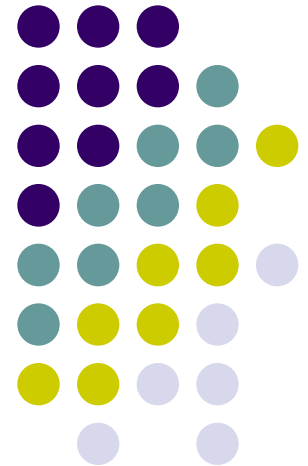
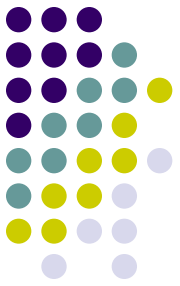


Polimetrics

Eight Assignment



Deadline: 26 November 2019

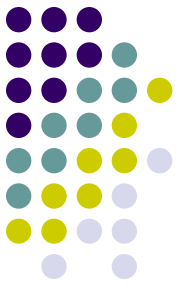


- Use once again the train and test set files including movie reviews
- Use as your training-set the reviews included in `train_review.csv`

PART 1:

- Run a 5-fold cross-validation using 3 different ML algorithms

Deadline: 26 November 2019



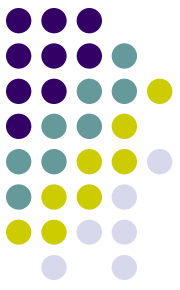
- HINT (part 1): keep in the analysis only those features that appear in at least 5% of the texts
- Otherwise your dfm will be too large, and when you run a ML algorithm you could get this error: «Error: protect(): protection stack overflow»
- It means that given that you use a VERY large number of variables in the dataframe, R refuses to go on with the analysis given the power of your laptop

Deadline: 26 November 2019



- HINT (part 2): when your output variable presents just two classes (Positive/Negative), the results you get via the *confusionMatrix* command are a bit more tricky
- In particular, you are required to estimate the value of F1 for both the Positive and the Negative class by hand (according to the formula we know pretty well!)

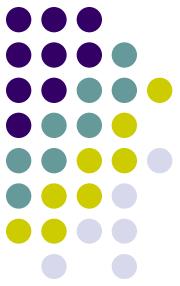
Deadline: 26 November 2019



- HINT (part 2): therefore, in the second part of the loop for CV, use the following command (below, as an example, I focus on the SVM estimation):

```
SVMPredict <- data.frame(col1=vector(), col2=vector(), col3=vector())
for(i in mget(ls(pattern = "conf.mat.sv"))) {
  col1 <- (i)$overall[1] # save in the matrix the accuracy value
  col2 <- (2*(i)$byClass[1]*(i)$byClass[3])/((i)$byClass[1]+(i)$byClass[3])
  # estimate and save in the matrix the F1 value for the negative class
  col3 <- (2*(i)$byClass[2]*(i)$byClass[4])/((i)$byClass[2]+(i)$byClass[4])
  # estimate and save in the matrix the F1 value for the positive class
  SVMPredict <- rbind(SVMPredict , cbind(col1, col2, col3))
}
```

Deadline: 26 November 2019



PART 2 (optional!):

- Then re-run 5-fold cross-validation using the same 3 different ML algorithms as well as iSAX. In this second case, compare the finale results in terms of the respective MAE