



From school mathematics to artificial neural networks: Developing a mathematical model to predict life expectancy

Stephan Kindler, Sarah Schönbrodt, Martin Frank

► To cite this version:

Stephan Kindler, Sarah Schönbrodt, Martin Frank. From school mathematics to artificial neural networks: Developing a mathematical model to predict life expectancy. Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13), Alfréd Rényi Institute of Mathematics; Eötvös Loránd University of Budapest, Jul 2023, Budapest, Hungary. hal-04410971

HAL Id: hal-04410971

<https://hal.science/hal-04410971v1>

Submitted on 22 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From school mathematics to artificial neural networks: Developing a mathematical model to predict life expectancy

Stephan Kindler¹, Sarah Schönbrodt¹ and Martin Frank¹

¹Karlsruhe Institute of Technology, Germany; stephan.kindler@kit.edu

In this paper, we show that it is possible to develop artificial neural networks building on school mathematical knowledge – initially avoiding AI terminology and comparisons with biological neurons since both are unnecessary to understand the underlying mathematical concepts. We present a didactical reduction of the mathematical foundations of a simple artificial neural network using the example of regression problems. It becomes clear that numerous connections to school mathematical content exist, not only from statistics but also from the area of analysis and linear algebra. As part of a design-based research project we developed digital teaching and learning material that builds on the presented didactical reduction. The material allows upper secondary students to develop the mathematical ideas of artificial neural networks in a problem-oriented way. The central building blocks of the material and first experiences with students are described.

Keywords: Artificial intelligence, machine learning, mathematics education, artificial neural networks, Jupyter Notebooks.

Artificial neural networks in mathematics education

Methods from the field of artificial intelligence (AI) and especially from its branch machine learning are the basis for numerous technologies in our daily lives (e.g., ChatGPT, DeepL, FaceID). Due to the significant relevance of AI for our society, AI education should already be anchored in schools. In fact, the topic of AI is already embedded in several computer science curricula, e.g., in parts of Germany or in the US (CSTA, 2017). The interdisciplinary nature of AI should certainly be also reflected in educational approaches and concepts in other school subjects (Micheuz, 2020). In addition to computer science skills, mathematical skills and a deep understanding of the field in which the AI application will be used are essential. Thus, AI education should not only take place in computer science classes but can also be integrated in mathematics education. In fact, various AI methods such as Support Vector Machines, Decision Trees or k Nearest Neighbour algorithms offer diverse connections to high-school mathematical contents, e.g., to vector calculation, statistics, and differential calculus (Schönbrodt et al., 2022; Biehler & Fleischer, 2021; Hazzan & Mike, 2022). Thus, AI education offers an authentic and modern application of school mathematical contents.

In this paper, the focus is on artificial neural networks (ANNs) – a well-known and frequently used method from the field of AI – more precisely from its branch machine learning. So far, research on ANNs in AI education on K12 level focuses mostly on computer science (Heinemann et al., 2020; Janssen, 2020). Furthermore, some papers exist that start with references to biological inspirations (Janssen, 2020; Swertz, 2020). We focus on the mathematical aspects. Therefore, we develop digital teaching and learning material as part of a design-based research project with the following research question: How can ANNs be presented such that the most important mathematical aspects are accessible to upper secondary high-school students? In the following, we discuss the objectives and

design principles of the developed digital learning material, shortly outlining the structure as well as the main didactical reductions. Finally, we discuss first experiences with students.

Objectives and design principles of the learning material

The learning material is developed following a design-based research approach (Prediger et al., 2012). Following this approach, we develop a prototype of the material based on the objectives and design principles described below. Then, design experiments are conducted where the material is tested and evaluated with students. Subsequently, the insights from the experiments are used to iteratively revise the material. The design project presented here is now in its first design cycle. A prototype of the material has been developed and the first design experiments are currently being conducted.

