

# SEMINÁRIO

26 de Setembro de 2008

(Departamento de Matemática, sala Sousa Pinto, 11:30-12:30)

**Título:** Integration of Data Mining Methods for Various Applications.

**Orador:** Olga Kurasova, Institute of Mathematics and Informatics, Vilnius, Lithuania (kurasova@ktl.mii.lt).

**Resumo:** Nowadays huge amounts of data have been collected. However comprehension of these data remains a topic problem so far. It is very difficult for a human being to make use of such an amount of information and to be able to find basic patterns, relations, or trends in the data. In such a manner data become ever less and less valuable. Data mining could be useful for comprehension of data and help to make decisions. The combination and integrated use of data mining methods (classification, clustering, visualisation, etc.) of a different nature are under a rapid development. The combination of different methods can be applied to make a data analysis, while minimizing the shortcomings of individual methods. There are investigated following data mining methods:

- dimension reduction methods (principal component analysis, multidimensional scaling (MDS), manifold learning);
- artificial neural networks (self-organizing maps (Kohonen maps) (SOM), combination of SOM and MDS algorithm);
- classification methods (support vector machine (SVM), Naive Bayes, k-NN classifier, classification tree, rule induction).

The great attention is paid to combination of self-organizing and a special case of multidimensional scaling - Sammon algorithm. This method is applied to some problem of a different nature: statistical analysis of

curricula, comparison of schools, and analysis of the economic and social conditions of Central European countries. Other problem on integration of classification and visualization for medical diagnosis decisions is investigated. The purpose of analysis of physiological data is to evaluate men's health state and their possibility of going in for sports. The target of the analysis of ophthalmological data is to evaluate how the vectors, consisting of the parameters of eye fundus, are distributed on the plane, whether they form specific groups. Is it possible to identify glaucoma using this system of parameters?