**Machine Learning manuscript No.**
(will be inserted by the editor)

# Supplementary Material for the paper: Instance Spaces for Machine Learning Classification

**Laura Villanova · Mario A. Muñoz · Davaatseren Baatar · Kate Smith-Miles**

Received: date / Accepted: date

## 1 Problem instances

The problem instances described in section 3.1 are listed below along with their source repository (UCI, KEEL, or DCol). The list comprises 209 datasets, 27 of which are investigated twice because of the presence of missing values (i.e. features are calculated on the dataset with missing values; features are calculated on the dataset with estimated missing values). These 27 datasets are highlighted with an $^*$. The data files (ARFF format) can be downloaded at the following link `http://users.monash.edu.au/~ksmiles/matilda/classification.zip`. The link also includes the code used to calculate the features in R.

K. Smith-Miles
School of Mathematical Sciences
Monash University, Australia
Tel.: +61-3-9905-3170
Fax: +61-3-9905 4403
E-mail: kate.smith-miles@monash.edu

S. Author
second address

| Problem instance | Repository |
|---|---|
| Abalone (28 classes) | UCI |
| Abalone (3 classes) | UCI |
| Appendicitis | KEEL |
| Asbestos | DCol |
| Audiology standardized | UCI |
| Australian | KEEL |
| Auto Univ 1 | UCI |
| Auto Univ 2 | UCI |
| Auto Univ 4 | UCI |
| Auto Univ 5 | UCI |
| Auto Univ 6.1 | UCI |
| Auto Univ 6.2 | UCI |
| Auto Univ 6.3 | UCI |
| Auto Univ 7.1 | UCI |
| Auto Univ 7.2 | UCI |
| Auto Univ 7.3 | UCI |
| Auto Univ 8 | UCI |
| Automobile | KEEL |
| Bach coral harmony | UCI |
| Balance scale | UCI |
| Banana | KEEL |
| Banknote | UCI |
| Bankruptcy | KEEL |
| Blogger | UCI |
| Blood transfusion | UCI |
| Breast | KEEL |
| Breast cancer | UCI |
| Breast cancer Wisconsin (diagnostic)* | UCI |
| Breast cancer Wisconsin (original) | UCI |
| Breast cancer Wisconsin (prognostic) | UCI |
| Breast cancer Wisconsin (prognostic 2) | UCI |
| Breast tissue | UCI |
| Breast tissue merged | UCI |
| Bupa | KEEL |
| Car evaluation | UCI |
| Cardio 10 | UCI |
| Cardio 3 | UCI |
| Chess (king-rook vs. king-pawn) | UCI |
| Chronic kidney disease | UCI |
| Chronic kidney disease (full)* | UCI |
| Cilinder bands* | UCI |
| Climate model simulation | UCI |
| CNAE-9 | UCI |
| Coil2000 | KEEL |
| Congressional voting records * | UCI |
| Connectionist bench (sonar, mines vs rocks) | UCI |
| Connectionist bench (vowel recognition) | UCI |
| Contraceptive method choice | UCI |
| Credit approval* | UCI |
| Crx | KEEL |
| Dermatology | UCI |
| Diabetic retinopathy | UCI |
| Dresses attribute sales | UCI |
| Echocardiogram* | UCI |
| Ecoli | UCI |
| Fertility | UCI |
| Firm teacher | UCI |
| First order theorem | UCI |
| Flare | KEEL |
| Forest type | UCI |
| Fourclass | DCol |
| Gesture phase segmentation (raw) | UCI |
| Gesture phase segmentation (va) | UCI |
| Glass identification | UCI |
| Grammatical Facial Expression (a1) | UCI |
| Grammatical Facial Expression (a2) | UCI |
| Grammatical Facial Expression (a3) | UCI |
| Grammatical Facial Expression (a4) | UCI |
| Grammatical Facial Expression (a5) | UCI |
| Grammatical Facial Expression (a6) | UCI |
| Grammatical Facial Expression (a7) | UCI |
| Grammatical Facial Expression (a8) | UCI |
| Grammatical Facial Expression (a9) | UCI |
| Grammatical Facial Expression (b1) | UCI |
| Grammatical Facial Expression (b2) | UCI |
| Grammatical Facial Expression (b3) | UCI |
| Grammatical Facial Expression (b4) | UCI |
| Grammatical Facial Expression (b5) | UCI |
| Grammatical Facial Expression (b6) | UCI |
| Grammatical Facial Expression (b7) | UCI |
| Grammatical Facial Expression (b8) | UCI |
| Grammatical Facial Expression (b9) | UCI |
| Haberman's survival | UCI |
| Hayes roth | UCI |
| Heart disease (cleverland)* | UCI |
| Heart disease (hungarian)* | UCI |
| Heart disease (switzerland)* | UCI |
| Heart disease (va)* | UCI |
| Hepatitis* | UCI |

