
MDL Tutorial

Peter Grünwald

We give a self-contained tutorial on the Minimum Description Length (MDL) approach to modeling, learning and prediction. We focus on the recent (post 1995) formulations of MDL, which can be quite different from the older methods that are often still called 'MDL' in the machine learning and UAI communities.

In its modern guise, MDL is based on the concept of a 'universal model'. We explain this concept at length. We show that previous versions of MDL (based on so-called two-part codes), Bayesian model selection and predictive validation (a variation of cross-validation) can all be interpreted as approximations to model selection based on 'universal models'. Modern MDL prescribes the use of a certain 'optimal' universal model, the so-called 'normalized maximum likelihood model' or 'Shtarkov distribution'. This is related to (yet different from) Bayesian model selection with non-informative priors. It leads to a penalization of 'complex' models that can be given an intuitive differential-geometric interpretation. Roughly speaking, the complexity of a parametric model is directly related to the number of distinguishable probability distributions that it contains. We also discuss some recent extensions such as the 'luckiness principle', which can be used if the Shtarkov distribution is undefined, and the 'switch distribution', which allows for a resolution of the AIC-BIC dilemma.

The talk assumes no prior knowledge of information theory. The menu is as follows:

1. Codes and Probability Distributions
2. Universal Coding
3. The Bayes, 2-part and Normalized Maximum Likelihood Universal Model
4. MDL Model Selection
5. Relation to Bayes factor model selection and Cross-Validation
6. The Luckiness Principle, The Switch Distribution