

Homogeneity hypothesis in discriminant analysis

Dmitriy Klyushin^a

^a*Taras Shevchenko National University of Kyiv, Ukraine, 03680, Kyiv, Akademika Glushkova Avenue 4D*

Abstract

One of the most important properties of a machine learning algorithm is its ability to generalize results of learning on finite training sets. This property is based on the compactness hypothesis stating that objects of the same class in the feature space, as a rule, are located closer to each other than to objects of other classes. The compactness hypothesis has a geometric nature and uses the concept of proximity in the feature space, which is most often expressed in terms of a metric. Meanwhile, this hypothesis does not fully take into account the probabilistic nature of the features. It is quite suitable for data with unimodal distributions that have a compact support, but in a general case it may not hold leading to incorrect generalizations. In the paper, an alternative approach is described in which the homogeneity hypothesis is used instead of the compactness hypothesis. Within the framework of this approach, objects are called homogeneous, if their features follow identical distributions. We propose as measures of homogeneity the Petunin's p -statistics and its versions, which is highly efficient in recognizing both disjoint and significantly overlapping samples that violate the compactness assumption. This approach has a rigorous mathematical foundation and high efficiency in practical applications.

Keywords ¹

Discriminant analysis, relational analysis, featureless pattern recognition, compactness hypothesis, homogeneity hypothesis

1. Introduction

The complexity of machine learning significantly depends on the compactness hypothesis, which allows generalizations based on finite training samples. Intuitively, the hypothesis states that in the feature space, similar objects should be closer to each other than to dissimilar ones. This definition appeals to the geometric concept of proximity and implicitly uses a metric. A typical example of a method based on such principles is the nearest neighbor method, which recognizes test objects by their closeness to training objects.

The compactness hypothesis ignores the probabilistic nature of the random training data. More precisely, it is only acceptable for classifying random data with unimodal distributions having compact support. In practice, such a condition is too burdensome. Therefore, it is necessary to develop a method that would estimate the proximity between random samples on different principles. To solve the problem, let us introduce the concept of object homogeneity, which means that objects are drawn from the same population, i.e. their features obey the same distribution. This allows objects to be classified using criteria to test static hypotheses of homogeneity.

The aim of the article is to describe a new approach to machine learning based on the homogeneity hypothesis as an alternative to the compactness hypothesis. Using a measure of homogeneity, not just a metric, allows generalizing relational discriminant analysis and increasing its efficiency.

CMIS-2021: The Fourth International Workshop on Computer Modeling and Intelligent Systems, April 27, 2021, Zaporizhzhia, Ukraine

EMAIL: dokmed5@gmail.com (D.A. Klyushin)

ORCID: 0000-0003-4554-1049 (D.A. Klyushin)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

2. Distance-based machine learning techniques

Duin, Pekalska et al. [1–4], Mottle, Seredin et al. [5–8] and others proposed the concept of featureless discriminant analysis. They suggested replacing the feature vector of an object with an estimate of its proximity to some training set using a metrics. Unfortunately, this approach is poorly suited to solving problems often encountered in biomedical research. Let us say a researcher studies the parameters of a set of cells. In this case, it gets samples of real numbers, not an ordered vector. In such cases, the metric is not applicable and the only useful tool is a homogeneity (similarity) measure. Among the numerous statistical tests for the homogeneity of samples, only the Kolmogorov-Smirnov test, the Wilcoxon test and the Klyushin-Petunin test allow us to assess robustly the homogeneity of samples in the form of the probability of belonging to the same general population [9].

As it was noted above, in the featureless discriminant analysis, objects are represented not as vectors in a feature space, but as a measure of proximity to a training set. As a result, the starting point of featureless analysis is a distance matrix filled with distances or labels characterizing the similarity between objects in the training set (reference points). As a basic distance, usually Euclidean and pseudo-Euclidean distances are used coupled with the kernel trick. Obviously, this approach leads to problems with generalization power and strong dependence from a training set. In addition, it is not valid for samples containing of independent identically distributed (i.i.d.) random values.

Recently, this approach was renewed as machine learning techniques such as Minimal Learning Machine [10] and the Extreme Minimal Learning Machine [11]. The main tool in these methods is the nonlinear distance regression, which estimates the dissimilarity between observations. Nowadays, various metrics and learning techniques are used in this field [12–21]. Excellent surveys of these methods may be found in [22, 23]. These methods have some useful advantages, but they use Euclidean distance-constructed probability distributions. Thus, they fail in situations when reference points are not vectors in some vector space but samples of i.i.d. random values.

