Hybrid Variable Selection and Support Vector Regression for Gas Sensor Optimization

Rebolledo, Margarita; Stoean, Ruxandra; Eiben, A. E.; Bartz-Beielstein, Thomas

published in
Bioinspired Optimization Methods and Their Applications
2020

DOI (link to publisher)
10.1007/978-3-030-63710-1_22

document version
Publisher's PDF, also known as Version of record

Hybrid Variable Selection and Support Vector Regression for Gas Sensor Optimization

Margarita Rebolledo[1(✉)], Ruxandra Stoean[2,3], A. E. Eiben[4], and Thomas Bartz-Beielstein[1]

[1] Institute for Data Science, Engineering, and Analytics, TH Köln, Köln, Germany
{margarita.rebolledo,thomas.bartz-beielstein}@th-koeln.de
[2] Romanian Institute of Science and Technology, Cluj-Napoca, Romania
[3] Grupo Ingeniería de Sistemas Integrados, E.T.S.I. Telecomunicación, Universidad de Malaga, Malaga, Spain
ruxandra.stoean@rist.ro
[4] Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
a.e.eiben@vu.nl

Abstract. The improvement of combustion processes in industry, especially in the automotive branch, is of great importance to maintain the environmental permitted limits. Carbon monoxide concentration in the exhaust gases can give an insight into the efficiency of the combustion taking place and for this reason, it is important to have sensors that can measure it accurately. First results of a long term study with one of the leading sensor manufactures showed high performance using genetic programming. However, this expensive approach is difficult to apply in real-world settings. Therefore a hybrid optimization that combines support vector regression (SVR) with variable pre-selection is proposed. Three different methods for variable selection are compared for this application, a genetic algorithm, and two methods from Bayesian statistics: statistical equivalent signatures and projection predictive variable selection. Furthermore, a multi-objective approach using the same hybrid definition is implemented for the cases in which several sensors need to be considered simultaneously. Our results show that the hybrid model is an improvement compared to the previous study, while delivering good performance when dealing with a multivariate formulation. Genetic algorithms in combination with SVR lead to enhanced variation on the groups of selected variables.

Keywords: Support vector regression · Feature selection · Projection predictive · Statistical equivalent signatures

1 Introduction

The automotive industry is increasingly concerned with building high performance cars while also adhering to the normative set in place for environment

B. Filipič et al. (Eds.): BIOMA 2020, LNCS 12438, pp. 281–293, 2020.
https://doi.org/10.1007/978-3-030-63710-1_22

protection. Among other things, this requires a cutback on the emission levels of carbon monoxide. The efficiency increase of motor combustion processes plays an important role in the reduction of pollutant levels. The different concentrations of gases resulting from combustion allow to make an approximate analysis of said efficiency and thus a reliable in-situ sensing system is required.

In addition to the already existent oxygen sensor, a reliable carbon monoxide in-situ sensor is also needed. The sensor should have a high sensitivity for carbon monoxide and be able to vary its output according to the proportion of gases found in the exhaust gases. Furthermore the interpretability of the system describing these sensors needs to be maintained in order to keep a clear understanding of how each different gas affects the system output.

Under these conditions, an accurate and informative computational model can be constructed to examine the dependency of the sensor output to the complex interactions of the exhaust gases. Therefore, an important trait of a model would be the ability to reveal if there are such synergies between the attributes. Finally, the underneath workings of the combustion process should be better revealed by more than one sensor at a time, which would require a multivariate approach for an adequate modelling of the entire problem. The findings can then determine if the designed sensors are robust or need to be further enhanced by the engineers.

Rebolledo et al. [17] proved the efficiency of Genetic Programming (GP) when modelling single carbon monoxide gas sensors. However, the expensive GP implementation and its excessive time demanding approach made the method infeasible for the real world setting. Instead the Least Absolute Shrinkage and Selection Operator (LASSO) [19] model was favored given its interpretability and ability to identify the effect each variable had on the output.

