Course 10	**Feature Subset Selection**

Program	1 Introduction
	2 Filter approaches
	  2.1 Introduction
	  2.2 Univariate filters
	    - Parametric (t-test, ANOVA, mutual information)
	    - Model-free (p metric, Mann-Whitney, Kruskal-Wallis, BSS/WSS, permutation test)
	  2.3 Multivariate filters
	    - CFS, Relief, Markov blanket
	3 Embedded methods
	  3.1 Attribute-weighted naïve Bayes
	  3.2 Classification trees
	  3.3 Random forest
	  3.4 Regularization
	4 Wrapper methods
	  4.1 Introduction
	  4.2 Classification algorithms
	  4.3 Performance measures
	  4.4 Search strategies
	5 Additional topics
	  5.1 Assessment
	  5.2 Stability

Bibliography	• H. Liu, H. Motoda (2008). Computational Methods of Feature Selection.Chapman and Hall/CRC • H. Liu, H. Motoda (1998). Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic Publishers. • Y. Saeys, I. Inza, P. Larrañaga (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23(19), 2507-2517.
Prerequisites	The students should install Weka 3.6 before the course starts.
Practical sessions	We will use both Weka and R for the practical sessions. Weka we to allow us to get a quick review of the concepts and apply them without worrying about programming. With R we will practice methods that are not available in Weka. We will not spend time on introducing Weka and R, only the bare minimum needed for a student with no prior knowledge to use them in this course.
Readings before coming	The student will benefit more from the course if she watches (some of) the following lectures before attending (these are not compulsory, only advisable): • "Introduction to feature selection" by Isabelle Guyon http://videolectures.net/bootcamp07_guyon_ifs/ • "Feature selection, fundamentals and applications" by Isabelle Guyon http://videolectures.net/mmdss07_guyon_fsf/ • "Dimensionality Reduction by Feature Selection in Machine Learning" by Dunja Mladenic http://videolectures.net/slsfs05_mladenic_drfsm/