One main objective of the learning material is to allow students to understand the key mathematical aspects of ANN. More specifically, they should understand that simple ANNs can be represented as a composition of mathematical functions. They should also be able to explain the role of the chain rule when optimizing the parameters of a simple ANN. There are numerous connections to school mathematics – most importantly functions, differential calculus (chain rule) and basic knowledge of vectors. The material was devised in such a way that it utilizes school mathematical knowledge wherever possible, making those contents prerequisite for processing the learning material. Therefore, the material is targeted at students in upper secondary education, more precisely in grades 11 to 13 (corresponding to ages of approx. 15 to 18). Further mathematical concepts that go beyond school mathematical knowledge are introduced in a problem-oriented way.

Secondly, the students should be able to develop a better understanding of the importance of mathematics in real-world applications. To achieve this, the material presents an authentic problem to be solved using mathematics and computers. In solving the problem, key steps of typical mathematical modelling cycles (Blum, 1985) are run through several times. In addition, the students get to know ANN as a mathematical model. This is intended to raise awareness of the importance of (school) mathematics for ANNs and AI in general.

The learning material should be based on modern digital tools and allow differentiation for heterogeneous learning groups. To that end, the material consists mainly of interactive, digital worksheets based on Jupyter Notebooks with Python. These are hosted on servers of CAMMP¹, so they can be easily accessed through a web browser. Since Jupyter Notebooks are widely used in science and research, not only the AI method and the real-world problem but also the digital tool used are authentic. The notebooks allow for a natural combination of text fields (including images, videos and links) and code fields. They include various means of differentiation, such as individual automatic feedback on students' solutions, staggered tips, and optional tasks for more advanced students. We designed the material for potential use in regular math lessons. This means that we do not want to assume any programming knowledge from the teacher or the student. A "fill in the gap" approach has been implemented within the code cells so that the students mostly only have to add newly developed mathematical formulas.

¹ CAMMP stands for Computational and Mathematical Modeling Program, see www.cammp.online/english/index.php

Insight into the learning material

The learning material includes presentations for plenary phases and digital worksheets for working phases. The plenary phases serve to structure the workflow, while also allowing for important discussions and securing results. In the working phase, the students work in pairs on the digital worksheets. In this section, we provide insight into the learning material and give an overview of the structure. We outline the central mathematical steps that the students perform while working with the material. In addition, we clarify at which steps students draw on school knowledge and at which points we introduce concepts beyond school mathematics.

Understanding the problem and exploring the data

The learning material is built on the problem of predicting life expectancy. For that purpose, we used a dataset from the Global Health Observatory data repository from the WHO. The dataset includes twenty features (i.e., economic, social, and health-related) as well as the life expectancy as target feature for 192 countries.

Using the interactive data analysis tool CODAP² the students first explore the dataset and identify features that can be used to make first predictions on the life expectancy (see Figure 1). The students graphically determine those features that seem to have a linear relationship with life expectancy (e.g., the number of years attending school).