The problem we try to solve is to extend the application field of the distance-based machine learning techniques using not metrics to estimate distances between objects but the homogeneity (similarity) measures described below, and propose an alternative way of similarity-based classification. These homogeneity measures do not dependent on underlying distributions of training samples and have useful properties of generalization. For example, in opposite to standard counterparts (the Kolmogorov–Smirnov statistics and the Wilcoxon statistics), they successfully work both with samples following distributions with different means and identical variance and with identical means and different variances.

3. Two-sample homogeneity measure

Consider training samples $x = (x_1, x_2, \dots, x_n) \in G_1$ and $y = (y_1, y_2, \dots, y_n) \in G_2$ from populations G_1 and G_2 following absolutely continuous distribution functions F_1 and F_2 . We reduce the classification of a test sample $z = (z_1, z_2, \dots, z_n)$ to testing of homogeneity z and x from the one side, and z and y from the other side. There are many nonparametric tests for two samples homogeneity: Kolmogorov–Smirnov test, Wilcoxon sign rank test etc. (see, for example, [24]). However, as it will be shown, the most effective tool for testing homogeneity of two samples is the Petunin's p -statistics [25]. This is explained by the fact that the p -statistics has similar high significance and sensitivity independently in both cases when samples are disjoint or almost overlapped.

3.1. Original Klyushin–Petunin test

The Klyushin–Petunin test [25] is non-parametric one and use only assumption that distribution functions are absolutely continuous. This test uses the Hill's assumption $A_{(n)}$ [26] stating that for

exchangeable random values $x_1, x_2, \dots, x_n \in G$ following to an absolutely continuous distribution function we have:

$$P\left(x \in (x_{(i)}, x_{(j)})\right) = \frac{j-i}{n+1}, \quad j < i, \quad (1)$$

where $x_{(i)}$ and $x_{(j)}$ are the i -th and j -th order statistics. The Hill's assumption was proved as for i.i.d. random values [27] and for exchangeable i.d. random values [28]. Finding the relative frequency h_{ij} of the event $z_m \in (x_{(i)}, x_{(j)})$ for the elements of z , we can estimate a proximity between h_{ij} and $\frac{j-i}{n+1}$. This may be made using numerous confidence intervals for binomial proportion. For definiteness, we use the Wilson confidence interval $I_{ij}^{(n)} = (p_{ij}^{(1)}, p_{ij}^{(2)})$ where

$$p_{ij}^{(1)} = \frac{h_{ij}n + g^2/2 - g\sqrt{h_{ij}(1-h_{ij})n + g^2/4}}{n + g^2}, \quad (2)$$

$$p_{ij}^{(2)} = \frac{h_{ij}n + g^2/2 + g\sqrt{h_{ij}(1-h_{ij})n + g^2/4}}{n + g^2}.$$

The significance level of this interval is the function of g . When $g = 3$ the significance level of $I_{ij}^{(n)}$ does not exceed 0.05 [25]. P -statistics, estimating the homogeneity of samples x and z , is defined by the equation

$$h = \# \left\{ p_{ij} = \frac{j-i}{n+1} \in I_{ij}^{(n)} \right\} / \left(\frac{n(n-1)}{2} \right), \quad (3)$$

As we see, the p -statistics is the estimation of the probability that the samples are homogeneous, therefore, using (2) we may formulate the following test: the null hypothesis is accepted if h is greater than 0.95, else it is rejected.

When the null hypothesis is true, the events $\left\{ p_{ij} = \frac{j-i}{n+1} \in I_{ij}^{(n)} \right\}$ generate a generalized Bernoulli scheme [29, 30]. When the alternative hypothesis is true, they generate a modified Bernoulli scheme. When the null hypothesis can be either true or false, they generate the Matveichuk–Petunin scheme [31]. When the null hypothesis holds, $\lim_{n \rightarrow \infty} \frac{j-i}{n+1} \in (0,1)$, and $\lim_{n \rightarrow \infty} \frac{i}{n+1} \in (0,1)$, then the asymptotic significance level β of a sequence of confidence intervals $I_{ij}^{(n)}$ is less than 0.05 [25].

