Classifying neural networks and methods of their illogical behaviour revealing

N Y Grigoryeva¹, T R Zhangirov², A S Perkov², S A Ivanova², A A Liss²

¹ Saint-Petersburg Research Center for Ecological Safety, Russian Academy of Sciences, ul. Korpusnaya, 18 197110 Saint-Petersburg, Russia
² Saint-Petersburg Electrotechnical University “LETI”, ul. Prof. Popov, 5 197376 Saint-Petersburg, Russia

E-mail: renes3@mail.ru

Abstract. In this article a feedforward error backpropagation artificial neural network is investigated and the analysis of its illogical behaviour is presented. The problem of illogical behavior arises in various models of artificial neural networks. In the presented work a classifying artificial neural network (CANN) is considered and several learning algorithms were implemented and compared. CANN was designed for automatic differentiation of cyanobacterial strains during environmental monitoring and some of trained networks demonstrated illogical behavior in further testing. Several original techniques were elaborated for estimation of the quality and accuracy of classification in addition to the traditional ones. Novel visualization methods were suggested for classification and generalization results representation.

1. Introduction

Artificial neural networks (ANNs) are now widely used for modeling and simulating in almost all fields related to processing large volumes of information: in energy sources, chemical engineering, water treatment, food chemistry, control domain, etc.

One of the most common problem, solved by means of ANN, is the classification problem. Neural networks are quite actively used in the classification of numerical and time series, which are utilized in physical fields. During the last years, the ANNs have been frequently used to process spectral signals [1–3].

In general, the classification problem is one of the data processing tasks, which matches the considered observation with one of the known target classes [4]. Each observation is characterized by a set of features. The classification of observations is a vital application problem in most scientific areas. The solution of this problem is associated with the elaboration of a correct model with minimal classification errors, as well as with verification of the results of applying such model to the known and unknown observations.

The wide application of CANN is due to several advantages, firstly, the ability of CANN to create a model with a high level of nonlinearity and, secondly, the ability of CANN to work with samples that cannot be correctly statistically analyzed by conventional methods. On the other hand, CANNs have one great disadvantage related to its intrinsic nature, it is a randomness of the resulting trained model. Using one dataset one can obtain a large number of different models that can give an approximately equal good results. Unfortunately, these models follow their own inner logic, which is different for each model, thus some of them may display illogical behaviour in specific non-standard situations. This leads to the necessity of accurate selection of the best model from a set of similar ones.
Selection of a correctly trained and correctly classifying neural network is an important problem of the machine learning. There are various ways to verify the results of the neural network operation, and some of them have become conventional. However, the conventional methods cannot sometimes detect the illogical behavior of a trained model. One of the examples of such conventional method is the calculation of generalization errors for new observations, belonging to one of target classes of the already trained classifier. But sometimes the illogical behavior can become obvious only while the dataset of observations, not belonging to the considered classes, was used for testing [5].

In this paper a new original technique for verifying the classification results, obtained by a multilayer feed-forward artificial neural network, is presented and new methods for identifying of ANN’s illogical behavior are suggested, including the method based on a comparison with the results of the linear discriminant analysis (LDA).

The proposed technique is based on the analysis of entropy and probability distribution of the neural network output. We describe some problems in correctly trained neural network identification, and suggest possible ways to overcome these problems. The problem of generalization of data not belonging to the classes, used in training process, is discussed. The effectiveness of the presented technique in solving a specific classification problem is demonstrated on the ordinary ecological problem of classification of 16 strains of blue-green algae (cyanobacteria). Such problem arises during the environmental monitoring of toxical “blooms” of water bodies. The considered classification model is based on a 63 features, extracted from a series of 7 single-cell self-fluorescence spectra for each observation. The cases of illigical behaviour of one of the trained CANN are demonstrated on the base of the comparison of CANN results with the results of LDA.

2. Material and methods

One of the aims of the environmental monitoring of water bodies is the investigation and control of toxic cyanobacterial “blooms”. This task involves the study of biological diversity of cyanobacteria and the identification and continuous monitoring of the development of potentially harmful strains. Until now, this problem has been difficult to formalize due to the ambiguous and labour-consuming methods of data collection. However, recently the authors of this article developed a novel technique based on a single-cell data obtained by means of fluorescent microscopic spectroscopy [6–7]. The intrinsic single-cell fluorescence spectra of cyanobacteria, recorded by means of confocal laser scanning microscopy, are convenient for statistical processing and for subsequent computer analysis via various mathematical methods. In this work 307 single cyanobacterial cells, belonging to 23 different strains, were analyzed. For each sample a set of 7 fluorescent spectra were obtained. The classification problem was solved on the base of artificial neural network modelling and the results were verified by comparing with the results of linear discriminant analysis.