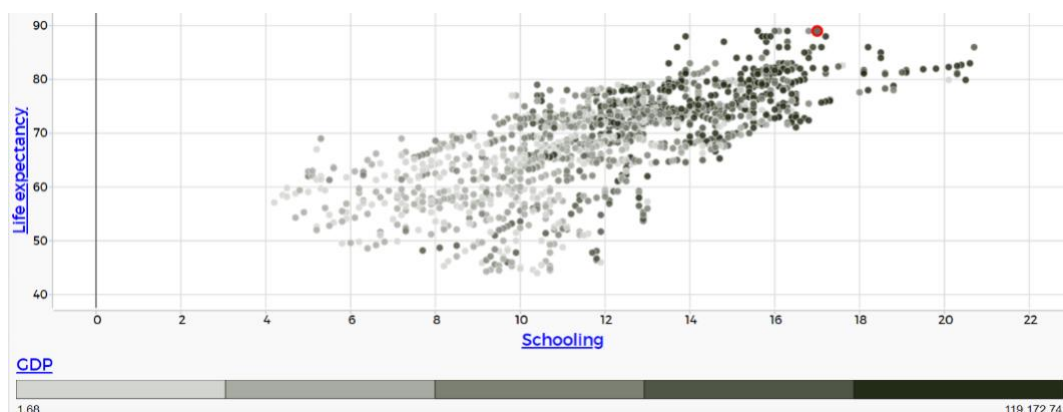


Figure 1: Data exploration using CODAP embedded in a Jupyter Notebook

During the second plenary discussion phase, a crucial aspect of many AI workflows is discussed: Dividing the given data into training and test data. The training data is used to develop a mathematical model that can be applied to predict life expectancy. The test data is then used to evaluate the model.

Developing a first mathematical model

After exploring the dataset, a first mathematical model is developed. The aim is to find a linear model f with $f_{a,b}(x) = a \cdot x + b$ and $a, b \in \mathbb{R}$ that best predicts the unknown life expectancy y based on the previously identified feature x . Therefore, students use sliders to determine the best values for parameters a and b in an interactive graph such that the line best fits the data points based on visual,

² For further information, see <https://codap.concord.org> last accessed: 12 February 2023

graphical judgment. In the following second plenary phase, the focus is on how life predictions can be made with the help of this model. Ideas for criteria to compare different models are discussed and the mean squared error (MSE) is introduced as an error function L .

For data points $(x^{(i)}, y^{(i)})$, with $i = 1, \dots, N$ and N the number of datapoints, the MSE gets

$$L(a, b) = \frac{1}{N} \left((ax^{(1)} + b - y^{(1)})^2 + \dots + (ax^{(N)} + b - y^{(N)})^2 \right).$$

Since students are not familiar with the sum symbol $\sum$, we introduce the sum in its written-out form. The function is implemented by the students in the following working phase.

Formulating and solving an optimization problem

The aim is to find the values of the parameters such that the error function is minimal, thus getting to an optimal linear function to predict life expectancy. To this end, the students work with an interactive 3D graph of the error function. Again, they use sliders to find the optimal values for the parameters, which are compared in the following plenary phase. Since the starting values are randomised, students will most likely arrive at different values. This leads to the discussion that the search for the minimum cannot be done by trial and error or by referring to normal school mathematics. The concept of the gradient descent method is then introduced as a solution. In letting the students set the parameters manually one by one, the concept has been motivated. Part of the algorithm are partial derivatives, which the students compute manually using the chain rule. Following that, they adapt a numeric representation of the algorithm on the digital worksheets.

$$a^* \rightarrow a - \alpha_a \cdot \frac{\partial L}{\partial a} \qquad b^* \rightarrow b - \alpha_b \cdot \frac{\partial L}{\partial b}.$$

Here a^* and b^* are the updated versions of a and b and $\alpha_a, \alpha_b \in \mathbb{R}$ the so-called step sizes.

After calculating new values, students receive a plot of the updated regression line and error function. For subsequent updates, partial derivatives are recalculated using new values for a and b , and a new line is plotted. The students decide how many updates they perform: stop after a certain number of updates or set a minimum improvement rate. They also experiment with different initial values as well as step sizes and observe how these affect the regression line. It is discussed that several individual decisions must be made for modelling even this reduced problem: the initial values, the step sizes, and the number of updates must be determined manually.

Finally, the students apply the current regression model to the test data. The results are assessed using a statistical measure, namely the MSE. In a plenary phase it is concluded that the prediction does not work sufficiently well. Further model improvements are discussed.

Adding more features to the model

So far, only one feature has been considered. Now we gradually include further features to improve the model and thus the regression results. Since students have already dealt with functions with multiple input variables as part of this learning material, these steps are tangible.

