Feature Selection for Classification

Jyh-Shing Roger Jang (張智星)

MIR Lab, CSIE Dept.

National Taiwan University

jang@csie.ntu.edu.tw , *http://mirlab.org/jang*

2023/11/29

Outlines

- Introduction to feature selection
- Heuristic search
  - One-pass ranking
  - Sequential forward selection
- Exhaustive search
- Examples

Intro. to Feature Selection

- Feature selection
  - Also known as input selection
- Goal
  - To select a subset from the original feature set for better accuracy
- Items to be specified before feature selection Quiz!
  - Classifier, such as KNNC
  - Performance index, such as accuracy
  - Performance evaluation method, such as k-fold CV
- Benefits
  - Better accuracy
  - Less computation
  - Explainability between features and outputs

Feature Selection vs. Extraction

- Common part
  - Both known collectively as dimensionality reduction
  - Goal: Reduced model complexity with improved accuracy
- Feature selection: select the best subset from the original features
- Feature extraction: Extract new features by a linear or nonlinear combination of all original features
  - Extracted features may not have physical meanings
  - Examples of linear feature extraction
    - PCA (unsupervised)
    - LDA (supervised)

Heuristic Search

- A number of heuristic search for feature selection
  - One-pass ranking
  - Sequential forward selection (SFS)
  - Sequential backward selection (SBS)
  - Generalized sequential forward selection
    - Select the best k features at each iteration (k=1 for SFS)
  - Generalized sequential backward selection
    - Delete the best k features at each iteration (k=1 for SBS)
  - Sequential forward floating selection (SFFS)
  - Sequential backward floating selection (SBFS)
  - 'Add m, remove n' selection
  - Generalized 'add m, remove n' selection

One-pass Ranking

- Steps
  - Sort the given d features in descending order of their accuracy based on a single feature only
  - Select the top m features from the sorted list that has the best performance
- Complexity

  Quiz!
  - If the dataset has d features, we need to perform $2d$-1 CV.
- Properties Quiz!
  - Advantage: Extremely fast
  - Disadvantage:
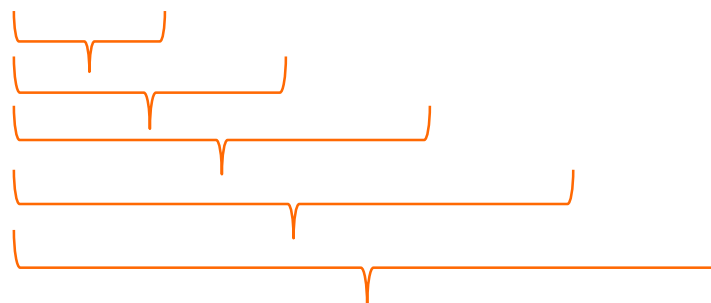   - Feature correlation is not considered
    - Selected features are not always optimal

Example of One-Pass Ranking

o SFS with 4 features

Original order ➡ [x1] [x2] [x3] [x4] [x5]

After ranking ➡ [x3] [x1] [x5] [x2] [x4]

Sequential Forward Selection (SFS)

- Steps for sequential forward selection
    1. Select the first feature that has the best accuracy.
    2. Select the next feature (among all unselected features) that, together with the selected features, gives the best accuracy.
    3. Repeat the previous step until all features are selected.
- Complexity
    - If the dataset has d features, we need to perform $d(d+1)/2$ CV.

        Quiz!
- Properties   Quiz!
    - Advantage: Fast
    - Disadvantage: Selected features are not always optimal.