2.1. Data acquisition and pre-processing

23 cyanobacterial strains, used in this investigation, were provided by Core Facility Center for Culture Collection of Microorganisms of the Research Park of St. Petersburg State University. In the CALU collection of this facility center cyanobacterial strains are numbered and for the clarity of further narration the CALU numbers for corresponding strains were used below in all illustrations. To record self-fluorescence spectra of living cyanobacterial cells confocal laser scanning microscope Leica TCS-SP5 was used. In the presented investigation fluorescence emission spectra of the intact cells were measured at 7 excitation wavelengths: 405, 458, 476, 488, 496, 514, 543 nm.

The main problem of the considered classification task resulted from the high non-uniformity of the initial data and non-ballansed class sizes (different number of observation in each class). A smallness of initial dataset, as well as the sophisticated nature of the experimental data requires complex pre-processing procedure. The original experimental data represents a set of self-fluorescence spectra of cyanobacterial cells, belonging to 23 different strains. Each observation from a data set is described by a series of seven spectra taken from a single cell by means of confocal laser scanning microscopy. Each initial spectrum is an array of 38-45 numbers, which correspond to the fluorescence intensities on specific emission frequencies of the visible range from 520 to 785 nm. In contrast to the previous investigations, that utilized
for classification a full spectrum of the samples [8-10], we used a set of integral and statistical characteristics, describing the shape of each spectrum. To extract a set of classification features from initial data a computer program has been developed in mathematics system MATLAB [11]. By means of this program, interpolation, extrapolation and smoothing of the raw spectra were carried out. All spectra were reduced to the same scale and size of data array, the first derivative was taken over initial spectra and the fast Fourier transform was performed. The specific values characterizing the shape of obtained curves and the spectral composition of their derivatives were calculated. After additional investigation, a set of 63 features, describing the shape and statistical characteristics of seven-spectra set, has been determined. All selected classification features can be divided into three groups: asymmetry and excess, deviation between 0 and 1 and allows to analyze FFAN.

In our case cross-entropy is better than mean square error for training of classifying neural networks, because it allows to measure distance between two distributions of probability. In our case cross-entropy

\[ H(i|\lambda) = \frac{1}{L} \sum_{i=1}^{L} -y_i \ln y_i, \]

where \( y_i \) and \( y_i \) is obtained and expected value on \( i-th \) neuron of the output layer, respectively. It should be noted, that cross-entropy is better than mean square error for training of classifying neural networks.

2.2. Artificial neural network

There is a lot of different types and architectures of neural networks varying fundamentally. In this paper a feed-forward ANN (FFANN) is used for solving considered classification problem [13]. The architecture of FFANN may contain two or more layers. An input layer contains the input variables and output layer contains the solution of the problem. An additional intermediate (hidden) layers are employed to handle the problem’s nonlinearity and complexity. Due to the simplicity of the problem to be solved, a multi-layer feed-forward neural network (FFANN) with one hidden layer was considered (figure 1).

The size of the input layer is equal to the number of classification features – 63 neurons. The number of neurons on the output layer corresponds to the number of target classes (16 in our case). During the investigation, several variants of network architecture with different number of neurons in hidden layer have been tested. Finally, the number of hidden-layer neurons was established as 31 [14]. An additional bias neuron with a signal equal to unity is added on input and hidden layers, which helps to obtain trained model with higher classification accuracy.

While FFANN is activated by feeding the input vector, this vector propagates through the layers and transforms. The way of vector transformation is given by weights of neural network and activation functions at each layer. At each layer, the input signal is converted according to the following rule:

\[ z_j = F_i \left( \sum_{j=1}^{L} w_{ji} h_j \right), \]

where \( z_j \) - state of \( i-th \) neuron, \( h_j \) - value of \( j-th \) input signal, \( w_{ji} \) - weight between \( i-th \) and \( j-th \) neuron of two adjacent layers, \( L \) – number of neurons in previous layer, \( F_i \) - activation function on \( i-th \) neuron.

