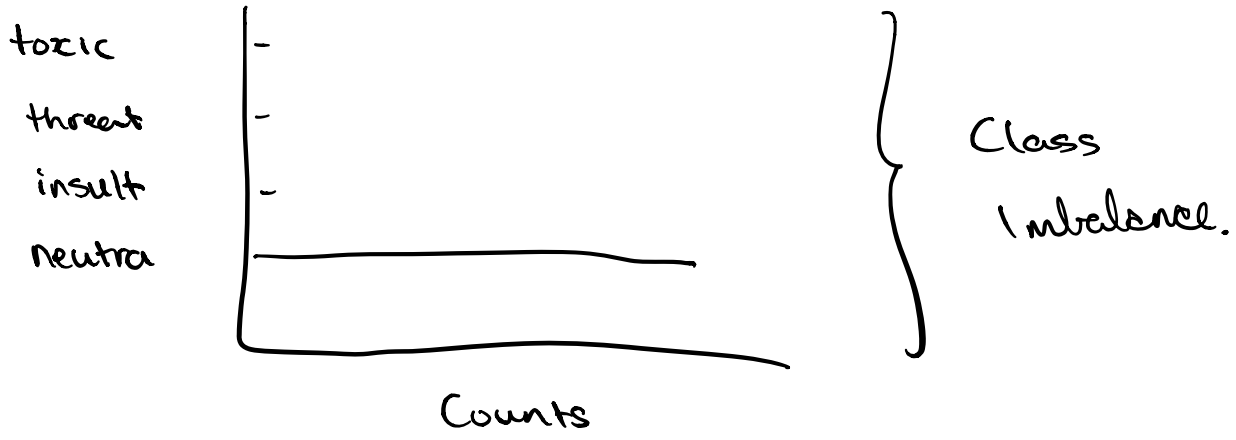
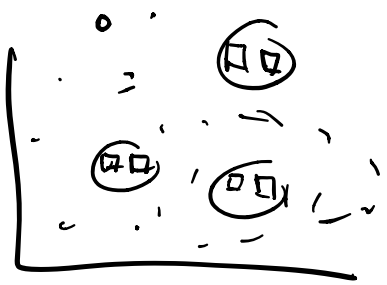
IMRSV – does NLP and Speech analysis

How to classify Sentences



Problems:

Overlapping Classes



Small Disjuncts

True Positive rate = TP/(TP+FN)

- ROC-AUC.

Solutions

- Data vs Model

— DATA —

- augmentation using translations (en → fr → en)
- random understampling (discard overpop
- random oversampling (magnify small sample)

- MODEL -

Embeddings

↓

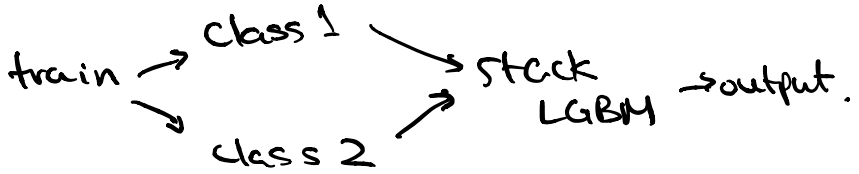
Bi-LSTM

↓

Bi-LSTM

↓

Softmax



How to find data we Out of Fold (OOF) predictions.