Machine learning handbook

Feature selection

Taeyang Yang

Update: August 14, 2018

BCILAB, UNIST

Feature selection

- Feature selection = select a subset of features

- Necessity of feature selection process
  1. Redundancy
  2. The number of features ~ the number of parameters to be optimized
  3. A few features → easier to interpret the result
  4. Faster prediction
  5. Storage efficiency

Feature selection approaches

- Three approaches
  - Filter: **measure of relationship between each feature and target class**
    - Pearson correlation coefficient: Linear dependencies (Hastie *et al.*, 2001)
    - Pearson correlation coefficient (square/log of features): Non-linear dependencies
    - Mutual information
    - Maximum Relevance Minimuim Redundancy (mRMR) (Peng et al., 2005; Mühl et al., 2014)
    - R^2 (Muller et al., 2006; Vaughan et al., 2006)
  - Wrapper: **use a classifier to obtain a subset of features**
    - Support vector machine for channel selection (Lal *et al.*, 2004)
    - Linear regressor for knowledge extraction (Liang and Bougrain, 2012)
    - Genetic algorithm for spectral feature selection (Corralejo *et al.*, 2011)
    - Evolutionary algorithm for feature selection (Ortega *et al.*, 2016)
  - Embedded approach: **integrate feature selection and the evaluation**
    - Decision tree (Quinlan, 1986; Breiman et al., 1994)
    - Multilayer perceptron with optimal cell damage (Cibas *et al.*, 1994)
    - Stepwise linear discriminant analysis (Krusienski *et al.*, 2006)