The main goal of this paper is to put forward a hybrid model strategy to improve the modelling of single sensors and go one step further into the formulation and solution of the multi-objective case. This in order to handle the situation when there is more than one sensor required to be investigated. The well performing support vector regression (SVR) technique is combined with variable selection methods to accelerate the sensor optimization and at the same time to enable the understanding of parameter importance. Furthermore, a lower number of parameters will limit SVR high flexibility, reducing in this way its variance and avoiding overfitting [10].

Three different methods of variable selection are compared for this application: a genetic algorithm (GA) and two methods from Bayesian statistics, i.e. the statistically equivalent signature (SES) and projection predictive variable selection (Projpred). By this we want to answer two questions:

Q-1. Can these methods be applied as a feasible approach to single and multivariate sensor optimization scenarios yielding minimal mean square error (MSE)?

Q-2. Which method performs best among the ones tested?

The findings will help determine whether the designed sensors are robust to work in single and/or multiple arrays, while maintaining a clear presentation of the influence that the different gases have on the sensor reading.

The paper is structured in the subsequent manner. Section 2 presents the description of the problem and the available data set. Section 3 explains the setup of our experiments and the workings of the three tested hybrid algorithms. Section 4 presents the results obtained. Finally, Sect. 5 discusses results and presents answers to questions (Q-1) and (Q-2).

2 Problem Description

The efficiency of motor combustion processes can be indirectly measured by monitoring the concentrations of carbon monoxide and other harmful gases releases into the atmosphere. This paper focuses on the modelling and optimization of a sensor that is able to discern carbon monoxide concentration apart from other six exhaust gases. This is difficult because the sensor is exposed and influenced by these gases. Thus, the sensor output will represent the underneath process influenced by all the other gases and not directly the carbon monoxide concentration.

The optimization task can be given in a general formulation as two problems ranging from single to multi-objective. Given the data set $\{\mathbf{A}_i, \mathbf{B}_i\}$, $i = 1, 2, ..., m$, where each $\mathbf{A}_i = (A_i^1, A_i^2, ..., A_i^n)$ refers to the concentrations of the X_1, X_2, ..., X_n gases involved in the combustion process, with $A_i^j \in \mathbb{R}$, $j = 1, 2, ..., n$, and every $\mathbf{B}_i \in \mathbb{R}^p$ denotes the output of the p sensors Y_1, Y_2, ..., Y_p (i.e., a multivariate regression formulation), the two main objectives of the experiments can be named as:

1. *Find* the (combination of) predictors A_i^j that *minimizes* the MSE for each sensor measurement independently.
2. *Find* the (combination of) predictors A_i^j that *minimizes* the MSE for all sensor measurements $Y_1, Y_2, ..., Y_p$ simultaneously.

A first simple way to address the presence of several sensor is to appoint a naive approach where the fitness evaluation of an individual is a summation of several objective function, i.e. sum of SVR estimated MSE for each output. $f(c) = f_1(c) + f_2(c)... + f_p(c)$.

At this point the hybrid approaches GA-SVR, SES-SVR and Projpred-SVR can be appointed to solve both the single and the multi-objective task.

2.1 Data Description

The data set for this work was presented in [17] and is included on the R package SPOT [2] where it can be freely accessed for comparative studies.

During four years, the data was collected from an extensive real-world project in cooperation with one of the leading sensor manufacturers. The data was

recorded from a series of experiments following a response surface design. This design constraints itself to the maximum and minimum expected concentration values of each gas under normal working conditions. Given the cost and time consumption required for the experiments, only a limited amount of samples could be measured.

The data contains $m = 140$ samples and every record is defined by seven attributes, X_1 to X_7, which represent the concentrations of each of the measured exhaust gases, here anonymized due to confidentiality reasons. Carbon monoxide is identified as X_1. Two sensor outputs, denominated $Y1$ and $Y2$, were recorded. The data is standardized, meaning that every sample had its mean subtracted and was then divided by the standard deviation. Following the nomenclature of the above defined optimization task, the variables m, n and p have the following particular values: $m = 140$, $n = 7$, $p = 2$.

3 Experiments

Following the standard procedure in machine learning described in detail in [7], the data is divided into training, test, and validation sets.