| Problem instance | Repository |
|---|---|
| Hill-Valley (with noise) | UCI |
| Hill-Valley (without noise) | UCI |
| HIV-1 protease cleavage (1625) | UCI |
| HIV-1 protease cleavage (746) | UCI |
| HIV-1 protease cleavage (impens) | UCI |
| HIV-1 protease cleavage (schilling) | UCI |
| Horse colic (outcome) | UCI |
| Horse colic (lesion) | UCI |
| Human activity recognition using smartphones | UCI |
| ILPD (Indian Liver Patient Dataset)* | UCI |
| Image segmentation | UCI |
| Internet advertisements | UCI |
| Ionosphere | UCI |
| Iris | UCI |
| Isolet | UCI |
| Japanese credit screening* | UCI |
| Leaf | UCI |
| LED7digit | KEEL |
| Lenses | UCI |
| Libras movement | UCI |
| Liver disorders | DCol |
| LSVT voice rehabilitation | UCI |
| Lung cancer* | UCI |
| Lymphography | KEEL |
| Madelon | UCI |
| Mammographic mass* | UCI |
| Mechanical analysis | UCI |
| Mice protein expression* | UCI |
| Molecular biology (promoter gene sequences) | UCI |
| Molecular biology (splice-junction gene sequences) | UCI |
| Monk's problems (1) | UCI |
| Monk's problems (2) | UCI |
| Monk's problems (3) | UCI |
| Multiple features | UCI |
| Mushroom | UCI |
| One-hundred plant species leaves (mar) | UCI |
| One-hundred plant species leaves (sha) | UCI |
| One-hundred plant species leaves (tex) | UCI |
| Optical recognition of handwritten digits | UCI |
| Ozone level detection (1 hour) * | UCI |
| Ozone level detection (8 hours) * | UCI |
| Page blocks classifiation | UCI |
| PAMAP2 physical activity | UCI |
| Parkinson speech | UCI |
| Parkinsons | UCI |
| Pen-based recognition of handwritten digits | UCI |

| Problem instance | Repository |
|---|---|
| Phishing websites | UCI |
| Phoneme | KEEL |
| Pima indians diabetes | UCI |
| Pittsburgh bridges 1.1 | UCI |
| Pittsburgh bridges 1.2 | UCI |
| Pittsburgh bridges 1.3 | UCI |
| Pittsburgh bridges 1.4 | UCI |
| Pittsburgh bridges 2.1 | UCI |
| Pittsburgh bridges 2.2 | UCI |
| Pittsburgh bridges 2.3 | UCI |
| Pittsburgh bridges 2.4 | UCI |
| Planning relax | UCI |
| Post-operative patient | UCI |
| Primary tumor | UCI |
| QSAR biodegradation | UCI |
| Qualitative bankruptcy | UCI |
| Ring | KEEL |
| Robot execution failures (lp1) | UCI |
| Robot execution failures (lp2) | UCI |
| Robot execution failures (lp3) | UCI |
| Robot execution failures (lp4) | UCI |
| Robot execution failures (lp5) | UCI |
| Saheart | KEEL |
| Satimage | KEEL |
| SECOM | UCI |
| Seeds | UCI |
| Seismic-bumps | UCI |
| Soybean (large)* | UCI |
| Soybean (small) | UCI |
| Spambase | UCI |
| SPECT heart | UCI |
| SPECTF heart | UCI |
| Statlog (australian credit approval) | UCI |
| Statlog (german credit data) | UCI |
| Statlog (heart) | UCI |
| Statlog (image segmentation) | UCI |
| Statlog (land satellite) | UCI |
| Statlog (vehicle silhouettes) | UCI |
| Tae | KEEL |
| Tao | DCol |
| Teaching assistant evaluation | UCI |
| Texture | KEEL |
| Thyroid disease (allbp)* | UCI |
| Thyroid disease (allhyper) | UCI |
| Thyroid disease (allhypo)* | UCI |
| Thyroid disease (allrep)* | UCI |
| Thyroid disease (ann)* | UCI |
| Thyroid disease (dis)* | UCI |
| Thyroid disease (hypothyroid) | UCI |