Example of SFS

- SFS with 4 features
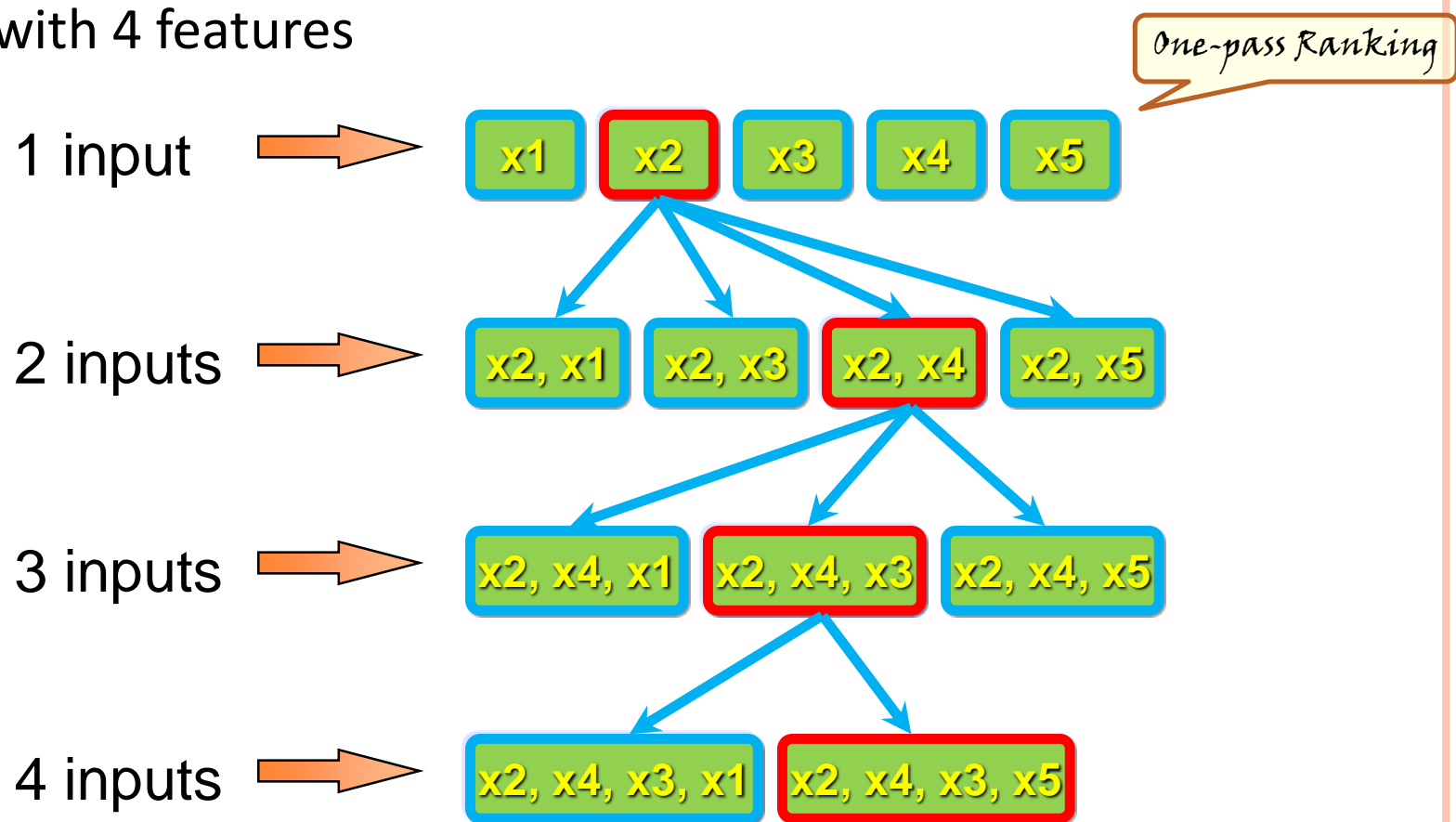
Exhaustive Search

- Steps for exhaustive search (ES)
    1. Generate all combinations of features and evaluate them one-by-one
    2. Select the feature combination that has the best accuracy.
- Drawback

    Quiz!
    - d features ➜ 2^d-1 CV for performance evaluation
    - d = 10 ➜ 1023 CV for evaluation ➜ Time consuming!
- Properties
    - Advantage: Optimal feature set can be identified.
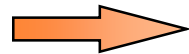   - Disadvantage: Extremely slow if no. of features is large.

Exhaustive Search

○ Direct exhaustive search

One-pass Ranking
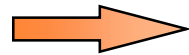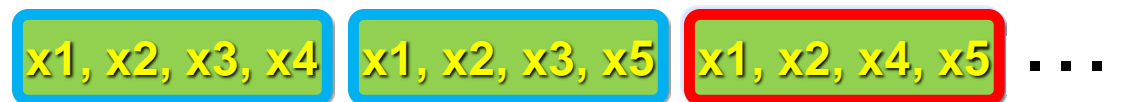
1 input ➡️ x1 x2 x3 x4 x5

2 inputs ➡️ x1, x2 x1, x3 x1, x4 x1, x5 x2, x3 ...

3 inputs ➡️ x1, x2, x3 x1, x2, x4 x1, x2, x5 x1, x3, x4 ...

4 inputs ➡️ x1, x2, x3, x4 x1, x2, x3, x5 x1, x2, x4, x5 ...