To give stability and statistical significance to the results, several experimental runs were performed by drawing different compositions for the three sets. 30 partitions are constructed by repeatedly selecting 80/25/35 samples correspondingly. The selection is done randomly.

The variable selection and model building will be evaluated in terms of MSE using the validation set. The best obtained result will be evaluated on the test set to acquire the final result.

All the executed experiments use the R package *e1071* [12] to implement the SVR. All the instances of SVR use the radial basis function (RBF) kernel, $\exp(-\gamma||x - x'||^2)$, with parameter $\gamma = 0.1$ indicating the spread of the kernel.

The exact implementation of the different hybrid methods differs according to their inner workings and their exact experimental settings are presented in more detail in the following sections.

3.1 GA-SVR

In this approach a genetic algorithm (GA) [5] selects the predictors X_i that influence the output for each available sensor and learns from the training data.

Evolutionary and swarm computation have been often successfully partnered with support vector machines for variable or parameter selection, as demonstrated by other application areas in industry, medicine and biology [8,9,14,18].

Since the current task is variable selection, a binary representation was chosen. An individual is a binary vector $c \in \{0, 1\}^n$, where $c_n = 1$ signifies that the corresponding gas n influences the sensor signal and $c_n = 0$ that its effect on the sensor output is insignificant.

This approach is depicted by Algorithm 1. Here $no\_of\_repeats$ corresponds each of the 30 train/validation/test sets partitions.

Evolution at each repeat follows the standard cycle of parent selection, variation, and survivor selection. The GA binary encoded individual c indicates the attributes that will be included in the modelling step. A SVR is trained on the obtained attribute collection and the MSE is computed on the validation data. The MSE value is returned as the current individual's fitness. At the end of each GA run, the best individual $best_l$, is retained. After the 30 trials are finished, the number of times a feature l was selected in the preserved best individuals is counted in $score_l$. Therefore, a ranking of the involved attributes is achieved. Additionally, the MSE of each of the preserved best individuals is computed on its respective test set. The final test MSE is obtained as the average of the results over these runs.

```
for i = 1 : no_of_repeats do
    use train/validation/test set partition i ;
    initialization of population pop_GA;
    evaluate pop_GA by calculating MSE;
    for j = 1 : no_of_generations do
        parent selection in pop_GA;
        variation in pop_GA;
        obtain offspring population off ;
        evaluate off by calculating MSE;
        survival selection in pop_GA;
    end
    store the best individual of pop_GA in best_i;
end
for l = 1 : n do
    score_l = sum of selected attributes in (best);
end
rank variables according to score;
```

Algorithm 1: Hybrid GA-SVR algorithm. The algorithm accepts the different partitions of the data set and returns a score of most important variables.

The GA-SVR meta-heuristic can be easily extended to be able to simultaneously handle several sensors. The objective function will be the summation of each several objective function as stated on the problem definition. The problem can be therefore defined as a multi-objective discrete and combinatorial optimization problem [4], where every objective refers to the measurement of one sensor and needs to be minimized.

The GA was implemented using the R package *genalg* [22]. The GA population size is set to 20, with 50 generations. Bit flip mutation with a probability of 0.3 was used with elitist selection.

3.2 SES-SVR

Statistically equivalent signature (SES) [11] is a constraint-based feature selection algorithm with roots in causal analysis, where the optimal set of predictors consist in the Markov Blanket (MB) of the variable in the Bayesian Network (BN) representing the data distribution [21]. SES has already been proved to work on several high-dimensional gene-expression data sets including temporal data [20] and text mining applications [1].

Given a subset of variables, **W**, an statistical independence test, *ind*(), is used to test the null hypothesis that a variable **X** is conditionally independent on the output **T** given **W**, *ind*(**X,T**|**W**). Variables that cannot be proven as independent, that is they show a connection (functional relation) to the output, are selected.

