

SOM Jeremy (2017): Comparison of search strategies for feature selection using ELM

Abstract

In geosciences, natural phenomena can be described from tens to hundreds of features. Feature selection is needed to reduce this high dimensionality and to have a better understanding of the relationships between the features and the phenomena.

There are three main families of search strategies in feature selection: exhaustive, greedy and random methods. Exhaustive search ensures finding the best subset of features by evaluating every possibilities, but is time consuming and may even be impossible to compute when the dimensionality exceeds 15.

With thirteen variables, the proposed feature space for wind speed in Switzerland has a relatively low dimension. Performing an exhaustive search on this dataset allows having a comparison basis for the other methods. In particular, its results are compared to those of greedy methods (sequential forward selection and sequential backward selection) and random search without optimization. An artificial dataset is used to implement the algorithms and to help in the interpretation of the results provided by the neural network.

The aim of this work is to compare the ability of four algorithms to search for an optimal subset of features by using a recently developed neural network algorithm called extreme learning machine, which has shown good generalization performances at very fast learning speeds.