Statistical Pattern Recognition
Assignment 5

Ali Gholami

Department of Computer Engineering & Information Technology
Amirkabir University of Technology

https:// aligholamee.github.io
aligholami7596@gmail.com

Abstract

Keywords. *Dimensionality Reduction, Principal Component Analysis, Fisher Linear Discriminant Analysis, Feature Subset Selection, Sequential Feature Selection, Data Visualization & Representation.*

1 Sequential Feature Selection

Given the following objective function, use **SFS**, **SBS** and **Plus-2 Minus-1 Selection** to select 3 features:

$$J(x) = 5x_1 + 7x_2 + 4x_3 + 9x_4 + 3x_5 - 2x_1x_2 + 2x_1x_2x_3 - 2x_2x_3 - 4x_1x_2x_3x_4 + 3x_1x_3x_5$$

Solution

SFS

In this method, we start feature selection from an empty set. We add features one by one and compute the value of the objective function with respect to each of the features being added. The feature with the largest objective function will be selected. The iteration goes on until all features all covered. We then select ideal features (subset with k features and maximum objective). The actual algorithm is as following:

1. Start with the empty set $Y_0 = \emptyset$
2. Select the next best feature $x^+ = argmax[J(Y_k + x)]$
3. Update $Y_{k+1} = Y_k + x^+; \quad k = k + 1$
4. Go to 2

Below is the demonstration of iterations taken to completely explore the search space. The first iteration is:

- $J(x_1) = 5$
- $J(x_2) = 7$
- $J(x_3) = 4$
- $J(x_4) = 9$
- $J(x_5) = 3$

According to the heuristic nature of sequential subset selection, we'll choose x_4 as the first best feature. We'll then generate subsets containing combination of features with x_4:

- $J(x_4x_1) = 14$
- $J(x_4x_2) = 16$
- $J(x_4x_3) = 13$
- $J(x_4x_5) = 12$

Thus, features x_4 and x_2 are selected until now. We'll drive the 3 sized subsets:

- $J(x_4x_2x_1) = 19$
- $J(x_4x_2x_3) = 18$
- $J(x_4x_2x_5) = 22$

Three best features selected by the algorithm are x_4, x_2 and x_5.

SBS

This method initiates the feature selection procedure using a complete subset of features. It then removes each feature and evaluates the objective function. The feature that causes the lowest decrease in the objective function will be remove (useless feature!). We'll stop when we reach a satisfying 3 sized feature subset. The algorithm is formally working as follows:

1. Start with the full set $Y_0 = X$
2. Remove the worst feature $x^- = argmax[J(Y_k - x)]$
3. Update $Y_{k+1} = Y_k - x^-; \quad k = k + 1$
4. Go to 2

Applying this algorithm on the given objective function yields the following results:

- $J(x_1x_2x_3x_4x_5) = 25$