Once the variables with the most expected predictive power have been selected, a SVR model is built on the training set including only these variables. The group of variables with the smallest MSE in all validation sets is selected as the best. For the final test MSE, the SVR models are generated again on all 30 data partitions using the best variables and the average MSE on the test sets is computed. Algorithm 2 illustrates the steps for the SES-SVR approach.

for $i = 1$ *: no_of_repeats* **do**
  use train/validation set partition i ;
  initialize variable selection algorithm ;
  Select variables with the configured criteria ;
  estimate MSE in validation set ;
end
store selected variables with best MSE in *best* ;
for $l = 1 : n$ **do**
  $score_l$ = sum of selected attributes in (*best*);
end
for $i = 1$ *: no_of_repeats* **do**
  use test set partition i ;
  estimate MSE using *best* in test ;
end
rank variables according to *score* ;

Algorithm 2: General algorithm for SES-SVR and Projpred-SVR. The algorithm accepts the different partitions of the data set and returns a score of most important variables.

The multi-objective formulation needs only a new definition of the independence test used while selecting the variables. In this case the multivariate regression test is applied. The MSE of the best variables will be defined as the MSE sum of all the sensors outputs.

SES was implemented using the R package *MXM* [11]. Since both variables and outputs have continuous values, the Fisher test (*testIndFisher*) is computed

for the single objective case and the multiple regression test (*testIndMVred*) for the multiobjective case. A maximum of four variables is used as the conditioning set and the threshold for the p-value is set at 0.05 because this is considered as a standard value.

3.3 Projpred-SVR

The projection predictive variable selection (Projpred) [16] is a Bayesian model selection method, in which the posterior information of a reference model that includes all possible variables (M_*) is projected onto candidate models ($M_\perp$) containing only a subset of the variables. The goal is to find a submodel $M_\perp$ whose predictive distribution is as close as possible to that of M_*. The Kullback-Leibler (KL) divergence is used to determine the divergence between both distributions.

The method works as follows: a Gaussian linear model is used to build the reference model M_* with input variables f_i as in Eq. 1.

$$\begin{aligned} f_i &= \mathbf{W}^\top \mathbf{X}_i \\ y_i &= f_i + \epsilon_i,\ \epsilon \sim N(0, \sigma^2) \end{aligned} \tag{1}$$

To encourage sparsity an extra prior is added to the weights $\mathbf{W} = (w_1, ...w_n)$ to count for their relevance or irrelevance to the output. The Horseshoe prior [3] accomplishes this by introducing a global scale, τ, inferred by the data, and a local scale, λ, inferred by $\mathbf{W}$, as seen in Eq. 2, where t_v^+ refers to the half-Student-t prior with $v = 1$ degrees of freedom. Both scale parameters are unknown quantities and will be inferred during the Markov chain runs. The scale parameter λ will be high for inputs with high relevance and small for those with low or no relevance.

$$\begin{aligned} w_i|\lambda_i, \tau &\sim N(0, \lambda_i \tau) \\ \lambda_i &\sim t_v^+(0, 1) \end{aligned} \tag{2}$$

After model fitting is finish, variable selection starts searching for important variables using L1-search, a LASSO related method in which a subspace to project the model is defined using L1 constraints on the parameters of the full model [13]. The variables that achieve the most similitude to the predictive distribution of the original model are selected.

The algorithm works following the same steps as illustrated in Algorithm 2. First the full Gaussian linear model is fitted using the horseshoe prior. To specify the prior beliefs about the number of relevant variables, the results from [17] are used. According to their findings four variables showed a higher influence on the model output. This information will be transmitted to the model through the prior definition. After the full model is fitted, the variable selection starts using L1-search. The chosen variables are the ones that most decrease the KL-divergence between the predictive distribution of the full model and the one of the candidate model. At the end of the runs the variables with minimum MSE on the validation set are chosen as the best. The SVR model uses only the best

variables to give the final result on the test set. Moreover, the best variables are encoded as a binary vector to allow to calculate the ranking for each attribute.

The implementation of the described algorithm is implemented using the *projpred* R package [15]. The model fitting was done using the *rstanarm* package [6]. The definition of the Horseshoe prior uses a global scale parameter ≈ 0.149. Four Markov chains are run, each with 1000 iterations and leaving 500 iteration as burn-in.

