6.0 Binary calssification structure

[Bernhard83] published a method to increase the classificators and thus he obtained a better classification result. The classification result was better about the factor 100 when using linear basis classificators both with the rejection and with the error rate instead the classification with a linear least square method classifier for several classes. The presented classification problem is based on printed capital letters with a random sample of 26.000 samples and 26 different classes.

The structure is given in fig. 6.01. The names of the classes are showed and it is evident, the number of classifiers is calculated with k classes with the formular:

$$\text{number} = \frac{k(k-1)}{2}.$$

In our example we use the linear least square method classifier [Schürmann77] for 2 classes and calculate 15 classifiers.

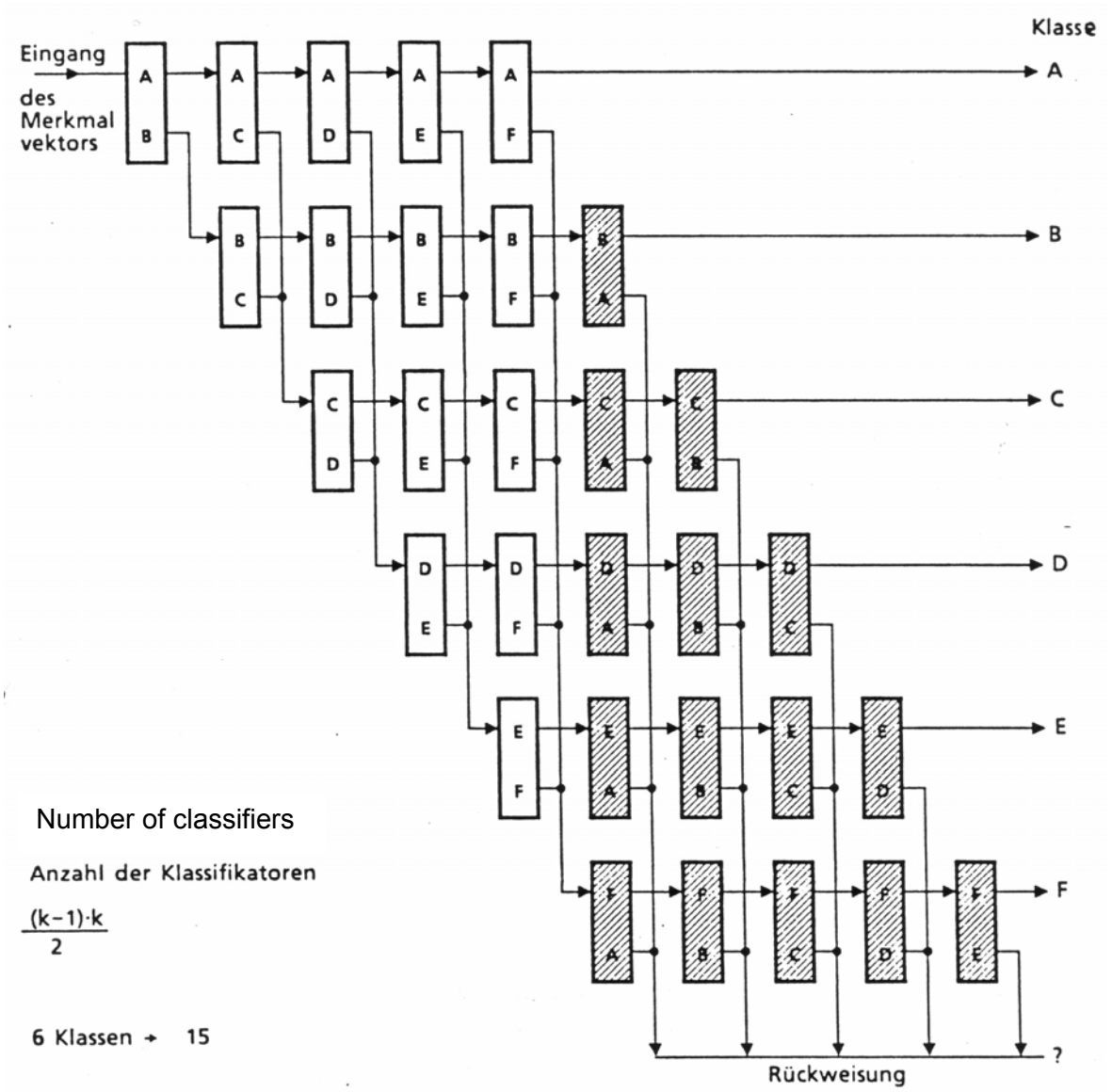


Figure 6.01: Binary classification structure for 6 classes A, B, C, D, E and F

The structure indicates, that the system gives a rejection by a contradictory decisions.