3.2. Modified Klyushin–Petunin test

In practice, samples, as a rule, contain rounded numbers and duplicates (ties). Thus, we must distinguish a hypothetical sample drawn from a hypothetical population G of precise measurements and an empirical sample drawn from an empirical population \tilde{G} of rounded measurements. Let us introduce a sample $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ approximating a hypothetical sample $x = (x_1, x_2, \dots, x_n)$ and let the variational series $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ and $\tilde{x}_{(1)} < \tilde{x}_{(2)} < \dots < \tilde{x}_{(m)}$ be variational series of hypothetical and empirical samples.

If a number x^* is drawn from G independently from x then

$$p\left(x^* \in [x_{(k)}, x_{(k+1)}]\right) = \frac{1}{n+1}, \quad (4)$$

where $k = 0, 1, \dots, n$, $x_{(0)} = -\infty$, and $x_{(n+1)} = \infty$. Thus,

$$p\left(\tilde{x}^* \in [\tilde{x}_{(i)}, \tilde{x}_{(j)}]\right) \approx \gamma_i + \gamma_{i+1} + \dots + \gamma_{j-1} + \frac{j-i}{n+1}, \quad (5)$$

where $t_i = t(\tilde{x}_{(i)})$ is the multiplicity of $\tilde{x}_{(i)}$. If \tilde{x} does not contain ties then $\gamma_i = 0$.

Suppose, that the hypothetical population G follows a hypothetical absolutely continuous distribution function F . Then, (4) holds. Consider empirical samples $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$ and $\tilde{z} = (\tilde{z}_1, \dots, \tilde{z}_n)$. Using the Wilson confidence interval $I_{ij} = (p_{ij}^{(1)}, p_{ij}^{(2)})$ for the probability (5) of the event $\{\tilde{z}_k \in (\tilde{x}_{(i)}, \tilde{x}_{(j)})\}$ we find an observed relative frequency. Let us denote $N = \#I_{ij} = \frac{n(n-1)}{2}$ and compute the empirical p -statistics $h = \frac{1}{N} \#\left\{\frac{j-i}{n+1} \in I_{ij}\right\}$. Then, we can formulate the following test: the null hypothesis is accepted if the h (the probability that the samples are homogeneous) is greater than 0.95, else the null hypothesis is rejected.

3.3. Exact Klyushin–Petunin test

As we see, the versions of the Klyushin–Petunin test based on the Wald confidence interval depend on the parameter g , that varies from 1.96 for the normal distribution to 3 for an general unimodal distribution. To avoid this uncertainty, we propose to use the exact confidence interval for the unknown probability p on the basis of the proportion h in the Bernoulli model consisting of n trials [32]. To do this, consider two functions depending on $p \in [0, 1]$:

$$\varphi(p) = |h - p|$$

and

$$\psi(p) = \frac{1}{2n} + \frac{\lambda}{n} \sqrt{np(1-p) + \frac{1}{12}}.$$

where $\lambda > \sqrt{\frac{8}{3}}$ is the parameter of the Vysochansky–Petunin inequality [33]

$$p\left(|y - m(y)| \geq \lambda \sigma(y)\right) \leq \frac{4}{9} \frac{1}{\lambda^2}.$$

Denote

$$\tilde{\psi}(p) = \sqrt{np(1-p) + \frac{1}{12}}, \quad p \in R^1.$$

The graph of $\tilde{\psi}(p)$, $p \in R^1$ is the upper half of the ellipse E passing through the points

$$A = \left(\frac{1}{2n} \left(n + \sqrt{\frac{n}{3} + n^2} \right), 0 \right), \quad B = \left(\frac{1}{2}, \sqrt{\frac{1}{12n} + \frac{1}{4}} \right),$$

$$C = \left(\frac{1}{2n} \left(n - \sqrt{\frac{n}{3} + n^2} \right), 0 \right), \quad D = \left(\frac{1}{2}, -\sqrt{\frac{1}{12n} + \frac{1}{4}} \right)$$

with the center $\left(\frac{1}{2}, 0\right)$. The graph of $\psi(p)$ is the restriction of the graph of $\tilde{\psi}(p)$ on the segment $[0, 1]$ stretching or shrinking the graph by $\frac{\lambda}{n}$ and shifting it by $\frac{1}{2n}$.