4 Results

The constructed algorithms were applied to the gas sensor data described in Sect. 2. The experiments were first conducted for the single objective formulation using the output given by sensor Y_1 and subsequently by sensor Y_2. The results obtained on the test data are shown in Table 1. These correspond to the average MSE value across all 30 test set partitions. To enable comparison with previous results, the results obtained using LASSO in [17] are also presented as a baseline.

Table 1. MSE with standard deviation obtained for GA-SVM, SES-SVM and Projpred-SVM on the single-objective formulation experiments. As a baseline the results obtained in [17] using LASSO are also presented. Smaller values mean better performance.

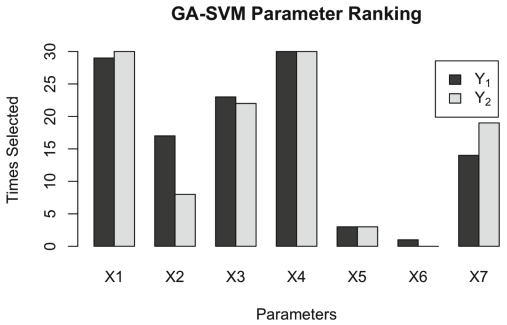
	GA-SVM	SES-SVM	Projpred-L1	Baseline
Y1	0.3321 ± 0.0875	0.3367 ± 0.1062	0.3928 ± 0.1375	0.56
Y2	0.2827 ± 0.0633	0.2868 ± 0.0797	0.2880 ± 0.0721	0.27

The variable ranking given by all the three different methods coincide on clearly pointing parameters X_1 and X_4 as the ones with the strongest influence on outputs Y_1 and Y_2. Results start to diverge when observing the other parameters. While GA-SVR algorithm finds parameters X_3 and X_7 as the second most influential parameters this trend, although visible, is not as marked for SES-SVR or Projpred-SVR. Figure 1 shows the ranking differences between two of the algorithms. It is clear that the GA method includes more variation in variable selection.

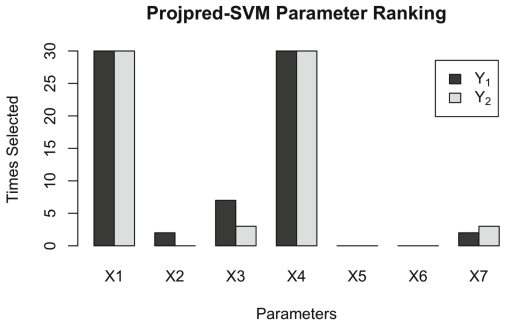
The MSE results for the multi-objective formulation are given separately for the two sensors in Table 2.

It is interesting to observe that the performance of the compared methods is not greatly affected when changing between single and multi-objective formulations. It is again significantly improved when compared to the baseline in Table 1.

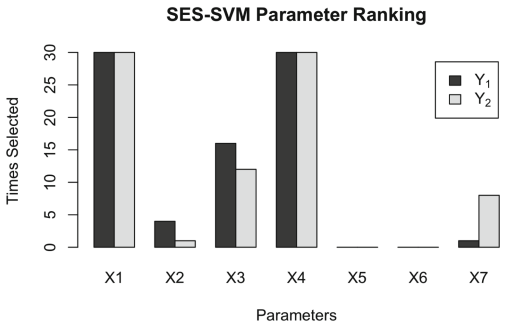
The parameter ranking for the multi-objective formulation of all three methods is shown in Fig. 2. The same most influential parameters are identified and the least important X_5 and X_6 are hardly considered. These two parameters



(a)



(b)



(c)

Fig. 1. Variable importance ranking for the (a) GA-SVR, (b) Projpred-SVR, and (c) SES-SVR. All three algorithms are in their single objective formulation for Y_1 and Y_2.

Table 2. MSE with standard deviation obtained for GA-SVM, SES-SVM and Projpred-SVM on the multi-objective formulation. Here smaller values mean better performance.

	GA-SVM	SES-SVM	Projpred-L1
Y1	0.3619 ± 0.0917	0.3460 ± 0.0882	0.3525 ± 0.0939
Y2	0.2993 ± 0.0641	0.3010 ± 0.0759	0.3022 ± 0.0721

are only selected by GA-SVR on a low number of occasions. It is interesting to note that the variable selection implemented by the GA maintained the same behavior in the single- and multi-objective case.