Adding further features is done using the same mathematical considerations. However, the notations become cumbersome and possibly confusing. Therefore, we introduce a vector notation. The features

are combined in a D -dimensional input vector $\vec{x}$, and the parameters $a_1, \dots, a_D$ in a D -dimensional vector $\vec{a}$, where D is the number of features. Thus, the notation of the model simplifies to

$$f_{\vec{a},b}(\vec{x}) = \vec{a} \cdot \vec{x} + b.$$

Putting it all together: a simple ANN

At this point we introduce a flowchart representation of $f_{\vec{a},b}$ to help understand the process so far and to make the next steps and the idea of networks comprehensible (see Figure 2).

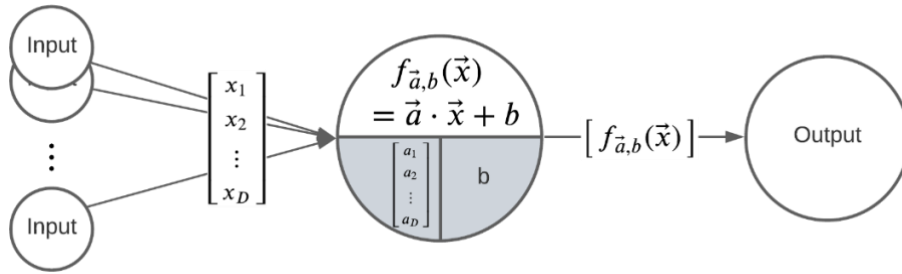


Figure 2: Flow chart representation of the function $f_{\vec{a},b}$

This model could be considered a basic building block of ANNs. It receives an input vector, applies the function f and passes on the function value as output. In the learning material this structure is described as a node. We deliberately do not introduce the usual term "neuron" here, as this could be misleading at this point. The concept of a node has been worked out purely mathematically and lacks some of the basic properties of biological neurons.

We have arrived at a simple, linear model that considers all the features of the dataset and predicts the life expectancy. However, the students observe that the predictions are still not very good. Approaches to further improve the model are discussed with the students:

Combining several nodes: It is discussed that various nodes could be used in a network of nodes, whereby the outputs of previous nodes are used as inputs for following nodes. This approach results in a network which is again visualized within a flowchart (see Figure 3). The students are encouraged to experiment with different network architectures and apply these to the problem at hand.

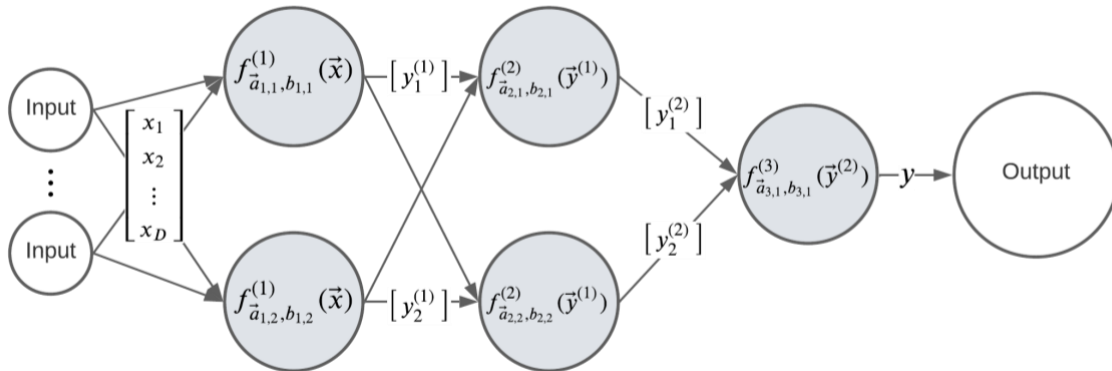


Figure 3: Example for a network of multiple nodes

From the linear to the nonlinear case: Connecting several nodes consecutively corresponds to a composition of the functions within the nodes. Moreover, since the functions in all nodes are linear, the resulting composition function is also linear. Therefore, our current model cannot represent nonlinear relations properly. Through appropriate examples and visualizations, students come to this conclusion on their own. Now, to tackle the identified problem, we introduce some nonlinear function $\sigma: \mathbb{R} \rightarrow \mathbb{R}$. This function is called an activation function and is applied to the result of each node

$$f_{\vec{a},b}(\vec{x}) = \sigma(\vec{a} \cdot \vec{x} + b).$$

The activation functions discussed are the logistic sigmoid function and the rectified linear unit.

