Credit Card Attrition Classification Through Neuronets

Spyridon D. Mourtas[1,2,a], **Vasilios N. Katsikis** [2,b], **Romanos Sahas** [2,c]

[1] Laboratory "Hybrid Methods of Modelling and Optimization in Complex Systems",
Siberian Federal University, Prosp. Svobodny 79, 660041 Krasnoyarsk, Russia

[2] Department of Economics, Division of Mathematics-Informatics and Statistics-Econometrics,
National and Kapodistrian University of Athens, Greece

E-mail: [a]spirosmourtas@gmail.com, [b]vaskatsikis@econ.uoa.gr, [c]romanossahas@econ.uoa.gr

2022

Introduction

The following two publications served as the foundation for this presentation:

- T.E. Simos, V.N. Katsikis, and S.D. Mourtas, “Multi-input bio-inspired weights and structure determination neuronet with applications in European Central Bank publications,” Mathematics and Computers in Simulation 193, 451-465 (2022)
- T.E. Simos, V.N. Katsikis, S.D. Mourtas, “A multi-input with multi-function activated weights and structure determination neuronet for classification problems and applications in firm fraud and loan approval,” Applied Soft Computing, 127, 109351 (2022)

Introduction

- The main topic of this presentation is the use of **neuronet (also known as neural network) classification processes** in the field of finance.
- A novel weights-and-structure-determination (WASD)[1] algorithm, called **power softplus WASD (PS-WASD)**, is used to train neuronet models.
- To classify credit card customers that are likely to attrite, a **PS-WASD based neuronet model** for binary classification problems is used.

[1] Y. Zhang, D. Chen, C. Ye, "Deep Neural Networks: WASD Neuronet Models, Algorithms, and Applications," CRC Press, 2019

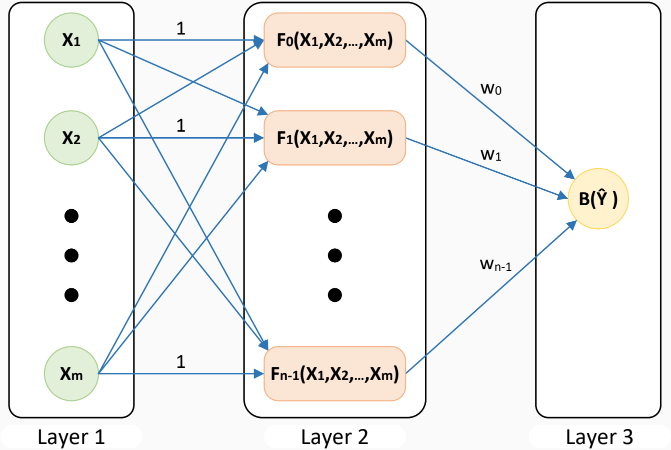
Neuronets Training Algorithms

The feed-forward neuronet (FFN) based on the error back-propagation (BP) training algorithm is one of the most important and popular FFN models, with numerous theoretical analyses and real-world applications. However, BP-based neuronets have the following weaknesses:

- the probability of being trapped in some local minima;
- difficulty selecting suitable learning rates (or, say, speed of training);
- inability to design the optimal or smallest neuronet structure in a deterministic manner (or, say, high computational complexity).

As a result, many improved BP-type algorithms, as well as many alternative neuronet training algorithms, such as WASD, have been developed and explored.

The PS-WASD neuronet model



Layer 1: Data Normalization

The data requires normalization to a range of $[-0.5, -0.25]$ before their input in the neuronet model. We achieve that by using a linear transformation as below,

$$X_{nor} = \frac{X - X_{min}}{4(X_{max} - X_{min})} - \frac{1}{2},$$

where X_{max} and X_{min} are the maximum and minimum values of the data X, respectively. It is worth noting that the PS-WASD neuronet can deal with over-fitting in this way.

Layer 2: WDD Process

The relationship between the input $X_1, X_2, \ldots, X_m$ and the target Y can be described with the following function:

$$Y = f(X_1, X_2, \ldots, X_m). \tag{1}$$

The K-order Taylor polynomial $P_K(X_1, X_2, \ldots, X_m)$ can map (1) as follows:

$$P_K(X_1, X_2, \ldots, X_m) = \sum_{d=0}^{n-1} q_d W_d, \tag{2}$$

where $q_d = F_d(X_1, X_2, \ldots, X_m) \in \mathbb{R}^{1\times m}, d = 0, 1, \ldots, m-1$, implies a power function of all inputs and $W_d \in \mathbb{R}^m$ denotes the weight (or coefficient) for q_d.