Therefore, the graph of the function $\psi(p)$ which does not depend on h is an arc of ellipse Γ passing through the points $(0, \psi(0))$, $\left(\frac{1}{2}, \psi\left(\frac{1}{2}\right)\right)$, $(1, \psi(1))$, such that the function $\psi(p)$ reach the minimum at the point $p = \frac{1}{2}$ and it is symmetrical with respect to this point.

The lower confidence limit p_1 is a root of the quadratic equation

$$\left(1 + \frac{\lambda^2}{n}\right)p^2 - \left(\frac{\lambda^2}{n} - \frac{1}{n} + 2h\right)p + h^2 - \frac{h}{n} + \frac{1}{4n^2}\left(1 - \frac{\lambda^2}{3}\right) = 0. \quad (6)$$

If $h > \psi(0) = \frac{1}{2n} + \frac{\lambda}{n\sqrt{12}}$, then the lower confidence limit p_1 is the least root of (6). If $h \leq \psi(0)$, then $p_1 = 0$.

Similarly, the upper confidence limit p_2 is a root of the square equation

$$\left(1 + \frac{\lambda^2}{n}\right)p^2 - \left(\frac{\lambda^2}{n} + \frac{1}{n} + 2h\right)p + h^2 + \frac{h}{n} + \frac{1}{4n^2}\left(1 - \frac{\lambda^2}{3}\right) = 0. \quad (7)$$

If $1 - h > \psi(1)$, then the upper confidence limit p_2 is the largest root of (4). If $1 - h \leq \psi(1)$, then $p_2 = 1$.

Remark. Since $p_1 \leq h \leq p_2$, the proportion of successes always is in the confidence interval $[p_1, p_2]$.

For the generalized Bernoulli model similar reasoning gives the following quadratic equation for lower confidence limit:

$$\left(1 + \frac{(m+n+1)\lambda^2}{(n+2)m}\right)p^2 + \left(\frac{1}{m} - \frac{(m+n+1)\lambda^2}{(n+2)m} - 2h\right)p + h^2 - \frac{h}{m} + \frac{1}{4m^2}\left(1 - \frac{\lambda^2}{3}\right) = 0 \quad (8)$$

If $h > \frac{1}{2m} + \frac{\lambda}{m\sqrt{12}} = \gamma$, then the lower confidence limit p_1 for the generalized Bernoulli model is the least root of (8). If $h \leq \gamma$, then $p_1 = 0$.

Similar, the upper confidence limit p_2 for the generalized Bernoulli model is the root of the equation

$$\left(1 + \frac{(m+n+1)\lambda^2}{(n+2)m}\right)p^2 - \left(\frac{1}{m} + \frac{(m+n+1)\lambda^2}{(n+2)m} + 2h\right)p + h^2 + \frac{h}{m} + \frac{1}{4m^2}\left(1 - \frac{\lambda^2}{3}\right) = 0 \quad (9)$$

If $1 - h > \gamma$, then the upper confidence limit p_2 is the largest root of (9). If $1 - h \leq \gamma$, then $p_2 = 1$. By virtue of the previous results the significance level of the confidence interval does not exceed $\frac{4}{9} \frac{1}{\lambda^2}$ (in particular, 0.05 for $\lambda = 3$).

4. Experiments and results

To assess the true positive and true negative rates of the proposed tests, we performed numerical experiments using samples from the normal distribution $N(\mu, \sigma)$ of various degree of overlapping. We considered 100 samples of 40 random numbers having different averages and the same variance (location shift) and as well as 100 samples of 40 random numbers having the same average value and different variance (scale shift). We calculated the average p -statistics and its lower and upper confidence limits, the average Kolmogorov–Smirnov statistics and its p -value, and the average Wilcoxon statistics and its p -value. To estimate the true positive rate of the Klyushin–Petunin test we used the relative frequency of an event when the p -statistic is less than 0.95 for different distributions. The true positive rate of the Kolmogorov–Smirnov and Wilcoxon sign rank tests is the relative frequency of an event when the corresponding p -value is less than 0.05, when the distributions are different. The true negative rate of the Klyushin–Petunin test is the relative frequency of an event when the upper confidence limit of the p -statistic is greater than 0.95 for identical distributions. The true negative rate of the Kolmogorov–Smirnov tests and Wilcoxon signed ranks is the relative frequency of an event when the value of p is less 0.05, when the distributions are identical. Thus, we tested two statistical hypotheses: location shift and scale. The null location shift hypothesis means that the mathematical expectations of both distributions are identical. The null scale hypothesis means that the variances of both distributions are identical. Alternative hypothesis, in contrast, asserts that the distribution functions are different. The results are presented in Tables 1-11.