The principle of optimization continues to work the same way. The MSE between predictions of the model and the target variables is to be minimized, so the partial derivatives are still needed. These are now the partial derivatives of multiply chained functions, which means that the chain rule must be applied repeatedly. Here we differentiate: Interested students can determine the necessary formulas themselves, while these are pre-implemented for others.

After incorporating the nonlinear functions, the model created has the typical structure of a basic ANN. Only now is the commonly used AI terminology (e.g., neuron, layers, activation function) introduced for the components of the model and the fitting process.

Critical analysis of the model

After understanding the basics of ANNs, the students try out different network architectures as well as nonlinear functions and explore their respective influence on the prediction results for life expectancy. These results are again assessed with statistical measures, first the MSE. Other measures (e.g., mean absolute error) are presented as further differentiation. The influence of the size of the network, step size and number of iterations can be examined by the students.

Lastly, the developed model is critically reviewed for its suitability for actual application. For example, the model currently processes missing input values as zeros or not meaningful values, such as 50 years of schooling or negative values for GDP. Limitations of the dataset are also discussed, as it only contains data from 2010 to 2015.

The material concludes with a critical discussion of socially problematic implications of AI systems, both in relation to the presented problem of life expectancy and more generally with respect to further AI applications.

First experiences with students

The first design experiment is currently being conducted as part of a weekly students working group in a German high school. Nine students from grades 11 to 12 registered voluntarily for this working group. Therefore, we can only describe first impressions here. The following observations stood out: Discussions about using a model to predict life expectancy were very diverse. During data exploration, students expressed numerous ideas about how possible trends in the data could be justified with respect to the real-world situation. They discussed possible correlations between different features and the resulting impact on life expectancy.

Summary and outlook

The designed teaching and learning material presented here aims to make the mathematical foundations of ANNs accessible to students in a problem-oriented way. The material guides students through the fundamental building blocks of ANNs, starting with a simple regression problem that is progressively developed into a basic ANN. Numerous school mathematical contents are used, amongst them linear functions, vectors and the chain rule. The result is a highly relevant and authentic application of central school mathematical contents, given that ANNs are commonly used in popular AI technologies like DeepL (Vos, 2011). Content that goes beyond school mathematics (e.g., MSE or partial derivatives) is introduced as part of the solving process. During the introduction, typical AI terminology was avoided. Nevertheless, key AI vocabulary is introduced at the end so that students can apply and categorize what they have learned in future discussions around ANNs.

The digital teaching and learning material was developed based on Jupyter Notebooks. It includes various means of differentiation such as automatic feedback, tips, and optional additional tasks. Since Jupyter Notebooks are used in science, the digital tool used is also authentic. The material is structured such that no programming knowledge is required on the part of the teacher or student.

In summary, this paper highlights two things in particular: Firstly, the mathematical background of fundamental components of ANNs can be understood starting only with high-school mathematical knowledge.

Secondly, the development of AI is not only extra-mathematical, but also intra-mathematical interdisciplinary. In this paper, and also in the learning material developed, the potential for intra-mathematical connectivity is illustrated by the following steps:

- *Understanding the problem and the data:* statistics, data exploration
- *Developing a mathematical model (regression model):* statistics, analysis, linear algebra
- *Formulating an error function and the optimization problem:* analysis, optimization
- *Application of an optimization algorithm:* analysis, numerical mathematics
- *Validation of the model based on test data:* statistics.

Thus, integrating ANNs into the mathematics classroom provides an opportunity to experience that AI methods heavily rely on mathematics and that inner-mathematical interconnections are often central when solving real-world problems.

