Introduction to Machine Learning

Feature Selection: Wrapper methods

Feature Selection: Wrapper methods

Learning goals

- Understand how wrapper methods work
- Understand how they can help in feature selection
- Know their advantages and disadvantages

Learning goals

- Understand how wrapper methods work
- Forward + backward search, EAs
- Advantages and disadvantages

OBJECTIVE FUNCTION

Given p features, **best-subset selection problem** is to find subset $S \subseteq \{1, \dots p\}$ optimizing objective $\Psi : \Omega \to \mathbb{R}$:

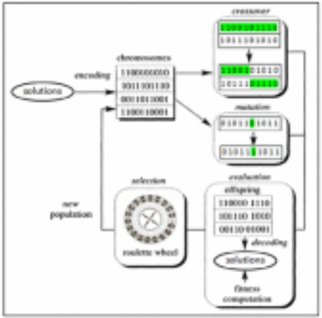
$$S^* \in \arg\min_{S \in \Omega} \{\Psi(S)\}$$

- Ω = search space of all feature subsets $S \subseteq \{1, \dots, p\}$. Usually we encode this by bit vectors, i.e., $\Omega = \{0, 1\}^p$ (1 = feat. selected)
- Objective Ψ can be different functions, e.g., AIC/BIC for LM or cross-validated performance of a learner
- Poses a discrete combinatorial optimization problem over search space of size = 2^p, i.e., grows exponentially in p (power set)
- Unfortunately can not be solved efficiently in general (NP hard; see, e.g., Natarajan, 1995)
- Can avoid searching entire space by employing efficient search strategies, traversing search space in a “smart” way

EXTENSIONS: GENETIC ALGORITHMS FOR FS

Example Template for $(\mu + \lambda)$-Evolutionary Strategy applied to FS

1. Initialization: μ random bit vectors (feature inclusion/exclusion)
2. Evaluate model performance for bit vectors
3. Select μ fittest bit vectors (parents)
4. Generate λ offspring applying crossover and mutation
5. Select μ fittest bit vectors from $(\mu + \lambda)$ options for next generation
6. Repeat steps 2-5 until stopping criterion is met



- Use CV/validation set for evaluation to avoid overfitting
- Choice of μ and λ allows some control over exploration vs. exploitation trade-off
- See our sjds-lmu2021 for further information

MitsuoGen1996