

Assignment 3

L665/B659: Applying Machine Learning Techniques in CL; Sandra Kuebler

- For this assignment, you will use the English lexical sample data from the SensEval-3 competition. This data set is attached to the assignment and will also be used for the final project.
- For the current assignment, we will concentrate on 3 words out of the set: `arm.n`, `difficulty.n`, and `interest.n`.
- Extract all examples for those words from the data files, and from those examples, extract the following features: For each focus word w : you extract a vector with the following information:

w_{-2} w_{-1} w_{+1} w_{+2} $w_{-2}\%w_{-1}$ $w_{-1}\%w_{+1}$ $w_{+1}\%w_{+2}$ class

- Train 3 Timbl classifiers on the training set for one word, and test it on the word's test set.
- Optimize the classifiers by adjusting the parameters.
- Report the baseline (default setting) and the best results including the parameter settings used to generate these results. Did you need different settings for the three words?
- Work in groups of 2-3 people.
- **DUE DATE:** February 22 before class. Please submit via canvas.