Table 1

P -statistics for the location shift hypothesis without ties

Distribution	N(0,1)	N(1,1)	N(2,1)	N(3,1)	N(4,1)
N(0,1)	1.000	0.752	0.680	0.457	0.389
N(1,1)	–	1.000	0.846	0.584	0.424
N(2,1)	–	–	1.000	0.680	0.442
N(3,1)	–	–	–	1.000	0.570
N(4,1)	–	–	–	–	1.000

Table 2

Exact p -statistics for the location shift hypothesis without ties

Distribution	N(0,1)	N(1,1)	N(2,1)	N(3,1)	N(4,1)
N(0,1)	1.000	0.646	0.459	0.376	0.374
N(1,1)	–	1.000	0.990	0.522	0.418
N(2,1)	–	–	1.000	0.859	0.505
N(3,1)	–	–	–	1.000	0.959
N(4,1)	–	–	–	–	1.000

Table 3

P -value of the Kolmogorov–Smirnov test for the location shift hypothesis without ties

Distribution	N(0,1)	N(1,1)	N(2,1)	N(3,1)	N(4,1)
N(0,1)	1.000	0.0001	<0.0001	<0.0001	<0.0001
N(1,1)	–	1.000	0.0003	<0.0001	<0.0001
N(2,1)	–	–	1.000	<0.0001	<0.0001
N(3,1)	–	–	–	1.000	<0.0001
N(4,1)	–	–	–	–	1.000

Table 4*P*-value of the Wilcoxon signed rank test for the location shift hypothesis without ties

Distribution	N(0,1)	N(1,1)	N(2,1)	N(3,1)	N(4,1)
N(0,1)	1.000	0.002	<0.0001	<0.0001	<0.0001
N(1,1)	–	1.000	0.005	<0.0001	<0.0001
N(2,1)	–	–	1.000	<0.0001	<0.0001
N(3,1)	–	–	–	1.000	<0.0001
N(4,1)	–	–	–	–	1.000

Note that the *p*-statistic is monotonically decreasing as the location shift increases. As expected, in this case the Kolmogorov-Smirnov and Wilcoxon sign rank tests work well. However, when the distribution functions are largely overlapped the discrepancy between them is not very significant. Moreover, the Wilcoxon signed-rank test poorly recognizes the inversions between largely overlapped samples. These statements are justified by the following results (Table 5–8).

Table 5*P*-statistics for the scale shift hypothesis without ties

Distribution	N(0,1)	N(0,2)	N(0,3)	N(0,4)	N(0,5)
N(0,1)	1.000	0.726	0.641	0.581	0.427
N(0,2)	–	1.000	0.819	0.753	0.620
N(0,3)	–	–	1.000	0.979	0.976
N(0,4)	–	–	–	1.000	0.998
N(0,5)	–	–	–	–	1.000

Table 6Exact *p*-statistics for the scale shift hypothesis without ties

Distribution	N(0,1)	N(0,2)	N(0,3)	N(0,4)	N(0,5)
N(0,1)	1.000	0.725	0.556	0.585	0.613
N(0,2)	–	1.000	0.899	0.785	0.706
N(0,3)	–	–	1.000	0.988	0.996
N(0,4)	–	–	–	1.000	0.998
N(0,5)	–	–	–	–	1.000

Table 7*P*-value of the Kolmogorov–Smirnov test for the scale shift hypothesis without ties

Distribution	N(0,1)	N(0,2)	N(0,3)	N(0,4)	N(0,5)
N(0,1)	1.000	0.011	0.003	<0.0001	<0.0001
N(0,2)	–	1.000	0.027	0.012	0.084
N(0,3)	–	–	1.000	0.752	0.398
N(0,4)	–	–	–	1.000	0.742
N(0,5)	–	–	–	–	1.000

Table 8*P*-value of the Wilcoxon signed rank test for the scale shift hypothesis without ties