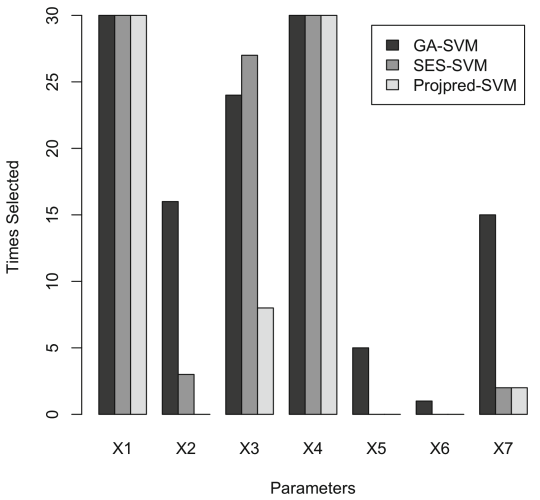


Fig. 2. Ranking of the variable importance of all three methods in the multi-objective formulation.

Lastly, it is of interest to know the difference in wall-clock time each of these methods need in order to complete the described experiments. Using a dual core 1.4 GHz Intel Core i5 processor the time required for the multi objective formulation was 10 min, 12 s and 15 min for GA-SVR, SES-SVR, and Projpred-SVR respectively.

5 Conclusions

The current data analysis considers seven gases resulting from a combustion process and two built-in sensors to measure the concentration of each gas as provided by an industrial testing station. The optimization problem required to find the minimal MSE while preserving the system interpretability.

Our proposed solution makes use of the well performing SVR to model the sensor output. To avoid overfitting and to maintain model interpretability only a subset on the input parameters is used when building the model. Three variable selection methods are tested in order to select the input subset: Projection predictive method, binary genetic algorithm, and statistical equivalent signatures. There three methods use different strategies when selecting the most important variables.

The experiments show there is clear improvement over the results presented in [17] and prove the hybrid approach has good performance for both the single- and multi-objective formulation of the gas sensors.

To answer the first question (Q-1), can these methods be applied as a feasible approach to single and multi-variate sensor optimization scenarios yielding minimal MSE?, single- and multi-objective solutions were tested. All three methods showed an increase in the performance when compared to the baseline in the single objective approach. This performance level was maintained when testing the methods on the multi objective scenario.

In the single-objective formulation the GA-SVR approach showed slightly better results than the other two competing methods. Even though LASSO was the preferred method in [17], the analog implementation used in Projpred did not show any solid advantage. In the multi-objective formulation all methods presented a performance comparable to that obtained on the single sensor approach. Here GA-SVR showed again a slightly better result.

Regarding the second question (Q-2), which method performs best among the ones tested?, the dynamics of the variable selection were observed across the experiments. As seen in the results all three approaches showed similar performances but there were two significant differences. On the one hand, GA-SVR shows more variation between the best variable groups while the other two methods find only a couple different best options and repeat them for several models. On the other hand, SES-SVR allows an implementation that is by far faster that the other two methods. Following these observations GA-SVR is our preferred method in scenarios where there is no time pressure. Here the variability on the best variable groups can be beneficial for data sets with complex input interactions.

References

1. Adamou, M., Antoniou, G., Greasidou, E., Lagani, V., Charonyktakis, P., Tsamardinos, I.: Mining free-text medical notes for suicide risk assessment. In: Proceedings of the 10th Hellenic Conference on Artificial Intelligence, SETN 2018, pp. 47:1–47:8. ACM, New York (2018). https://doi.org/10.1145/3200947.3201020, http://doi.acm.org/10.1145/3200947.3201020
2. Bartz-Beielstein, T., Lasarczyk, C., Preuss, M.: Sequential parameter optimization. In: Proceedings Congress on Evolutionary Computation 2005 (CEC 2005), Edinburgh, Scotland, p. 1553 (2005). http://www.spotseven.de/wp-content/papercite-data/pdf/blp05.pdf