...
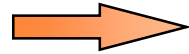
Summary of Computational Complexity

○ No. of CV required for feature selection in a dataset of d features

- One-pass ranking ➔ $2d$-1
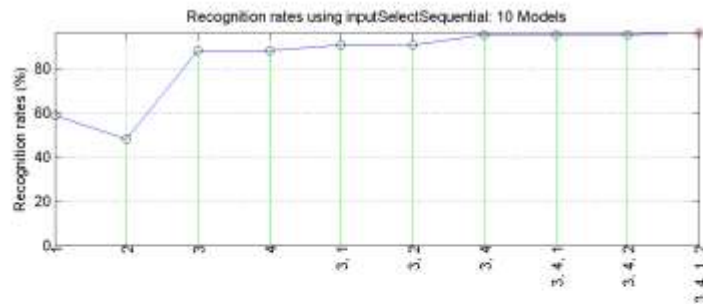- Sequential forward selection ➔ $d(d+1)/2$
- Sequential backward selection ➔ $d(d+1)/2$
- Exhaustive search ➔ 2^d-1

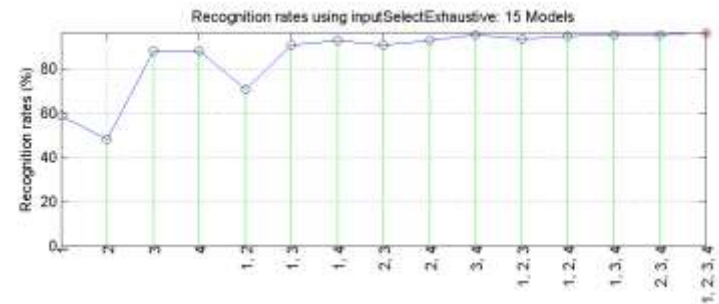○ No. of CV required for selecting up to m features in a dataset of d features

- One-pass ranking ➔ d+m-1
- Sequential forward selection ➔ ???
- Sequential backward selection ➔ ???
- Exhaustive search ➔ ???
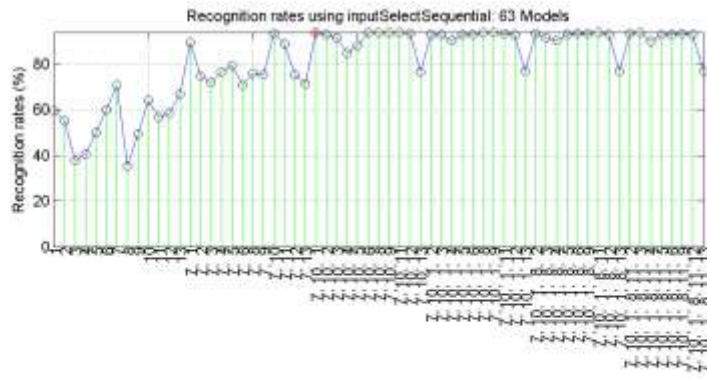
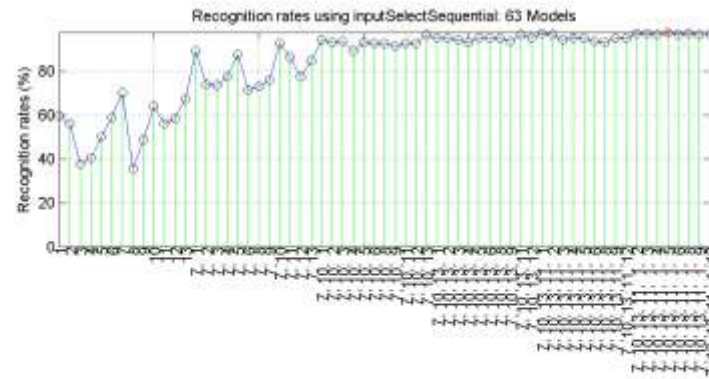# Feature Selection for Iris Dataset

○ SFS

○ Exhaustive search

Feature Selection for Wine Dataset

○ SFS

○ SFS with input normalization

○ Summary

- SFS ➔ 3 selected features, LOO accuracy=93.8%

- SFS with feature normalization ➔ 6 selected features, LOO accuracy=97.8%

- ES with feature normalization ➔ 8 selected features, LOO accuracy=99.4%

Proper Use of Feature Selection

- Common use of feature selection
  - Increase model complexity sequentially by adding more features
  - Select the model that has the least validation error
- Typical curve of error vs. model complexity
  - Determine the model's complexity with the least validation error