| Problem instance | Repository |
|---|---|
| Thyroid disease (new thyroid)* | UCI |
| Thyroid disease (sick)* | UCI |
| Tic-tac-toe | UCI |
| Titanic | KEEL |
| Trains | UCI |
| Turkiye student evaluation | UCI |
| Twonorm | KEEL |
| Urban land cover | UCI |
| User knowledge modeling | UCI |
| Vertebral column (2) | UCI |
| Vertebral column (3) | UCI |
| Wall-following robot navigation data (2 sensors) | UCI |
| Wall-following robot navigation data (4 sensors) | UCI |
| Wall-following robot navigation data (24 sensors) | UCI |
| Waveform database generator (version 1) | UCI |
| Waveform database generator (version 2) | UCI |
| Weight lifting exercises | UCI |
| Wholesale customer | UCI |
| Wilt | UCI |
| Wine | UCI |
| Wine quality | UCI |
| Winsconsin | UCI |
| Wpbc* | DCol |
| Yeast | UCI |
| Zoo | UCI |

## 2 Altered datasets for feature analysis

The instance alterations described in Section 4.1 are described below.

| | Comparison | |
|---|---|---|
| Challenge | Problem instance 1 | Problem instance 2 |
| Non-normality within classes | original | for each class $k = 1, \cdots, K$, the within class data are simulated from a multivariate normal distribution $N(\mu_k, \Sigma_k)$, where $\mu_k$ and $\Sigma_k$ is as from the original data |
| Unequal covariance within classes | for each class $k = 1, \cdots, K$, the within class data are simulated from a multivariate normal distribution $N(\mu_k, \Sigma_k)$, where $\mu_k$ and $\Sigma_k$ are as from the original data | for each class $k = 1, \cdots, K$, the within class data are simulated from a multivariate normal distribution $N(\mu_k, \Sigma)$, where $\mu_k$ is as from the original data and $\Sigma$ is the same for all of the classes |
| Redundant attributes | original | select attribute $x_i$, $i = 1, \cdots, q$ and add to it a normally distributed random error $\epsilon \sim N(0, \sigma \cdot 0.01)$ where $\sigma$ is the standard deviation of the original attribute $x_i$; finally, include the derived attribute in the original dataset |
| Type of attributes | original | discretize one continuous attribute and substitute the original attribute with the discretized one |
| Unbalanced classes | original | randomly remove instances from larger classes in order to obtain classes with the same number of instances |
| Constant attribute within classes | original | for a given attribute assign a constant value to the instances belonging to the same class |
| (Nearly) Linearly dependent attribute | original | given two continuous attributes from the original data, obtain an additional attribute as a linear combination of the two |
| Non-linearly separable classes | original | draw a line or hyperplane that approximately separates the original classes; change class labels to assign all instances on either sides of the line/hyperplane to the same class |
| Missing values | original | for each attribute, randomly select 25% of the original data values and consider them as missing values |
| Data scaling | original | each attribute is standardized to have mean equal to 0 and standard deviation equal to 1 |
| Redundant instances | original | double the size of the original data by replicating all of the instances |
| Lack of information | original | randomly select 50% of the instances and remove them from the original dataset |

## 3 Selected features and their correlation matrix

The set of 10 features selected in Section 4.1 are described below. The correlation matrix for features and algorithms performance is also reported.

- $H(\mathbf{X})'_{\max}$: *maximum normalized entropy of the attributes*. Given nominal attributes $X_i$, $i = 1, \ldots, m$, taking on values $\{x_j\}$, $j = 1, \ldots, m_i$, let $p_j^i = P(X_i = x_j)$. The entropy of $X_i$ is a measure of randomness in the attribute. It is maximal when all $p_j^i$ are equal, and its maximal value is $\log_2 m_i$. It is minimal when $X_i$ does not vary, and its minimal value is 0 [5]. An attribute with low entropy is likely to contains little information to discriminate between classes; conversely, an attribute with high entropy is likely to contain much information to discriminate between classes. The entropy of $X_i$ is defined as $H(X_i) = \sum_j p_j^i \log_2(p_j^i)$ and can be normalized in $[0, 1]$ using $H(X_i)' = -H(X_i)/\log_2 m_j$. The maximum of the normalized entropies

