

Stochastic simulation for biomarker validation

Olaf Tietje¹

¹HSR Hochschule für Technik Rapperswil, Oberseestrasse 10, 8640 Rapperswil

Biomarkers – such as patterns for lung cancer found in multivariate data – are complicated patterns found in few samples. Thus they are prone to overfitting, i.e. several variables are found that falsely indicate a biomarker. Because simple significance testing does not work when the number of variables exceeds the number of samples, a strong validation procedure is necessary. Stochastic simulation shows the strengths of leave-one-out validation, test set validation, stochastic validation, and repeated validation techniques. In several projects to diagnose cancer stochastic simulation shows that the exclusion of variables increases the risk of falsely positive variables and how much a repeated validation technique can improve the selection of truly positive variables.