The next steps in this research project are to conduct further design experiments and test the material with further and more heterogeneous groups of students. We plan to implement the material both within mathematical modeling days as well as part of a lesson series in regular mathematics classes.

References

Blum, W. (1985). Anwendungsorientierter Mathematikunterricht in der didaktischen Diskussion? [Application-oriented mathematics teaching in the didactical discussion?] In Kahle et al. (Eds.), *Mathematische Semesterberichte. Zur Pflege des Zusammenhangs zwischen Schule und Universität*, 32(2), 195–232. Vandenhoeck & Ruprecht.

- Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43, 133–142.
- Brinda, T., Diethelm, I., Gemulla, R., Romeike, R., Schöning, J., & Schulte, C. (2016). Dagstuhl-Erklärung: Bildung in der digitalen vernetzten Welt. Gesellschaft für Informatik eV. https://gi.de/fileadmin/GI/Hauptseite/Themen/Dagstuhl-Erklärung_2016-03-23.pdf
- CSTA (2017). About the CSTA K-12 Computer Science Standards. www.csteachers.org/page/standards
- Hazzan, O., & Mike, K. (2022). Teaching core principles of machine learning with a simple machine learning algorithm: the case of the KNN algorithm in a high school introduction to data science course. *ACM Inroads*, 13(1), 18–25. <https://doi.org/10.1145/3514217>
- Heinemann, B., Budde, L., Schulte, C., Biehler, R., Frischmeier, D., Podworny, S., & Wassong, T. (2018). Data Science and Big Data in Upper Secondary Schools: What Should Be Discussed From a Perspective of Computer Science Education? *Archives of Data Science, Series A*, 5(1), 1–18.
- Janssen, D. (2020). Machine learning in der Schule [Machine learning in school]. *Science on Stage Deutschland e.V.*, 325–334.
- Micheuz, P. (2020). Approaches to Artificial Intelligence as a Subject in School Education. In Brinda, T., Passey, D., Keane, T. (Eds.), *Empowering Teaching for Digital Equity and Agency*. OCCE 2020. IFIP Advances in Information and Communication Technology (pp. 3–13). Springer. https://doi.org/10.1007/978-3-030-59847-1_1
- Prediger, S., Link, M., Hinz, R., Hußmann, S., Thiele, J., & Ralle, B. (2012). Lehr-Lernprozesse initiieren und erforschen–fachdidaktische Entwicklungsforschung im Dortmunder Modell. *Der mathematische und naturwissenschaftliche Unterricht*, 65(8), 452–457.
- Schönbrodt, S., Camminady, T. & Frank, M. (2022). Mathematische Grundlagen der Künstlichen Intelligenz im Schulunterricht [Mathematical foundations of artificial intelligence in the school classroom]. *Math Semesterberichte*, 69, 73–101. <https://doi.org/10.1007/s00591-021-00310-x>
- Seegerer, S., Lindner, A., & Romeike, R. (2019). AI Unplugged–Wir ziehen Künstlicher Intelligenz den Stecker [AI Unplugged – We're pulling the plug on artificial intelligence]. In A. Pasternak (Ed.) *Informatik für alle* (pp. 325–334). <https://doi.org/10.1007/s00591-021-00310-x>
- Swertz, C. (2020). Big Data als datenbasierte Programmierung: Eine medienpädagogische Analyse künstlicher neuronaler Netzwerke [Big Data as Data-Based Programming: A Media Pedagogical Analysis of Artificial Neural Networks]. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 93–119. <https://doi.org/10.21240/mpaed/jb17/2020.04.27.X>
- Vos, P. (2011). What is “authentic” in the teaching and learning of mathematical modelling? In G. Kaiser, W. Blum, R. Borromeo Ferri & G. Stillman (Eds.), *Trends in Teaching and Learning of Mathematical Modelling* (ICTMA14) (pp. 713–722). Springer.