$$H(\mathbf{X})'_{\max} = \max_i H(X_i)' \tag{1}$$

  quantifies the highest amount of information contained in the data assuming independent attributes.
- $H'_c$: *normalized entropy of class attribute*. Given the class attribute $C$ with $K$ classes, let $q_k = P(C = k)$, $k = 1, \cdots, K$. The normalized class entropy

$$H'_c = -\sum_k q_k \log_2(q_k)/\log_2 K. \tag{2}$$

  is maximal when all the classes are equally probable; therefore, it is a measure of problem imbalance.
- $\overline{M}_{CX}$: *mean mutual information of attributes and class*. The mutual information between attribute $X_i$ and class $C$ is a measure of their shared information. It is minimal when $X_i$

Table 1: Correlation matrix for the set of relevant features and the investigated algorithms.

| | $H(X)'_{\max}$ | $H'_C$ | $\overline{M}_{CX}$ | $DN_{ER}$ | $SD(\nu)$ | $F3$ | $F4$ | $L2$ | $N1$ | $N4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $H(X)'_{\max}$ | 1.00 | 0.21 | 0.13 | 0.28 | -0.54 | 0.10 | 0.12 | 0.21 | 0.27 | -0.09 |
| $H'_C$ | 0.21 | 1.00 | 0.67 | 0.13 | -0.30 | 0.42 | 0.27 | -0.11 | -0.14 | -0.23 |
| $\overline{M}_{CX}$ | 0.13 | 0.67 | 1.00 | -0.16 | -0.07 | 0.58 | 0.38 | -0.15 | -0.44 | -0.36 |
| $DN_{ER}$ | 0.28 | 0.13 | -0.16 | 1.00 | -0.36 | -0.38 | -0.29 | 0.56 | 0.67 | 0.31 |
| $SD(\nu)$ | -0.54 | -0.30 | -0.07 | -0.36 | 1.00 | 0.12 | 0.07 | -0.35 | -0.44 | 0.09 |
| $F3$ | 0.10 | 0.42 | 0.58 | -0.38 | 0.12 | 1.00 | 0.70 | -0.29 | -0.57 | -0.41 |
| $F4$ | 0.12 | 0.27 | 0.38 | -0.29 | 0.07 | 0.70 | 1.00 | -0.28 | -0.44 | -0.47 |
| $L2$ | 0.21 | -0.11 | -0.15 | 0.56 | -0.35 | -0.29 | -0.28 | 1.00 | 0.64 | 0.32 |
| $N1$ | 0.27 | -0.14 | -0.44 | 0.67 | -0.44 | -0.57 | -0.44 | 0.64 | 1.00 | 0.41 |
| $N4$ | -0.09 | -0.23 | -0.36 | 0.31 | 0.09 | -0.41 | -0.47 | 0.32 | 0.41 | 1.00 |
| NB | 0.19 | -0.20 | **-0.49** | **0.57** | -0.25 | **-0.50** | -0.24 | **0.53** | **0.70** | **0.40** |
| LDA | 0.22 | **0.38** | -0.01 | **0.41** | -0.26 | 0.01 | -0.01 | 0.21 | 0.28 | -0.01 |
| QDA | 0.13 | **0.70** | 0.29 | 0.22 | -0.23 | 0.16 | 0.08 | -0.11 | 0.03 | -0.10 |
| CART | **0.34** | 0.00 | **-0.36** | **0.75** | **-0.38** | **-0.49** | **-0.36** | **0.59** | **0.83** | **0.37** |
| J48 | 0.26 | -0.05 | **-0.39** | **0.65** | **-0.38** | **-0.51** | **-0.35** | **0.62** | **0.85** | 0.39 |
| KNN | 0.24 | -0.05 | **-0.43** | **0.67** | **-0.36** | **-0.48** | **-0.37** | **0.57** | **0.90** | **0.36** |
| L-SVM | 0.28 | 0.00 | **-0.36** | **0.71** | **-0.33** | **-0.42** | **-0.32** | **0.59** | **0.74** | **0.36** |
| Poly-SVM | 0.29 | 0.05 | **-0.34** | **0.69** | **-0.43** | **-0.42** | **-0.36** | **0.53** | **0.72** | 0.28 |
| RB-SVM | **0.31** | 0.06 | **-0.35** | **0.69** | **-0.36** | **-0.38** | -0.27 | **0.56** | **0.76** | **0.34** |
| RF | 0.20 | 0.27 | -0.11 | **0.38** | -0.21 | -0.04 | -0.08 | 0.25 | **0.37** | 0.11 |