As an activation function, a hyperbolic tangent, bounded between -1 and 1, was used on the hidden layer, and Softmax function was used on the output layer. Softmax function at a specified neuron \( i \) of the considered layer have a following form:

\[ F_i(\hat{h}) = e^{\hat{h}_i} / \sum_{j=1}^{L} e^{\hat{h}_j}, \]

where the summation is performed over the neurons of the considered layer. Softmax function is bounded between 0 and 1 and allows to analyze FFANN output as a class likelihood.

Softmax is a conventional function used in classifying models, which allows to use cross-entropy for deviation calculation (i.e. as an error function):

\[ E = -\sum_{i=1}^{L} y_i \ln y_i, \]

where \( y_i \) and \( y_i \) is obtained and expected value on \( i-th \) neuron of the output layer, respectively. It should be noted, that cross-entropy is better than mean square error for training of classifying neural networks, because it allows to measure distance between two distributions of probability. In our case cross-entropy
measures the distance between expected and obtained by ANN probabilities of observation’s classification.

For the initial initialization of the neural network weights, Xavier initialization was used

\[ W \sim U \left(-\sqrt{\frac{6}{L_k+L_{k+1}}}, \sqrt{\frac{6}{L_k+L_{k+1}}}\right) \]

where \( U \) – uniform distribution, \( k \) – layer number. During the preliminary study LeCun and He weight initializations were rejected [15].

According to supervised learning the network is trained with a dataset of observations and optimized basing on its ability to predict a set of known outcomes. The deviation of the network solution from the target (true) value is computed, and the calculation of the error is propagated backward from the output layer to adjust the connection weights. Several first-order modifications of gradient descent method were tested as FFANN training algorithms. Finally, Adam method (method of adaptive moment estimation) was chosen due to the best performance [16]. This method approximates a second-order learning methods and analyzes value of error on previous epoch of training instead of analyzing only a current step, like in standard gradient descent method. The calculation of weight change in this case is similar to all first-order gradient methods:

\[ (\Delta W)_t = \eta \cdot (g_t + \rho \cdot W_{t-1}) + \nu \cdot (\Delta W)_{t-1} \]

(4)

where \( \eta \) - learning rate, \( \rho \) - regularization coefficient, \( \nu \) - momentum coefficient. Here learning rate \( \eta \) specifies the magnitude of changes in the weighting coefficients, and the coefficients of the momentum \( \nu \) and regularization \( \rho \) reduce the probability of getting stuck into a local minimum. According to the conducted investigation the values of these coefficients were chosen as follows: \( \eta = 0.05 \), \( \nu = 0.001 \) and \( \rho = 0.001 \).

The equation (4) differs from standard gradient descent method only by the presence of the function of error gradient \( g_t = D_t/(1-\beta)\sqrt{(1-\alpha)}S_t \), with \( S_t = \alpha \cdot S_{t-1} + (1-\alpha) \cdot (\nabla E)^2_t \), \( D_t = \beta \cdot D_{t-1} + (1-\beta) \cdot (\nabla E)^{t-1} \), where \( S_0 = 0 \), \( D_0 = 0 \), \( \alpha = 0.999 \), \( \beta = 0.9 \).

The ANN architecture presented in this paper (see figure 1), as well as the learning algorithm and its parameters were determined during the study of various configurations. The selected model after training consistently gives a classification accuracy of at least 90%.

The considered neural network architecture and selected training algorithm were implemented on MATLAB software.

3. Results and discussion

The selection of criterions for analysis of ANN classification quality in each specific case should be done according to the considered problem. In our case, final classifier should recognize all strains included in the training process and also should be able to distinguish new species (not belonging to the target classes) in percentage to already defined classes. In the frames of such approach the developed classifier

\[ \text{Figure 1. Artificial Neural Network model.} \]
must show high generalization quality (internal and external) in addition to high classification accuracy. It is well-known, that the analysis of generalization quality, especially the external generalization quality, is closely related to the problem of illogical behaviour of ANNs. Here under illogical behaviour we mean that the trained ANN classifies observation according to its own criterions that differ from normal human criterions. Therefore, according to these abnormal criterions ANN can make crucial errors in the nonstandard irregular situations. Thus, some specific additional techniques are required for revealing of such illogical behaviour. Here we present a set of conventional and new methods for classification quality estimation.