Layer 2: WDD Process

The steady-state weights of the K-order Taylor-polynomial neuronet can be obtained directly as proposed in the weights-direct-determination (WDD)[2]process:

$$W = Q^{\dagger} Y. \tag{3}$$

For a given number of samples $s \in \mathbb{N}$, $W = [W_1, W_2, \dots, W_n]^{\mathrm{T}} \in \mathbb{R}^{mn}$ implies the weights, $Y \in \mathbb{R}^{s}$ implies the desired-output vector and $Q = [q_1, q_2, \dots, q_n] \in \mathbb{R}^{s \times mn}$ denotes the input-activation matrix. Moreover, the following is the power sigmoid activation function:

$$F_d(X) = \ln\left(1 + e^{\odot X^{\odot d}}\right), \text{ with range } \left(0, \infty\right). \tag{4}$$

Note that $(\cdot)^{\odot}$ denotes the Hadamard (or element-wise) exponential.

[2]Y. Zhang, X. Yu, L. Xiao, W. Li, Z. Fan, W. Zhang, “Weights and structure determination of articial neuronets,” in: Self-Organization: Theories and Methods, New York, USA: Nova Science, 2013

Layer 2: Cross-Validation and Loss Function

The data input X is separated into two set of samples for fitting and validation. Supposing that J is the sample size of X, then the first $J_1 = pJ, p \in (0, 1] \subseteq \mathbb{R}$, samples of X are used for fitting the model and the last $J_2 = J - J_1$ samples for validation. **Cross-validation** tries to ensure that the model's performance generalizes beyond the training set because the validation set is separate from the training set.

The mean absolute error (MAE) is a measure of errors describing the same phenomenon between paired observations and, generally, is employed as a **loss function** for classification problems in machine learning, and is calculated as follows:

$$\mathrm{MAE} = \frac{1}{J} \sum_{k=1}^{J} \left| Y_k - \hat{Y}_k \right|, \tag{5}$$

where J implies the number of samples, Y denotes the target value and $\hat{Y}$ denotes the predicted value.

Layer 2: PS-WASD Algorithm

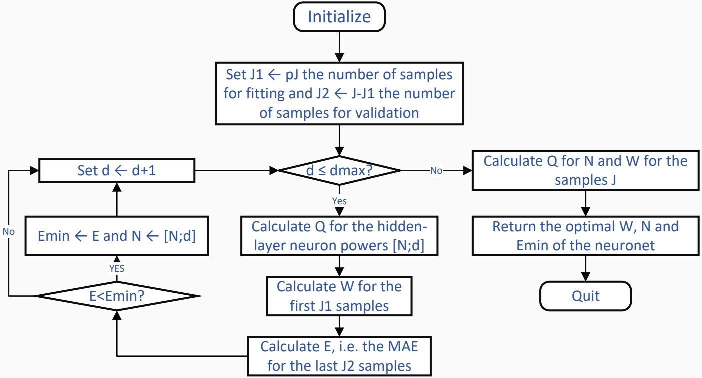


Figure 1: The PS-WASD algorithm.

Layer 3: PS-WASD neuronet Output

In the case of the third layer of the PS-WASD neuronet, the following nonlinear activation function is employed to the outcome $\hat{Y}$:

$$B(\hat{Y}) = \begin{cases} 1 & , \hat{Y} \geq -0.375 \\ 0 & , \hat{Y} < -0.375 \end{cases}, \tag{6}$$

where the numbers 0 and 1 stand for false and true, respectively, to classify something as true or false based on the corresponding input X of the first layer of the PS-WASD neuronet. Also, notice that the number -0.375 is the midpoint of the interval $[-0.5, -0.25]$.

Credit Card Attrition

The growth of a company depends significantly on the efforts that the entity makes towards: maintaining and growing its existing customer base, acquiring/keeping up with new technology, focusing on specific market segments, and improving its productivity and efficiency. This is especially true in highly competitive and mature business sectors, such as the banking sector.