Distribution	N(0,1)	N(0,2)	N(0,3)	N(0,4)	N(0,5)
N(0,1)	1.000	0.212	0.341	0.352	0.920
N(0,2)	–	1.000	0.065	0.144	0.350
N(0,3)	–	–	1.000	1.000	0.307
N(0,4)	–	–	–	1.000	0.751
N(0,5)	–	–	–	–	1.000

The Kolmogorov–Smirnov test fails when samples are largely overlapped in more than almost a half of the cases, and the Wilcoxon signed-rank test has failed at all. The Klyushin–Petunin test fails in almost a third of cases of very overlapped samples following the distributions $N(0,3)$, $N(0,4)$ and $N(0,5)$.

To simulate the ties in samples we rounded the samples from previous experiments to two decimal digits. Due to this, every sample contained four ties. The results are provided in Tables 9–12. The construction of the Kolmogorov–Smirnov and Wilcoxon signed rank tests do not depend on ties. Thus, we provide only results for the p -statistics.

Table 9

P-statistics for the location shift hypothesis with ties

Distribution	N(0,1)	N(1,1)	N(2,1)	N(3,1)	N(4,1)
N(0,1)	1.000	0.672	0.505	0.355	0.309
N(1,1)	–	1.000	0.831	0.305	0.323
N(2,1)	–	–	1.000	0.705	0.424
N(3,1)	–	–	–	1.000	0.573
N(4,1)	–	–	–	–	1.000

Table 10

Exact *P*-statistics for the location shift hypothesis with ties

Distribution	N(0,1)	N(1,1)	N(2,1)	N(3,1)	N(4,1)
N(0,1)	1.000	0.627	0.504	0.367	0.330
N(1,1)	–	1.000	0.804	0.561	0.348
N(2,1)	–	–	1.000	0.683	0.480
N(3,1)	–	–	–	1.000	0.581
N(4,1)	–	–	–	–	1.000

Table 11

P-statistics for the scale shift hypothesis with ties

Distribution	N(0,1)	N(0,2)	N(0,3)	N(0,4)	N(0,5)
N(0,1)	1.000	0.420	0.331	0.356	0.392
N(0,2)	–	1.000	0.799	0.682	0.588
N(0,3)	–	–	1.000	0.993	0.965
N(0,4)	–	–	–	1.000	0.982
N(0,5)	–	–	–	–	1.000

Table 12

Exact *P*-statistics for the scale shift hypothesis with ties

Distribution	N(0,1)	N(0,2)	N(0,3)	N(0,4)	N(0,5)
N(0,1)	1.000	0.726	0.640	0.495	0.524
N(0,2)	–	1.000	0.987	0.928	0.774
N(0,3)	–	–	1.000	0.967	0.953
N(0,4)	–	–	–	1.000	0.983
N(0,5)	–	–	–	–	1.000

The p -statistics monotonically decreases as the difference between the means increases. The Klyushin–Petunin test, like the Kolmogorov–Smirnov test, does not distinguish between the distributions of $N(0, 3)$, $N(0, 4)$, and $N(0, 5)$. At the same time, it turned out to be effective in cases where the Kolmogorov–Smirnov tests and the Wilcoxon sign rank test do not work. Thus, there is an advantage of the p -statistics over the Kolmogorov–Smirnov and Wilcoxon sign rank tests.

5. Conclusions

Correct generalization based on finite training sets depends on correctly chosen underlying hypotheses. Traditional discriminant analysis is based on the compactness hypothesis, which states that objects of one class in the feature space are located closer to each other than to objects from another class. This geometric hypothesis does not work when classifying random samples that differ from feature vectors. For samples, the concept of distance is meaningless. It should be replaced by the concept of homogeneity, meaning that features of objects have the same distribution function. The evaluation of the homogeneity of the samples is provided by the Petunin p -statistics and its variants, which demonstrate high sensitivity and specificity in experiments both when testing the hypothesis of a shift in the mean and in testing the hypothesis of a shift in the scale. The proposed method has a rigorous mathematical justification and high efficiency in practical applications.