3.1. Cross-entropy, classification accuracy and entropy

The analysis of the value of the error function allows one to select immediately those models that were trained incorrectly and which have no sense to analyze further. However, only on the basis of the value of error function is impossible to evaluate the quality of the trained model, therefore the next stage of the analysis should be evaluation of the classification accuracy.

The first criterion for assessing the quality of ANN operation is the value of the network error function – in our case it is cross-entropy. The dynamic changes of cross-entropy allows to determine whether the model has been correctly trained or not. As an example, the dependence of cross-entropy on the number of learning epochs for two model networks is presented in figure 2. These two network models (figure 2, a and figure 2, b) were chosen from a set of 10 as examples of “correctly” and “incorrectly” trained ANNs. Three curves red, blue and green correspond to learning, control and test samples. Here learning and control sample consist of 70 and 30 percent of the whole data set, respectively. And for the test sample the whole data set was used. This graph helps to identify the usual problems of the training process such as falling into a local minimum and overfitting.

Figure 2. Cross-entropy (left axis) and classification accuracy (right axis) versus number of epochs for two selected models: a) – model 1, b) – model 2. Three curves red, blue and green correspond to learning, control and test samples.

Usually falling into the local minimum is characterized by the fact that the value of error function for training sample begins to oscillate at a sufficiently high value of this function. Here both of sample networks have no such problems.

The occurrence of overfitting is detected as an increase of the error function (cross-entropy) for the learning, control and test samples. Overfitting is the state of the model when it gives a correct result only for the data on which it was trained. The ability of internal generalization for ANN with overfitting will be very low. According to figure 2 for both chosen networks overfitting arises after 500 epoch. Comparison of the curve shape for two models reveals that the second model has slightly better classification ability since it has less cross-entropy for learning sample. Moreover, the curve slope at the last 100 epochs also is less for the second model. So by this criterion it is better than first one.
Unfortunately, further investigation reveals that the generalization quality of the second model is worse, thus it cannot be selected as the best of these two ones. It should be noted here, that the separation of the whole dataset on learning and control samples was carried out randomly, thus the difference in the slope for the control samples can be explained by high non-uniformity of the initial data, and the analysis in this case should be made on the basis of learning and test samples.

The analysis of cross-entropy dynamics allows one to select immediately those models that were trained incorrectly as a whole and have some obvious problems. However, the best network model from a set of trained ones cannot be selected based on the error function solely. Thus further analysis should be carried out.

In our case it is more correct to calculate the classification accuracy for each class independently, as the ratio of the number of correctly classified class observations to the total number of observations in a given class. Then the average classification accuracy for all classes should be calculated. To determine minimal required number of epochs for learning the dependence of classification accuracy on a number of learning epochs should be plotted (figure 2 a, b, right axis, black curve). From figure 2 it is obvious, that after 500 epoch classification accuracy changes very slowly for both cases, so there is no reason to train the model further, because it can be overfitted. And again, following the conventional rules of model comparison the second model will be better of two ones, because it have higher classification accuracy 97.67%, in contrast to 95.73% for the first model.

The calculation of classification accuracy for each class independently allows to build a matrix of errors with size CxC (C – number of classes), which can be represented as a bar chart (Figure 3 c, d), thus the classification accuracy for each class can be visualized. Bar charts show the averaged probability distribution within each class (or averaged class likelihood) i.e. the averaged probability of assigning the observations of a specific class to corresponding target classes. The disadvantage of ANN analysis based on the classification accuracy assessment is the absence of a criterion for selecting the best model from a set, when all models have approximately the same classification accuracy. To solve this problem an additional criteria for selecting the best model from a set should be used.

The entropy of the network output can become such additional criteria for the case, when the models have similar classification accuracy:

$$h = -\sum_{i=1}^{L} y_i \cdot \ln y_i,$$

where $y_i$ – the value of the network output on the $i$-th neuron, $L$ – the number of neurons on the output layer.

While the cross-entropy, indicates the similarity of two probability distributions, and in the case of ANN, shows the closeness of the expected and obtained network output values, the conventional entropy (5) allows to estimate the degree of “chaos” for the probability distribution of each class. Entropy will be higher for a model that has more uniform distribution of output between target classes. Thus, the entropy decreases when the accuracy of the classification of observations of one class tends to unity.

