MDL

(Minimum Description Length)

Hannes Wettig

CoSCo

http://cosco.hiit.fi/

http://www.mdl-research.org/

MDL

The shortest Code is the best Model

$$P(D; M) = 2^{-L(D;M)}$$

$L(D; M)$ **approximates the Kolmogorov Complexity**

The Best Model

- compresses the data
- describes main features in short
- predicts
- denoises
- ...

Classic MDL

two-part code

$$L_2(D;M) = L(M,\Theta) + L(D;M,\Theta)$$

$\longrightarrow$ **redundancies**

Modern MDL

NML (Normalized Maximum Likelihood)

$$P_{NML}(D;M) = \frac{P(D;M,\widehat{\Theta}(D))}{\sum_{D'} P(D;M,\widehat{\Theta}(D'))}$$

$\longrightarrow$ **shorter code, better model**

$\longrightarrow$ **minimax optimal**

$\longrightarrow$ **loses** $\log \sum_{D'} P(D;M,\widehat{\Theta}(D'))$ **("regret") bits on any data against hind-sighted opponent**

NML: HowTo

$$P_{NML}(D; M) = \frac{P(D;M,\hat{\Theta}(D))}{\sum_{D'} P(D;M,\hat{\Theta}(D'))}$$

- no (parameter) prior needed
- use it for model class selection
- use it for clustering, prediction, denoising, ...

NML: Problems

$$P_{NML}(D;M) = \frac{P(D;M,\widehat{\Theta}(D))}{\sum_{D'} P(D;M,\widehat{\Theta}(D'))}$$

- regret typically incomputable
- still need to code the model class
- model class vs. parameters
- regret is a function of $|D|$

NML: Applications

$$P_{NML}(D;M) = \frac{P(D;M,\hat{\Theta}(D))}{\sum_{D'} P(D;M,\hat{\Theta}(D'))}$$

- **Image/Signal Denoising**
- **Clustering**
- **Density Estimation**
- **Bayesian Network Selection Criterion**
- **Hot Topic: Conditional/Supervised/Local NML**