6. References

- [1] R.P.W. Duin, D. de Ridder, D.N.J. Tax, Experiments with a featureless approach to pattern recognition, *Pattern Recognit Lett* 18 (1997) 1159–1166. doi: 10.1016/S0167-8655(97)00138-4.
- [2] R.P.W. Duin, E. Pekalska, D. de Ridder, Relational discriminant analysis, *Pattern Recognition Letters* 20 (1999) 1175–1181. doi: 10.1016/S0167-8655(99)00085-9.
- [3] E. Pekalska, R.P.W. Duin, On combining dissimilarity representations, in: J. Kittler, F. Roli (Eds.), *Multiple Classifier Systems, LNCS*, vol. 2096, Springer-Verlag, 2001, pp. 359–368. doi: 10.1007/3-540-48219-9_36.
- [4] E. Pekalska, R.P.W. Duin, *The Dissimilarity Representation for Pattern Recognition, Foundations and Applications*, World Scientific, Singapore, 2005.
- [5] V. Mottl, S. Dvoenko, P. Seredin, C. Kulikowski, I. Muchnik, Featureless pattern recognition in an imaginary Hilbert space and its application to protein fold classification. *Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science*, 2123 (2001) 322–336. doi: 10.1007/3-540-44596-X_26.
- [6] V. Mottl, O. Seredin, S. Dvoenko, C. Kulikowski, I. Muchnik, Featureless pattern recognition in an imaginary Hilbert space, in: *Object recognition supported by user interaction for service robots*, Quebec City, QC, Canada, 2002, pp. 88-91, vol.2. doi: 10.1109/ICPR.2002.1048244.
- [7] O. Seredin, V. Mottl, A. Tatarchuk, N. Razin, D. Windridge, Convex support and Relevance Vector Machines for selective multimodal pattern recognition, in: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, Japan, 2012, pp. 1647–1650.
- [8] V. Mottl, O. Seredin, O. Krasotkina, Compactness Hypothesis, Potential Functions, and Rectifying Linear Space in Machine Learning: International Conference Commemorating the 40th Anniversary of Emmanuil Braverman's Decease, Boston, MA, USA, April 28-30, 2017, Invited Talks. doi: 10.1007/978-3-319-99492-5_3.
- [9] R.I. Andrushkiw, N.V. Boroday, D.A. Klyushin, Y.I. Petunin, *Computer-aided cytogenetic method of cancer diagnosis*, New York, Nova Publishers, 2007.
- [10] B. Kulis. Metric learning: A survey. *Foundations and Trends in Machine Learning*, 5 (2013) 287–364. doi: 10.1561/22000000019
- [11] A. H. de Souza Junior, F. Corona, G. A. Barreto, Y. Miche, A. Lendasse, Minimal Learning Machine: A novel supervised distance-based approach for regression and classification. *Neurocomputing*, 164 (2015) 34–44. doi: 10.1016/j.neucom.2014.11.073.
- [12] D. P. P. Mesquita, J. P. P. Gomes, A. H. de Souza Junior, Ensemble of efficient minimal learning machines for classification and regression, *Neural Processing Letters*, 46 (2017) 751–766. doi: 10.1007/s11063-017-9587-5.
- [13] A. N. Maia, M. L. D. Dias, J. P. P. Gomes, and A. R. da Rocha Neto, Optimally selected minimal learning machine, in: H. Yin, D. Camacho, P. Novais, A. J. Tallón-Ballesteros (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL*, Springer International Publishing, Cham, 2018, pp. 670–678. doi: 10.1007/978-3-030-33617-2.