In the case of ANN, the entropy for each individual class can be interpreted as an indicator of how “confidently” the model determines the observations of this class. In the case of a large number of classes, the entropy histogram can be plotted (figure 3 a, b). The higher value of entropy indicates the class for which ANN can give a false positive result.

Actually, while analyzing the entropy of classes on which ANN was trained, the criterion of model correctness is the entropy tends to zero for all classes involved in learning process. At the same time, the entropy values for different classes should be close to each other. On the other hand, while estimating the external generalization quality, in the case when observations from a new class should be uniquely attributed to one of the known classes with similar characteristics, it is necessary to minimize the entropy, otherwise, when the new class have no similarities in target classes, the entropy should be maximal.

On the base of the averaged class likelihood (figure 3, c, d), it is possible to estimate the quality of ANN training, as well as the quality of internal generalization quality. The two models under consideration have similar classification accuracy, but second model is a little bit better with 97% against...
95% for the first model. However, from figure 3, a, b it follows, that “the confidence” of the second model is lower, because the entropy is higher for most of classes. Thus, the first model is better despite having lower classification accuracy, because it evaluates the internal generalization better.

It should be pointed out, that in the initial data set there are three very similar classes 1713, 1715, 1750. These are three cyanobacterial strains of one species. Obviously, that the classification accuracy for them is very low and these classes have maximum entropy. Consequently, the false-positive results may be observed in these classes. However, the second model also have the problems with classes 1416 and 1712. Thus, the analysis of the network output makes it possible to determine incorrectly trained ANN models and to detect the probability of false positive result (type I error). Also on the basis of such analysis the incorrectly determined source data can be revealed.

3.2. External generalization quality

usually, the generalization quality is evaluated on a set of observations, that belong to the classes included in model, so called internal generalization. However, in the areas such as microbiology and ecology, new classes of observations may appear during the research process, i.e. the classes that were not used in the model training. In this case, the developed ANN should be tested on the ability of the external generalization, i.e. the response of the ANN on observations belonging to classes unknown for the model should be analyzed. In some cases, such formulation may lead to the illogical behavior of ANN, in other words, ANN, while having high classification accuracy and good ability for internal generalization, at the same time fails in the external generalization [5].

In the considered classification problem, the operation quality of trained ANN should be determined not only by the absolute value of classification accuracy, but also by the ability of the designed network to recognize and properly classify unknown data classes, which were not included in training process. Thus, for verification of the external generalization quality the averaged class likelihood and entropy of the network output for unknown classes, should be calculated. The results of the external generalization
are verified on the base of a priori knowledge about new classes, which were obtained from an expert or from calculation of different metrics in the original feature space (e.g., Euclidian distances, Mahalanobis distances, etc.).

Figure 4. External generalization quality. Entropy histograms (a, b) and averaged class likelihood bar charts (c, d) for two models under consideration. Results for observations from 7 additional classes, not involved in the training process, are presented. The numbers on the x-axis and in the legend indicate numbers of strains from CALU collection.

In figure 4 the averaged class likelihood and entropy for 7 classes, that were not presented in learning process, are presented. It can be seen, that trained models are very different and the first model shows better external generalization quality because the entropy is lower for all classes and the classification results are closer to the expected values. The main goal of ANN classifier was to determine to which of the 16 known classes the observations from 7 unknown classes could be attributed.

Figure 5. LDA classification results plotted in first three dimensions. The unknown classes in the legend are marked by red. Closed curves on the plot indicate corresponding new observation sets.

In table 2 closely related strains for 7 unknown ones are anumerated and several numeric results of classification are given. Low averaged probability in classification of strains 756 and 550 can be explained by the fact that in the space of classification features they lie between the areas occupied by known classes, approximately at equal distances from 2 or 3 nearest ones. Therefore the ANN cannot make a correct
decision. This can be clearly demonstrated by the results of Fisher’s LDA [17] for a complete set of 23 strains presented in figure 5. Closed curves show areas occupied by seven new strains.