and $C$ are independent (i.e. there is no shared information), and its minimal value is 0. It is maximal when $X_i$ perfectly predicts $C$ (i.e. $X_i$ contains all the information to predict $C$), and its maximal value is $\min(H(C), H(X_i)$. Let $\pi^i_{jk}$ be the joint probability of observing class $k$ and the $j$-th value of attribute $X_i$; then, the mutual information of $C$ and $X_i$ is $M(C, X_i) = \sum_{jk} \pi^i_{jk} \log 2 \left( \frac{\pi^i_{jk}}{p^i_j q_k} \right)$ [5]. The mean mutual information of attributes and class

$$\overline{M}_{CX} = m^{-1} \sum_i M(C, X_i) \tag{3}$$

measures the average shared information between class attribute and attributes.

- $DN_{ER}$: *error rate of the decision node.* A decision node is a decision tree consisting of a single node, the root. The splitting attribute and value in the root are selected so as to maximize the information gain ratio [2]. We use 10-fold cross validation to train the decision node and test its performance (error rate). The error rate averaged over the 10 folds provides with an indication of linear separability in the data [1].
- $SD(\nu)$: *standard deviation of the weighted distance.* The weighted distance is a measure of the sparsity of the instances in a problem. Denote with $x^r = [x^r_1 \cdots x^r_m]^T$ the $r$-th instance of a problem with $m$ numeric attributes, where $r = 1, \ldots, n$ [7]. The distances between any two instances $x^r$ and $x^s$ is defined as $D(x^r, x^s) = \sqrt{\sum_{l=1}^m d(x^r_l, x^s_l)^2}$, where $d(x^r_l, x^s_l) = |x^r_l - x^s_l|/(\max x_l - \min x_l)$. The corresponding weights are defined as

$$W(x^r, x^s) = \left( 2^\alpha \cdot \frac{D(x^r, x^s)}{\sqrt{m} - D(x^r, x^s)} \right)^{-1}$$

in [0,1]. Therefore, closer instances are assigned exponentially higher weight. The parameter $\alpha$ in $W(x^r, x^s)$ controls the exponential decay of the weights as $D(x^r, x^s)$ grows larger. We set $\alpha = 2$ as recommended in [7]. For any given instance $x^r$, the weighted distance between $x^r$ and all other instances of the problem is defined as

$$\nu^r = \frac{\sum_{r=1, r \neq i}^n W(x^r, x^s) \cdot D(x^r, x^s)}{\sum_{r=1, r \neq s}^n W(x^r, x^s)}.$$

Small values of $\nu^r$ indicate that $x^r$ is close to the other instances, whereas large values of $\nu^r$ indicate that $x^r$ is further apart from the remaining instances. The standard deviation of the weighted distances $\nu^1, \ldots, \nu^n$ is derived using

$$SD(\nu) = \sqrt{\frac{\sum_{t=1}^n (\nu^t - \overline{\nu})^2}{n-1}} \tag{4}$$

where $\overline{\nu} = \sum_t \nu^t / n$. Small values of $SD(\nu)$ indicate that instances are homogeneously distributed in the feature space, whereas large values indicate non-homogeneous instances (e.g. clusters).

- $F3$: *maximum feature efficiency.* The feature efficiency measures the ability of a feature to linearly discriminate two different classes. Class discrimination is assessed using hyperplanes that are perpendicular to the feature axes. If such hyperplanes perfectly separates two classes (binary classification problem) no overlap exists between the feature values of the classes; in this case there is no ambiguity and the feature efficiency is maximal. In all other cases the classes tend to overlap and there exists ambiguity between classes. The degree of ambiguity can be measured in terms of the number of instances lying within the overlapping region. The efficiency of feature $X_i$, $F3_i$, is calculated as the fraction of instances lying outside the overlapping region. The maximum feature efficiency observed over all features

$$F3 = \max_i F3_i \tag{5}$$

indicates the degree of linear separability achievable when using the most discriminative attribute.

