Extended Abstract for the
Symposium on Integrating AI and Data Science into School Education Across Disciplines 2025

WHAT UNIVERSITY-LEVEL MATHEMATICS NEEDED TO UNDERSTAND DEEP LEARNING CAN BE MADE ACCESSIBLE IN HIGH SCHOOL?

Martin Frank
Karlsruhe Institute of Technology, Germany, martin.frank@kit.edu

Focus Topics: AI and Data Science Curricula and Implementation in School

Motivation

In this contribution, we would like to analyze the university-level mathematics that the research community currently believes to be crucial for the mathematical understanding of when, how and why supervised deep learning works (Kutyniok, 2024), with respect how much of it ("how much" certainly needs to be specified) can in principle be taught in high school. The analysis assumes that opening the black box of AI and thereby demystifying it empowers students as individual human beings (German *Mündigkeit*). We focus on the non-statistical aspects, not because the statistical aspects are not important, but because they are covered better in other talks at the workshop.

Key mathematical aspects of understanding deep learning

We restrict ourselves to supervised deep learning. Deep learning (DL) here means that we consider deep neural networks. We look at the supervised setting, meaning that we assume that a data set of input-output pairs is given, which is approximated by the deep neural network (NN). It would be worthwhile to analyze other subfields like reinforcement learning (which is closely related to optimal control theory) adversarial training (related to game theory), diffusion models (related to inverse problems), embeddings (which are at the core of transformers) but this is out of our scope. So is the role of different NN architectures, and of other hyperparameters.

Expressivity studies neural networks as a class of functions from the perspective of approximation theory. It is important to understand that on an abstract level NNs are simply a class of parameterized functions, more precisely a concatenation of parameterized elementary functions. A basic question is whether any sufficiently regular function on a compact set can be approximated by an NN (it can). A more relevant question is why NNs in many cases seem to be able to approximate high-dimensional functions so well compared to other classes of functions. High-dimensional problems are notoriously hard for conventional methods. Studying how many parameters are needed for a certain accuracy is called *expressivity*.

To study *training*, one needs first to understand that in the first place this consists of solving an optimization problem. This optimization problem is high-dimensional and highly non-convex. It is typically solved by (versions of) stochastic gradient descent (SGD). SGD is in its core a very simple algorithm. For the computation of the gradients of the loss function with respect to the network parameters it relies on a method called backpropagation (which has in fact been known as the adjoint method). Historically, SGD was used out of necessity to be able to make the solution feasible, but now it appears to be crucial for generalization.

Generalization is arguably the most stunning property of NNs is the double descent curve. Against conventional wisdom, in many cases increasing the number of parameters in the NN beyond the number of data points (therefore using the NN in an overparameterized regime) ultimately leads to smaller testing errors. An interesting, related question is whether these trained NNs "have memorized" their training data.

From a mathematical perspective, *explainabilty* of why a deep neural network yields a certain output is often related to sensitivity analysis.

What can be taught in high school?

At first glimpse, it appears to be impossible to make concepts from ongoing mathematical research accessible to high school students. Our purpose is to demonstrate, though, that the key concepts can in fact be made at least plausible, and that they have multiple connections to standard elements of the German curriculum (but certainly many other international curricula as well). Some examples of practical proofs-of-concept are shown at this workshop.

In S. Podworny & S. Schönbrodt (Eds.), *Towards Fostering AI and Data Science Literacy in Schools Across Disciplines. Proceedings of the 1st Symposium on Integrating AI and Data Science into School Education Across Disciplines (AIDEA1 2025)*, Salzburg, Austria.

Understanding NNs as parameterized classes of functions requires the knowledge of the activation functions (e.g., ReLU, sigmoid, hyperbolic tangent); these are all not standard functions, but not out of scope. A second (standard) concept is concatenation, and then an index notation of a matrix/vector formulation for performing the activations and the affine-linear transformations many times (cf. the talk by Stephan Kindler).

The approximation properties of NNs can be understood in an even easier way than for polynomials, for example by considering ReLU networks of one variable, and realizing that they represent all piecewise affine-linear functions. One can show graphically (and even computationally) that any continuous function on a bounded interval can be approximated arbitrarily well.

Training consists of solving an unconstrained optimization problem; such problems are in principle formulated in calculus (usually in one dimension). We would argue that it is possible to introduce elements multi-dimensional calculus in high schools (it has been in the past), and/or to come up with a version of gradient descent directly when trying to approach the optimization problem algorithmically. We also believe that supervised learning can be effectively approached using linear least squares (Biehler et al., 2024). Teaching backpropagation (which we believe to be hard to teach) is the topic of a contribution by Orit Hazzan.

The problems that appear in high dimensions can also be addressed in the context of optimization problems. Students might want to solve an optimization problem by grid search; this way, they realize that one simply cannot build a nontrivial search grid in high dimensions, because the number of grid points scales exponentially with the dimension.

We currently have no idea how to make the double descent curve accessible, beyond conveying wonder at it. It can certainly be taught that when one increases the number of parameters one eventually enters the interpolation regime, which – according to conventional wisdom – is related to overfitting. But conveying wonder and mentioning an open question might be something interesting in its own right.

Outlook

Summarizing, we believe that many aspects of AI can and should be demystified in high-school mathematics education. This has a lot to do with language, showing that mystically sounding AI terms often have a down-to-earth classical Math equivalent.

Another interesting topic that we have not touched upon is the connection to mathematical modeling, which is the use of mathematics with a specific purpose to solve a real-world problem (Schönbrodt & Frank, 2024). Key to this is the active use of mathematics, rather than the mechanical reproduction of mathematical techniques, which is so often at the core of high-school mathematics.

Finally, we would like to stress that we believe that demystifying AI/fostering AI literacy is an interdisciplinary challenge that from the methods side requires statistics, mathematics, and computer science, but cannot be understood without its impact in the sciences and the humanities.

References

Biehler, R., Schönbrodt, S., & Frank, M. (2024). KI als Thema für den Mathematikunterricht. *mathematik lehren, Volume 244*, 2–7.

Kutyniok, G. (2024). The Mathematics of Reliable Artificial Intelligence. SIAM News, https://www.siam.org/publications/siam-news/articles/the-mathematics-of-reliable-artificial-intelligence/

Podworny, S., Biehler, R., & Fleischer, Y. (2025). Young students' engagement with data to create decision trees. ZDM–Mathematics Education, 1-17.

Schönbrodt, S., & Frank, M. (2024). Wie viel Mathe steckt in mathematischer Modellierung? – Eine Antwort am Beispiel der Optimierung. *Der Mathematikunterricht, Volume 70*, 35-44.

Schönbrodt, S., & Oldenburg, R. (2024). Mathematische Grundlagen als Schlüssel zu einem allgemeinbildenden Verständnis von KI: Theoretische Perspektiven und praktische Unterrichtsideen. *Mathematik im Unterricht, 15*, 17-36.