Because of the impact that even a slight increase in customer retention can have on the banks' income statement along with the well-established facts that **maintaining is significantly less expensive than re-acquiring lost customers**, financial institutions have been motivated to gradually change their focus from attracting new customers to retaining as many of their existing ones.

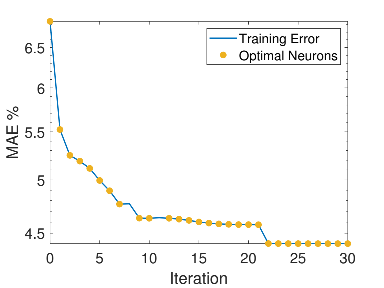
Application: Credit Card Attrition

In this application, the dataset used is the CCAD (credit card attrition dataset) taken from `https://www.kaggle.com/datasets/jaikishankumaraswamy/customer-credit-card-churn`.

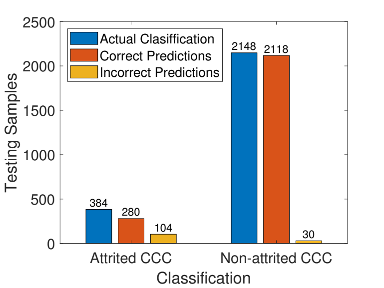
The CCAD is used for evaluating the neuronets performance and includes 33 variables and 1 target. It is worth mentioning that the CCAD is already split into training and testing sets, each of which contains 7596 and 2533 samples, respectively.

Application: Credit Card Attrition

Based on CCAD, we build a reliable classifier that can successfully take up the task of classifying credit card attrition. Note that we abbreviate "credit card customer" by CCC in the figures.



(a) neuronet training.



(b) neuronet testing.

Figure 2: Neuronet results on CCAD for $p = 0.75$ and $d_{\max} = 30$.

Application: Credit Card Attrition

Table 1: Performance comparison between classifiers.

Model	PS-WASD	KNB	Linear SVM	Fine KNN	Bernoulli WASD
MAE	**0.052923**	0.14652	0.092022	0.15008	0.19155
TP	0.98603	**1**	0.96648	0.92551	0.79562
FP	0.013966	**0**	0.03352	0.074488	0.20438
TN	0.72917	0.033854	0.58073	0.42708	**0.88021**
FN	0.27083	0.96615	0.41927	0.57292	**0.11979**
Precision	0.98603	**1**	0.96648	0.92551	0.79562
Recall	0.78452	0.50861	0.69744	0.61766	**0.86914**
Accuracy	**0.94708**	0.85348	0.90798	0.84992	0.80845
F-measure	**0.87381**	0.67428	0.81021	0.74088	0.83076

Application: Credit Card Attrition

Table 2: Results of McNemar test.

PS-WASD	McNemar test	
vs.	Null Hypothesis	p-value
KNB	Rejected	6.1951e-49
Linear SVM	Rejected	1.3301e-13
Fine KNN	Rejected	1.801e-45
Bernoulli WASD	Rejected	4.0024e-61

Conclusion

The key factors of this presentation are:

- we introduce and investigate a 3-layer PS-WASD neuronet model for credit card attrition classification;
- a new PS-WASD algorithm for training FFN neuronet models is proposed;
- an application on credit card attrition classification has demonstrated the learning and predicting performance of the PS-WASD neuronet model;
- the application results prove that neuronets trained with the PS-WASD algorithm have higher accuracy and F-measure than neuronet trained with traditional WASD algorithms, while their performance in comparison to conventional neuronet methods is pretty competitive.

References

- T.E. Simos, V.N. Katsikis, and S.D. Mourtas, “Multi-input bio-inspired weights and structure determination neuronet with applications in European Central Bank publications,” Mathematics and Computers in Simulation 193, 451-465 (2022)
- T.E. Simos, V.N. Katsikis, S.D. Mourtas, “A multi-input with multi-function activated weights and structure determination neuronet for classification problems and applications in firm fraud and loan approval,” Applied Soft Computing, 127, 109351 (2022)
- Y. Zhang, D. Chen, C. Ye, “Deep Neural Networks: WASD Neuronet Models, Algorithms, and Applications,” CRC Press, 2019
- Y. Zhang, X. Yu, L. Xiao, W. Li, Z. Fan, W. Zhang, “Weights and structure determination of articial neuronets,” in: Self-Organization: Theories and Methods, New York, USA: Nova Science, 2013

Thank you for your attention!