- $F4$: *collective feature efficiency*. It extends the concept in $F3$ to account for the discriminative ability of multiple attributes in the dataset. $F4$ measures the collective ability of the attributes to discriminate between classes when using hyperplanes that are perpendicular to the attribute's axes. It is calculated using an iterative procedure consisting in the following steps: (i) select the most discriminative attribute among the available ones, (ii) discriminate the instances using the selected attribute, (iii) remove from the dataset both attribute and instances that could be discriminated. The above procedure is repeated until either all instances are discriminated or all attributes are investigated [6]. The output is the proportion of instances that can be discriminated when using multiple independent attributes.
- $L2$: *training error of linear classifier*. Linear classification is used to assess linear separability of the classes in the training set. The linear classifier minimizes the error function

$$\begin{aligned} \min \quad & a^T \psi \\ \text{s.t.} \quad & U^T w + \psi \geq b \\ & \psi \geq 0 \end{aligned}$$

  where $a, b$ are arbitrary constant vectors (both chosen to be 1), $w$ is the weight vector, $\psi$ is an error vector, and $U$ is a matrix where each column $u$ is defined on an input vector $x$ (augmented by adding one dimension with a constant value 1) and its class $C$ (with value $c_1$ and $c_2$) [3]:

$$\begin{cases} u = +x & \text{if } C = c_1 \\ u = -x & \text{if } C = c_2 \end{cases} \tag{6}$$

- $N1$: *fraction of points on the class boundary*. $N1$ estimates the length of the class boundary and provides with a complexity measure of the boundary between classes [6]. The first step is to build a minimum spanning tree (MST). The nodes of the MST are the instances of the dataset and edges connect all the nodes disregarding the class information. When class information is superimposed to the MST, it is possible to count the number of nodes connected to nodes with different class label. The fraction of such nodes over all instances of the dataset gives $N1$. Small values of $N1$ indicate that there are only a few points along the boundary, whereas large $N1$ values indicate that the majority of points lay along the boundary. In both cases it may be difficult to accurately define the class boundary. Instead, moderate $N1$ values indicate a higher chance to accurately define the class boundary.
- $N4$ : *nonlinearity of the one-nearest neighbor classifier*. N4 estimates the nonlinearity of the class boundary using the one-nearest neighbor classifier and the concept of linear interpolation [3]. In particular, the classifier is first trained using all the instances in the problem instance. Then, $n$ instances $\mathbf{x}_j$ are randomly drawn with replacement. For each instance $\mathbf{x}_j$ a new instance $\mathbf{x}'_j$ is randomly generated, and pairs of instances $(\mathbf{x}_j, \mathbf{x}'_j)$ with the same class label are created. The linear interpolation $\alpha \cdot \mathbf{x}_k j + (1-\alpha) \cdot \mathbf{x}'_j$, with $0 \leq \alpha \leq 1$ is used to create a third object $\mathbf{x}''_j$. The class of $\mathbf{x}''_j$ is predicted using the previously trained classifier and compared to the class of $\mathbf{x}_j$ (or $\mathbf{x}'_j$). A discrepancy between class labels counts as one contribution to the nonlinearity. N4 is derived as the ratio between the number of nonlinearity contributions and $n$ [4].

## References

1. Bensusan, H., Giraud-Carrier, C.: Casa batló is in passeig de gràcia or how landmark performances can describe tasks. In: Proceedings of the ECML-00 workshop on meta-learning: Building automatic advice strategies for model selection and method combination, pp. 29–46 (2000)
2. Bensusan, H., Giraud-Carrier, C.: Discovering task neighbourhoods through landmark learning performances. In: Principles of Data Mining and Knowledge Discovery, pp. 325–330. Springer (2000)
3. Ho, T.K., Basu, M.: Complexity measures of supervised classification problems. Pattern Analysis and Machine Intelligence, IEEE Transactions on **24**(3), 289–300 (2002)

4. Hoekstra, A., Duin, R.P.: On the nonlinearity of pattern classifiers. In: Pattern Recognition, 1996., Proceedings of the 13th International Conference on, vol. 4, pp. 271–275. IEEE (1996)
5. Michie, D., Spiegelhalter, D.J., Taylor, C.C.: Machine learning, neural and statistical classification (1994)
6. Orriols-Puig, A., Macia, N., Ho, T.K.: Documentation for the data complexity library in c++. Universitat Ramon Llull, La Salle **196** (2010)
7. Vilalta, R.: Understanding accuracy performance through concept characterization and algorithm analysis. In: Proceedings of the ICML-99 Workshop on Recent Advances in Meta-Learning and Future Work, pp. 3–9 (1999)