- [14] W. L. Caldas, J. P. P. Gomes, D. P. P. Mesquita, Fast Co-MLM: An efficient semi-supervised co-training method based on the minimal learning machine, *New Generation Computing*, 36 (2018) 41–58. doi: 10.1007/s00354-017-0027-x
- [15] J. A. Florêncio, M. L. D. Dias, A. R. da Rocha Neto, A. H. de Souza Júnior, A fuzzy c-means-based approach for selecting reference points in minimal learning machines, in: G. A. Barreto and R. Coelho (Eds.), *Fuzzy Information Processing*, Springer International Publishing, Cham, 2018, pp. 398–407. doi: 10.1007/978-3-319-95312-0_34.
- [16] T. Kärkkäinen, Extreme minimal learning machine: Ridge regression with distance-based basis, *Neurocomputing*, 342 (2019) 33–48. doi:10.1016/j.neucom.2018.12.078.
- [17] H. Cao, S. Bernard, R. Sabourin, L. Heutte, Random forest dissimilarity based multi-view learning for radiomics application, *Pattern Recognition* 88 (2019) 185–197. doi: 10.1016/j.patcog.2018.11.011
- [18] J. A. Florêncio, S. A. Oliveira, J. P. Gomes, A. R. da Rocha Neto, A new perspective for minimal learning machines: A lightweight approach, *Neurocomputing* 401 (2020). doi: 10.1016/j.neucom.2020.03.088.
- [19] L. Nanni, A. Rigo, A. Lumini, S. Brahmam, Spectrogram Classification Using Dissimilarity Space, *Appl. Sci.* 10 (2020) 4176. doi: 10.3390/app10124176.
- [20] A. C. F. da Silva, F. Saïs, E. Waller, F. Andres, Dissimilarity-based approach for Identity Link Invalidation, in: *IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Bayonne, France, 2020, pp. 251–256, doi: 10.1109/WETICE49692.2020.00056.
- [21] M. Bicego, Dissimilarity Random Forest Clustering, in: *IEEE International Conference on Data Mining (ICDM)*, Sorrento, Italy, 2020, pp. 936-941. doi: 10.1109/ICDM50108.2020.00105.
- [22] Y. M. G. Costa, D. Bertolini, A. S. Britto, G. D. C. Cavalcanti, L. S. Oliveira, The dissimilarity approach: a review, *Artificial Intelligence Review*, 53 (2020) 2783–2808. doi: 10.1007/s10462-019-09746-z.
- [23] J. Hämäläinen, A. Alencar, T. Kärkkäinen, C. Mattos, A. Souza Júnior, J.P.P. Gomes, Minimal Learning Machine: Theoretical Results and Clustering-Based Reference Point Selection, *Journal of Machine Learning Research*, 21 (2020) 1–29.
- [24] B. Derrick, P. White, D. Toher, Parametric and non-parametric tests for the comparison of two samples which both include paired and unpaired observations. *Journal of Modern Applied Statistical Methods*, 18 (2019) eP2847. doi: 10.22237/jmasm/1556669520.
- [25] D. A. Klyushin, Yu. I. Petunin, A Nonparametric Test for the Equivalence of Populations Based on a Measure of Proximity of Samples, *Ukrainian Mathematical Journal* 55 (2003) 181–198. doi: 10.1023/A:1025495727612.
- [26] D. M. Hill, Posterior distribution of percentiles: Bayes’ theorem for sampling from a population. *Journal of American Statistician Association* 63 (1968) 677–691.
- [27] I. Madreimov, Yu. I. Petunin, Characterization of a uniform distribution using order statistics. *Theory of Probability and Mathematical Statistics*, 27 (1982) 96–102.
- [28] R.I. Andrushkiw D. A. Klyushin, Yu. I. Petunin, and V. N. Lysyuk, Construction of the bulk of general population in the case of exchangeable sample values, in: *Proceedings of the International Conference of Mathematics and Engineering Techniques in Medicine and Biological Science (METMBS'03)*, Las Vegas, Nevada, USA, 2003, pp. 486–489.
- [29] S.A. Matveichuk, Yu.I. Petunin, Generalization of Bernoulli schemes that arise in order statistics. I. *Ukrainian Mathematical Journal* 42 (1990): 459–466. doi: 10.1007/BF01058940.
- [30] S.A. Matveichuk, Yu.I. Petunin, Generalization of Bernoulli schemes that arise in order statistics. II. *Ukrainian Mathematical Journal*, 43(1991): 728–734. doi: 10.1007/BF01058940.
- [31] N. Johnson, S. Kotz, Some generalizations of Bernoulli and Polya-Eggenberger contagion models. *Statist Paper* 32 (1991) 1–17. doi: 10.1007/BF02925473.
- [32] R.I. Andrushkiw, D.A. Klyushin, Yu.I. Petunin, M. Yu. Savkina, The exact confidence limits for unknown probability Bernoulli models, in: *27th International Conference on Information Technology Interfaces*, 2005, pp. 164-168. doi:10.1109/ITI.2005.1491116.
- [33] D. Vysochanskij, Y. Petunin, Justification of the 3-sigma rule for unimodal distributions, *Theory of Probability and Mathematical Statistics* 21 (1980) 25–36.