MA4022/MA7022 DATA MINING and NEURAL NETWORKS

Computational Task 2

Due date Monday 02.03.2015, 23:59

1. Balance scale classification.

a) Read the description and download the dataset generated for modelling of psychological experiments

<http://archive.ics.uci.edu/ml/datasets/Balance+Scale>

b) Select randomly ~20% for the test set.

c) Use the learning set to create a decision tree for this classification problem using RIG for selection of the most informative attribute.

d) Test the decision tree after each branching step on the test and the training set. Find the test set and the training set errors (%). Represent these errors graphically (the “learning curve”).

e) Read the attached papers: psychological experiments with human learning [1] and machine learning [2] for this classification problem.

[1] Siegler, R. S. (1976). Three Aspects of Cognitive Development. Cognitive Psychology, 8, 481-520.

[2] Shultz, T., Mareschal, D., & Schmidt, W. (1994). Modeling Cognitive Development on Balance Scale Phenomena. Machine Learning, Vol. 16, pp. 59-88.

What are the main goals of these works? Which problems they solve?

f) Discuss the result. Compare your results to these papers.

2. Car evaluation

a) Read the description and download the dataset

<http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

b) Select randomly ~20% for the test set.

c) Use the learning set to create a decision tree for this classification problem using RIG for selection of the most informative attribute.

d) Test the decision tree after each branching step on the test and the training set. Find the test set and the training set errors (%). Represent these errors graphically (the “learning curve”).

e) Read the attached papers:

[3] M. Bohanec and V. Rajkovic: Knowledge acquisition and explanation for multi-attribute decision making. In 8th Intl Workshop on Expert Systems and their Applications, Avignon, France. pages 59-78, 1988.

[4] B. Zupan, M. Bohanec, I. Bratko, J. Demsar: Machine learning by function decomposition. ICML-97, Nashville, TN. 1997.

What are the main goals of these works? Which problems they solve?

f) Discuss the result. Compare your results to these papers.