Wrocław University of Technology

Randomness and Fuzziness in Bayes Multistage Classifier

Robert Burduk

robert.burduk@pwr.wroc.pl

Department of Systems and Computer Networks, Faculty of Electronics,

Contents

- Goal of paper
- The synthesis of multistage classifier
- Assumptions for analitical resalts
- Classification accuracy and example
- Conclusion

Goal of paper

- In the pattern recognition problem with fuzzy classes and fuzzy information we consider the following tree situations:
  - fuzzy classes and exact information,
  - exact classes and fuzzy information,
  - fuzzy classes and fuzzy information.

Goal of paper

- The paper deals with the probability of misclassification in a multistage classifier.
- The information on objects features is represented by fuzzy observations.

Goal of paper

- The paper deals with the probability of misclassification in a multistage classifier for exact classes and fuzzy information.
- The information on objects features is represented by fuzzy observations.

Goal of paper

- Our aim is to shown the difference between probability of misclassification for fuzzy and non-fuzzy features description.
- The obtained results are compared with the bound on the probability of error based on information energy of fuzzy events.

The synthesis of multistage classifier

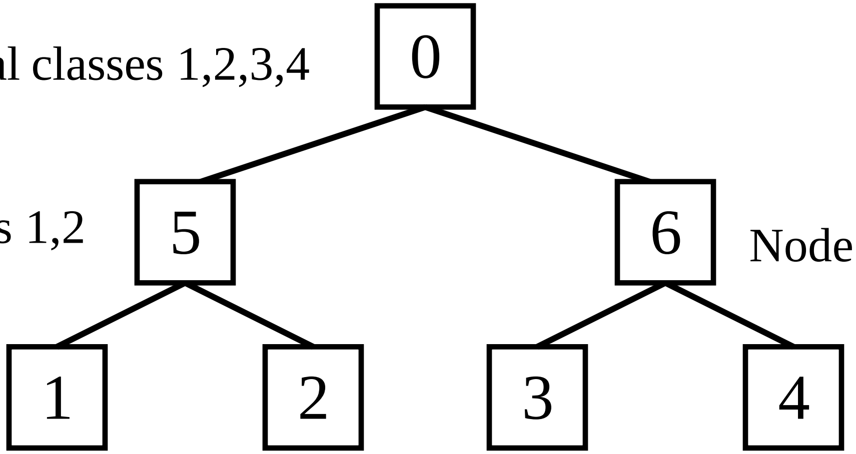
External nodes (label of classes)– 1, 2, 3, 4

