IPM-POLYTECHNIQUE-WPI Workshop on Bioinformatics and Biomathematics, April 11-21, 2005, IPM, Tehran

Molecular Diagnosis

(Part I)

F. Markowetz Department of Computational Molecular Biology Max Planck Institute for Molecular Genetics Berlin, Germany

I will give an introduction to supervised learning with a strong focus on classification of gene expression data. As background theory I will introduce: Risk, loss functions, overfitting, regularization, kernels, and the Representer theorem. The focus will then lie on Support Vector Machines (Vapnik, 1998).

References:

A good choice is to have a look into (Hastie et al., 2001) or (Scholkopf and Smola, 2002). (What gives the second book an competitive edge is the fact that the first third of it is available online for free.)

- [1] **T. Hastie, R. Tibshirani, and J. Friedman.** The Elements of Statistical Learning. Springer, 2001. URL http://www-stat.stanford.edu/ tibs/ElemStatLearn/.
- [2] B. Schölkopf and A. J. Smola. it Learning with kernels. The MIT Press, Cambridge, MA, 2002. URL http://www.learning-with-kernels.org/. first third of the contents is available online.
- [3] V. Vapnik. Statistical Learning Theory. Wiley, N.Y., 1998.