Therefore it is necessary to give the rejection rate δ the error rate ε and the recognition rate α. In principal:

$$\alpha + \varepsilon + \delta = 1$$

Plotting the the rejection and error rates over the error vector gives an exact statement about the efficiency of different procedures.
This procedure has following advantages

- The number of classes determines the minimum number of the samples in the learning set. That means, the representativeness of the learning set is easier to fulfill with 2 classes.
- The LMC classificiers determine the importance of the features individually adapted to the classes.
- The effort for classes, which are difficult to sepertate can be increased

The following comparison was accomplished with a language sample of **six** different classes and 20 features.

- linear least square method classifier
- quadratic least square method classifier
- binary classification structure

All three procedures are equivalent if only using the recognition rate. The efficiency of different classification procedures can be displayed more clearly calculating a quality measure with each decision and keeping records of the frequency statistics.

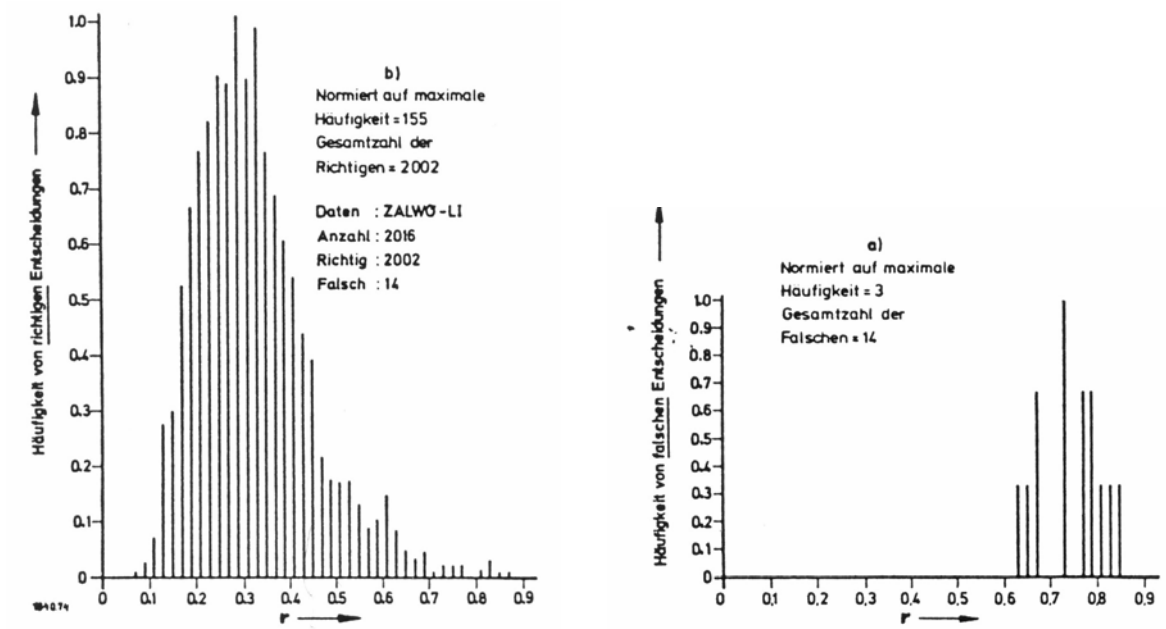


Figure 6.02: Frequency statistics of the error vector for the absolute value |r| separated according to correct and wrong decisions with the linear least square method classifier.

Figure 6.02 gives the results for the linear least square method classifier. The correct decisions on the left and the wrong decisions on the right. If a classifier improves the error vector |r| bzw. his distribution shifts toward zero.
Fig. 6.03 shows that for a quadratic least square method classifier the distribution of the correct decisions moves to the left.