Indernal nodes – 0, 5, 6

Node 0 – potential classes 1,2,3,4

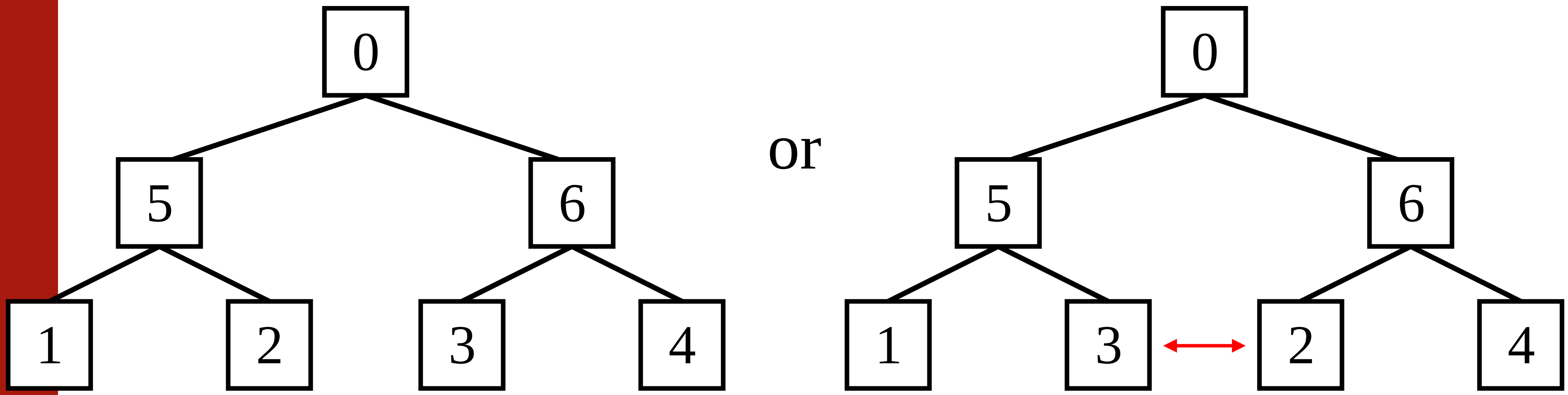
Node 5 – potential classes 1,2

Node 6 – potential classes 3, 4



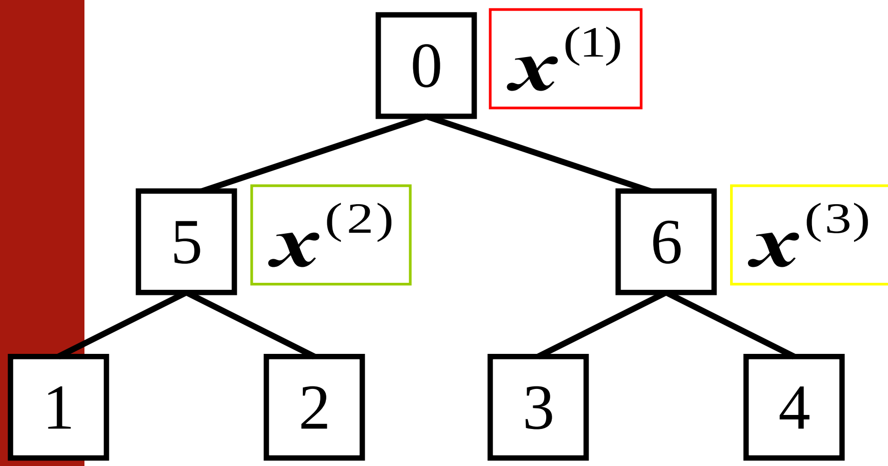
Thus multistage recognition means a successive narrowing of the set of potential classes from stage to stage, down to a single class.

The synthesis of multistage classifier

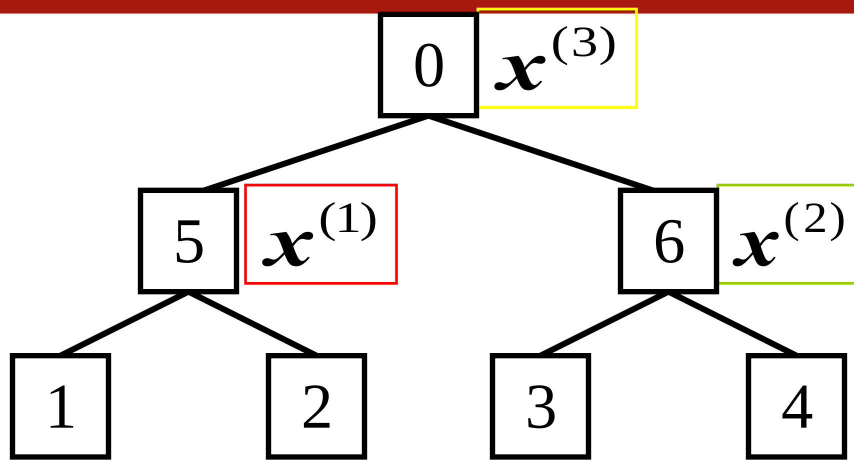


- the decision logic, i.e. assigning of class (or sets of classes) to nodes

The synthesis of multistage classifier

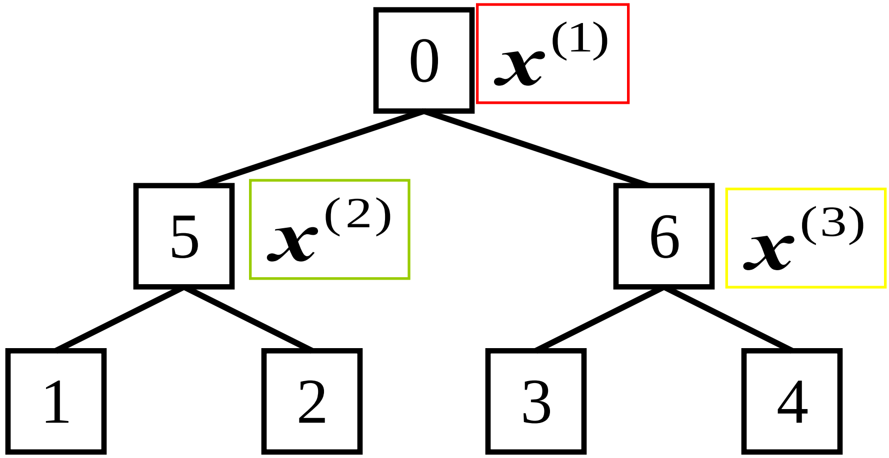


or



- feature used at each internal nodes of decision tree (feature selection)

The synthesis of multistage classifier



- the decision rules (algorithm) for performing the classification
(in this paper a Bayes rule in each nodes)

Assumption for analitical result

- The decision logic and features used at each internal nodes are known

- Our aim is to minimizes the mean risk function (the mean probability of misclassification of the whole multistage decision process – global optimal strategy of multistage classifier)