While comparing the results of two classifiers (LDA and ANN), the so-called intersection areas can be compared, i.e. those observations that are correctly and/or incorrectly classified in both models. If, for different ANN models, the area of incorrectly classified data coincides with the same area for LDA, then this indicates incorrectness in the source data. If for the ANN, the incorrect classification of observations that are correctly classified by the LDA model is observed, this indicates illogical behavior of a specific ANN. The second ANN model shows higher accuracy, but its areas of intersection did not match with LDA ones. Thus, the constructed model, even though it has a sufficiently high classification accuracy, is not entirely correct, and in its work, in some cases, may show illogical behavior.

4. Conclusion

In this paper, a novel technique is proposed to verify the quality of CANN operation and to identify the possibility of its illogical behavior. The presented method allows to select the best model from a number of trained ANNs. The advantage of this technique is a simple mathematical apparatus, which makes it easy to visualize the results. The elaborated technique was tested on two classifying ANN models, which solve the problem of differentiation of cyanobacterial strains using their self-fluorescence spectra. According to elaborated technique one of the compared models was rejected inspite its higher classification accuracy, because its external generalization quality turned out to be worse. The comparison with LDA result have confirmed the illogical behaviour of the second model, while classifying of the observations belonging to the unknown classes.

References

[1] Almhdi K H, Valigi P, Gulbinas V, Westphal R and Reuter R 2007 Classification with artificial neural networks and support vector machines: application to oil fluorescence spectra EARSeL eProceedings 6 (2) 115–129
[2] Goh A E C 1995 Back-propagation neural networks for modeling complex systems Artificial Intelligence in Engineering 9 (3) 143–151
[3] Heermann P D and Khazenie N 1992 Classification of multispectral remote sensing data using a back-propagation neural network IEEE Transactions on Geoscience and Remote Sensing 30 (1) 81–88
[4] Langford J 2005 Tutorial on practical prediction theory for classification Journal of machine learning research 6 273–306
[5] Altman E I, Marco G and Varetto F 1994 Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience) Journal of banking & finance 18 (3) 505–529
[6] Grigoryeva N Y, Chistyakova L V and Liss A A 2018 Spectroscopic Techniques for Estimation of Physiological State of Blue-green Algae after Weak External Action Oceanology 58 (6) 989–1000
[7] Grigoryeva N Y and Chistyakova L V 2018 Fluorescence Microscopic Spectroscopy for Investigation and Monitoring of Biological Diversity and Physiological State of Cyanobacterial Cultures In: Cyanobacteria, ed.by A. Tiwari (IntechOpen) 11–43
[8] Millie D F, Schofield O M, Kirkpatrick G J, Johnsen G and Evens T J 2002 Using absorbance and fluorescence spectra to discriminate microalgae European Journal of Phycology 37 (3) 313–322
[9] Escoffier N, Bernard C, Hamlaoui S, Groleau A and Catherine A 2014 Quantifying phytoplankton communities using spectral fluorescence: the effects of species composition and physiological state Journal of Plankton Research 37 (1) 233–247
[10] MacIntyre H L, Lawrenz E and Richardson T L 2010 Taxonomic discrimination of phytoplankton by spectral fluorescence In: Chlorophyll a fluorescence in aquatic sciences: methods and applications (Dordrecht: Springer) 129–169
[11] MATLAB MV. 8.0. 0.783 (R2012b). Natick, Massachusetts: The MathWorks Inc. 2012;82
[12] Zhangirov T R, Perkov A S, Ivanova S A, Liss A A, Grigoryeva N Y and Chistyakova L V 2019 Comparison of the efficiency of solution of classification task by methods of linear discriminant analysis and artificial neural networks Izvestiya SPbGETU “LETI” 5 64–73
[13] Haykin S 2009 Neural networks and learning machines (NJ: Pearson Education) p 938
[14] Perkov A S, Jhangirov T R, Grigoryeva N Y, Chistyakova L V and Liss A A 2019 Comparison of neural network training methods in the classification problem Izvestiya SPbGETU “LETI” 6 53–61
[15] Glorot X and Bengio Y 2010 Understanding the difficulty of training deep feedforward neural networks Proceedings of the thirteenth international conference on artificial intelligence and statistics 3 249–256
[16] Kingma D P and Ba J 2015 Adam: A method for stochastic optimization arXiv preprint arXiv:1412.6980. 2014 Available at: https://arxiv.org/abs/1412.6980 (accessed 21.02.19).
[17] Fisher R A 1936 The Use of Multiple Measurements in Taxonomic Problems Annals of Eugenics 7179–188