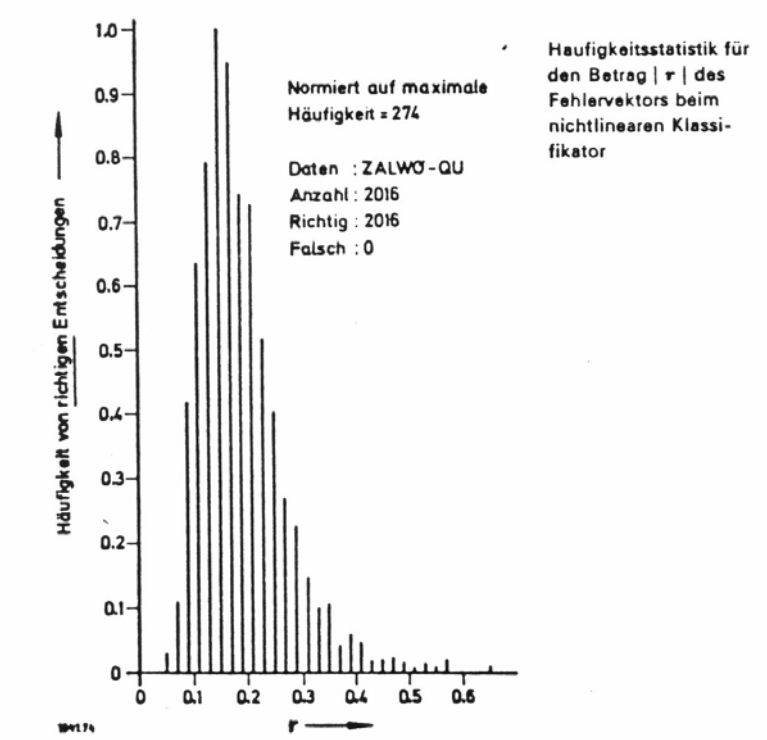
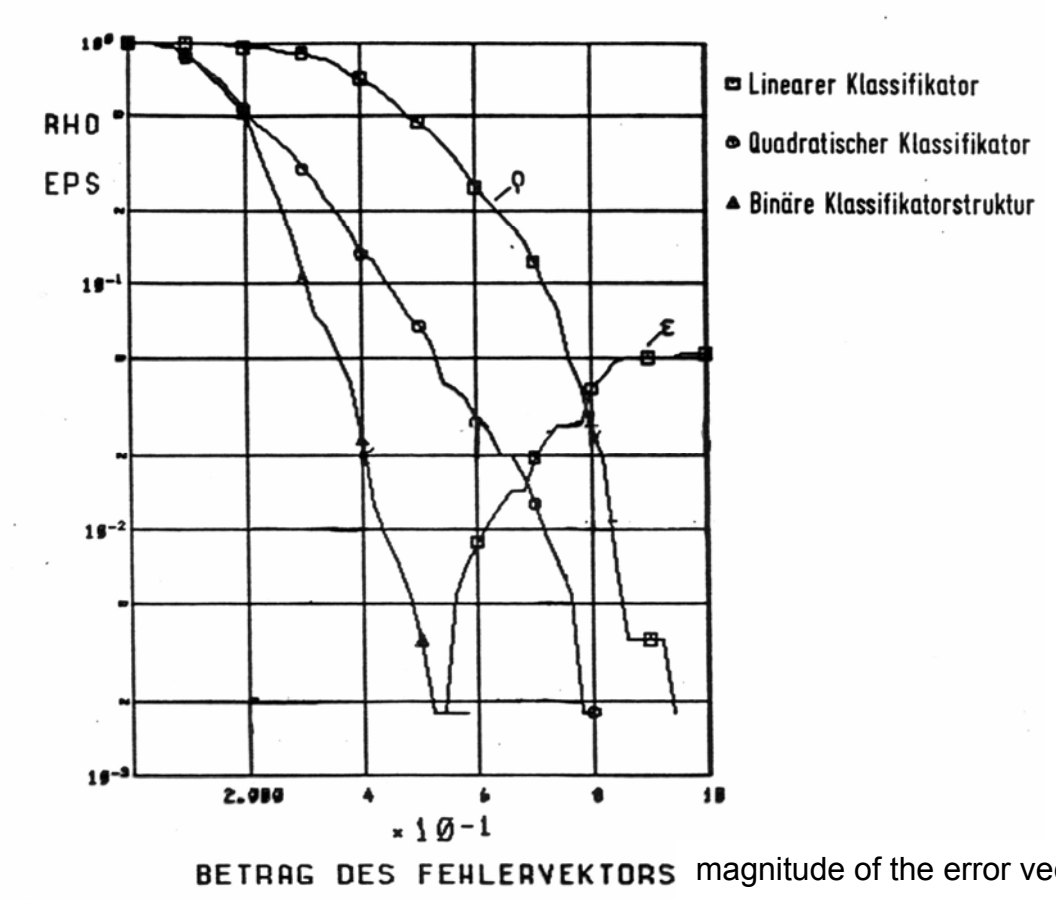


Figure 6.03: Frequency statistic of the error vector for the quadratic least square method classifier.

From the frequency distributions figure 6.02 and figure 6.03 can be won by summation over the appropriate subranges the functional connection between the rejection and the error rate.



magnitude of the error vector |r|

Figure 6.04: Rejection and error rate in dependency of the error vector for the three studied classification procedures.

Since the range of small errors and rejection rates are important, a logarithmic representation of the ordinate is appropriate. With this representation the classificfiers can be compared, even if all errors are zero.

Consequence: The adaptation ability to the learning set of the binary structure is phenomenal. During the investigations a representative test sample to make a final decision is missing.