3. Carvalho, C.M., Polson, N.G., Scott, J.G.: The horseshoe estimator for sparse signals. Biometrika **97**(2), 465–480 (2010). https://doi.org/10.1093/biomet/asq017
4. Ehrgott, M.: Multicriteria Optimization, 2nd edn. Springer, Heidelberg (2005). https://doi.org/10.1007/3-540-27659-9
5. Eiben, A.E., Smith, J.E.: Introduction to Evolutionary Computing. NCS. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-44874-8
6. Goodrich, B., Gabry, J., Ali, I., Brilleman, S.: rstanarm: Bayesian applied regression modeling via Stan (2018). http://mc-stan.org/. r package version 2.17.4
7. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. SSS. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
8. Hoang, T.T., Cho, M.Y., Alam, M.N., Vu, Q.T.: A novel differential particle swarm optimization for parameter selection of support vector machines for monitoring metal-oxide surge arrester conditions. Swarm Evol. Comput. **38**, 120–126 (2018). https://doi.org/10.1016/j.swevo.2017.07.006, http://www.sciencedirect.com/science/article/pii/S2210650217303942
9. Huang, H.L., Chang, F.L.: ESVM: evolutionary support vector machine for automatic feature selection and classification of microarray data. Biosystems **90**(2), 516–528 (2007). https://doi.org/10.1016/j.biosystems.2006.12.003, http://www.sciencedirect.com/science/article/pii/S0303264706002875
10. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. STS, vol. 103. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-7138-7
11. Lagani, V., Athineou, G., Farcomeni, A., Tsagris, M., Tsamardinos, I.: Feature selection with the R Package MXM: discovering statistically-equivalent feature subsets. arXiv e-prints arXiv:1611.03227, November 2016
12. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F.: e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien (2018). https://CRAN.R-project.org/package=e1071. r package version 1.7-0
13. Nott, D.J., Leng, C.: Bayesian projection approaches to variable selection in generalized linear models. Comput. Stat. Data Anal. **54**(12), 3227–3241 (2010). https://doi.org/10.1016/j.csda.2010.01.036
14. Perolini, A.: A fast approximated evolutionary approach to improve SVM accuracy. In: Bramer, M., Petridis, M., Hopgood, A. (eds.) Research and Development in Intelligent Systems XXVII SGAI 2010, pp. 193–206. Springer, London (2011). https://doi.org/10.1007/978-0-85729-130-1_14
15. Piironen, J., Paasiniemi, M., Vehtari, A.: projpred: Projection Predictive Feature Selection (2018). https://CRAN.R-project.org/package=projpred. r package version 1.1.0
16. Piironen, J., Vehtari, A.: Comparison of Bayesian predictive methods for model selection. Stat. Comput. **27**(3), 711–735 (2017). https://doi.org/10.1007/s11222-016-9649-y
17. Rebolledo Coy, M.A., Krey, S., Bartz-Beielstein, T., Flasch, O., Fischbach, A., Stork, J.: Modeling and optimization of a robust gas sensor. In: Papa, G., Mernik, M. (eds.) Bioinspired Optimization Methods and their Applications, pp. 267–278 (2016)
18. Stoean, C.: In search of the optimal set of indicators when classifying histopathological images. In: 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 449–455 (2016). https://doi.org/10.1109/SYNASC.2016.074

19. Tran, M.N., Nott, D.J., Leng, C.: The predictive lasso. Stat. Comput. **22**(5), 1069–1084 (2012). https://doi.org/10.1007/s11222-011-9279-3
20. Tsagris, M., Lagani, V., Tsamardinos, I.: Feature selection for high-dimensional temporal data. BMC Bioinform. **19**(1), 17 (2018). https://doi.org/10.1186/s12859-018-2023-7
21. Tsamardinos, I., Aliferis, C.: Towards principled feature selection: relevancy, filters and wrappers. In: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. Morgan Kaufmann Publishers (2003)
22. Willighagen, E., Ballings, M.: genalg: R Based Genetic Algorithm (2015). https://CRAN.R-project.org/package=genalg. r package version 0.2.0