 Ubungsblatt 3

Aufgabe 3.1

What filter techniques from the feature selection framework can be applied for a text

classification task? What Filters are not, or at least less, suitable for text? Justify your

answer under a careful reconsideration of the nature of the features in use! For simplicity

just assume a Bag-Of-Words representation of the text.

For Text classification the input and output variables are of categorical type specifically here more of nominal. As per the lecture slides Chi-Square (more dominate is the impact of a feature out of bag of words more strong is the X in X Square) and Mutual information (Information gain is actually very powerful can be used for both numerical and categorical data) are suitable for this nature of features (features in bag of words are nominal and assuming the target classes are also nominal ).

There are other methods suitable for more of numerical input and output types such as t-test, ANOVA etc. Which doesn’t suits text classification /bag of words (collection of words occurring in a document or text) approach since data types are not numeric. For example we can’t use the mean of the occurrence of the words to determine how a word is relevant in determining its class.

However if the target class is of numeric type which isn’t typically a case in bag of words methods like ANOVA etc can be used.

I found this paper- Isabelle Guyon and Andre Elisseeff “[An Introduction to Variable and Feature Selection](http://jmlr.csail.mit.edu/papers/volume3/guyon03a/guyon03a.pdf)” . It provide an interesting approach.

Aufgabe 3.2

Implement suitable Filters (or use methods from Python Sk-learn or similar tool-kits) for

a classification task of your own choice! Report performance measures (you can consider a benchmark dataset like 20 Newsgroups) of your classifiers in respect to the filtered features, e.g. on a regular grid (50, 100, 150, 200, 250, ...) of feature set sizes. Provide

plots where ever possible to explain your answer.

Apply chi-square on 20newsgroup using sk-learn

Plot the graph

Apply information gain

Plot the graph

Apply ANOVA

Plot the graph

Plot the comparison between chi-square and information gain and ANOVA

Aufgabe 3.3

Does the proposed feature scaling a\_ect every Machine-Learning model? If not, which

models are a\_ected and which models are not? Justify your answer in respect to the

parametrization of your chosen models! (Gaussian-Naive-Bayes, Logistic-Regression and

non-linear Support-Vector-Machine models might be good candidates for your reasoning.)

Aufgabe 3.4

Use an iteratively trained classi\_er on a classi\_cation task of your own choice (you can

consider a benchmark dataset like 20 Newsgroups and a tool-kit of your choice). Compare

its convergence when trained on the original feature and on scaled features. Provide plots

where ever possible to explain your answer. What do you observe and how can this

phenomenon be explained?