- The loss function is so-called „zero-one” loss function

Assumption for analitical result

- The probability of fuzzy events

$$P(\tilde{A}) = \int_{\text{supp}(\tilde{A})} \mu_{\tilde{A}_i}(x_i) f(x) dx$$

- The set of all available fuzzy observations satisfies the orthogonality constraint:

$$\sum_{l=1}^{n_k} \mu_{A_k^l}(x_k) = 1 \qquad \forall x \in X.$$

Classification accuracy

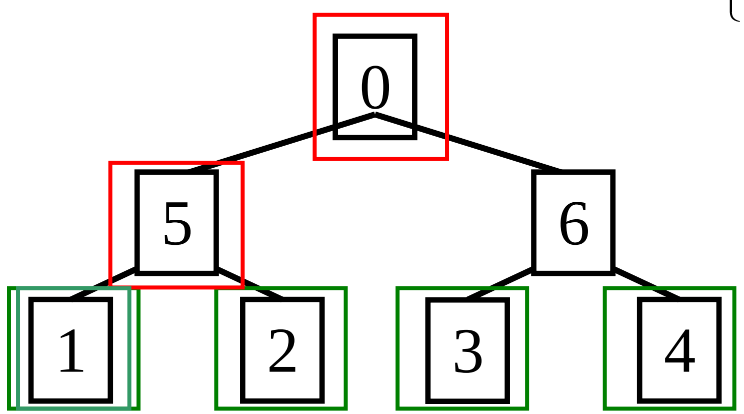
When we use fuzzy information on object features instead of exact information, we deteriorate the classification accuracy. The exact upper boundary of the difference between the probability of misclassification for the fuzzy and crisp data for the globally optimal strategy of multistage recognition is the following:

Exac bound of error for fuzzy data

$$Pe_F(\pi_N^*) - Pe(\pi_N^*) \le \sum_{j_N \in M(N)} p(j_N) \sum_{i_k \in s(j_N)-\{0\}} \varepsilon_{m_{i_k}}$$

where

$$\varepsilon_i = \sum_{A_i \in X_i} \left| \int_{\Re^i} \mu_{A_i}(x_i) \max_{k \in M^i} \{ f_k(x_i) \} dx_i - \max_{k \in M^i} \left\{ \int_{\Re^i} \mu_{A_i}(x_i) f_k(x_i) dx_i \right\} \right|.$$



Bound of error in therms of information energy

$$Pe^{IE}{}_{F}(\pi_N^*) - Pe(\pi_N^*) \leq \sum_{j_N \in M(N)} p(j_N) \sum_{i_k \in s(j_N) - \{0\}_{m_{i_k}}} \varepsilon_{m_{i_k}}$$

where

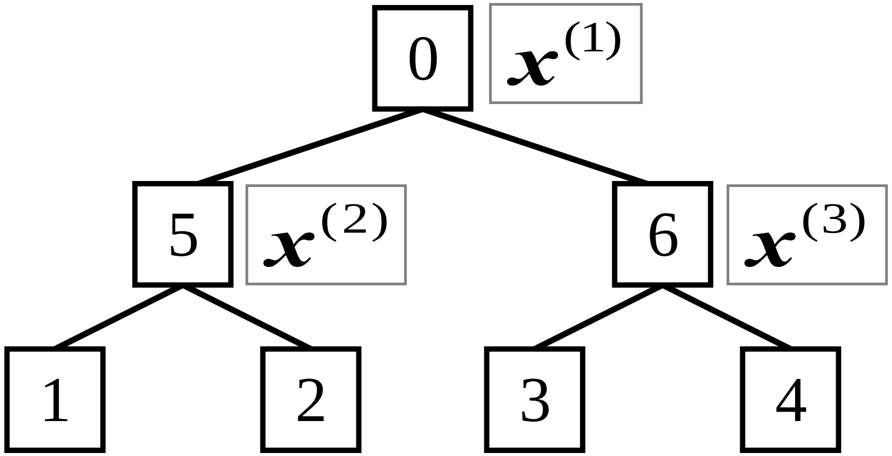
$$\varepsilon_i = 1 - E_{m_{i_k}}(A_{m_{i_k}}, M_{m_{i_k}})$$

- The conditional information energy for given fuzzy information

$$E_{m_{i_k}}(A_{m_{i_k}}, M_{m_{i_k}})$$

Wrocław University of Technology

Ilustrative example



- 3-dimensional feature $\boldsymbol{x} = [\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}]$
- Gaussian random variables with covariance matrices equal for every class

$$\sum_{j_2} = 2I, \quad j_2 \in \mathrm{M}(2),$$

Ilustrative example

- The exact difference between the probability of misclassification for fuzzy and non-fuzzy data in global optimal strategy and the same difference in terms of information energy

Shift of class-conditional probability density functions k=						
0	0,25	0.5	0.75	1	1.25	1.5
0.154	0.152	0.150	0.152	0.154	0.158	0.164
0.005	0.008	0.020	0.008	0.005	0.008	0.020

Canclusion

- The exact difeeerence the probability of error is always tighter than the differencebased on information energy

- The exact difference is periodical. In this case, the period is equal to 1.

- This period is the half of the width of the fuzzy number. This is the observation requires the analytic confirmation.

Thank you for Your attention