

Stephanie HOFMANN, Karlsruhe & Martin FRANK, Karlsruhe

Maschinelles Lernen im Schulunterricht am Beispiel einer problemorientierten Lerneinheit zur Wortvorhersage

Technologien aus dem Bereich des Maschinellen Lernens (ML) erleben derzeit durch den exponentiellen Anstieg der produzierten Daten und der enorm gestiegenen Rechenleistung ihren Aufstieg und werden zukünftig vermehrt Empfehlungen und Vorhersagen automatisiert treffen können. Die wachsende Bedeutung von ML für die Gesellschaft erfordert eine Thematisierung dieser Technologien und Fragestellungen auch im Unterricht. Die Relevanz von ML für Gegenwart und Zukunft der Lernenden sowie die Möglichkeit der didaktischen Reduktion einiger ML-Methoden auf schulmathematische Inhalte der Sekundarstufe unterstreichen diese Forderung. Gleichzeitig kann das Verständnis von und der Umgang mit Daten an Problemstellungen im Bereich ML geschult werden. Im Rahmen des Projektes CAMMP (www.cammp.online) entstand daher eine Lerneinheit, die Schüler*innen ab Klasse 10 am Beispiel der Wortvorhersage aufzeigt, wie maschinelle Lernmethoden genutzt werden können, um Vorhersagen aus Daten zu treffen.

Maschinelles Lernen als Unterrichtsthema

Die Forderung nach der Integration des Themas ML in den Bildungsplan wurde bereits mehrfach gestellt sowie diverse Argumente dafür sowohl aus stochastisch didaktischer Perspektive als auch aus Sicht der Modellierungs- didaktik dargelegt (Gould, 2021; Opel et al., 2019; Schönbrodt et al., 2022). Fest steht allerdings, dass viele Lehrkräfte sich nicht befähigt fühlen Kurse zu ML anzubieten (Davies & Sheldon, 2021, S. 64). Unser Ziel ist es deshalb Lehrkräfte mit erprobtem Lehr- und Lernmaterial zu unterstützen, welches maschinelle Lernmethoden an realen und authentischen Problemstellungen einführt. Im Vergleich zu digitalen Lernmaterialien zu ML von Biehler & Fleischer (2021) sowie Opel et al. (2019), welche insbesondere Kompetenzen aus Statistik und Informatik in den Fokus nehmen, konzentriert sich die in diesem Beitrag vorgestellte Lerneinheit auf die computergestützte mathematische Modellierung. Strategien des ML und andere mathematische Inhalte werden ausgehend von einem Problem eingeführt.

Um an die Alltagserfahrungen der Schüler*innen anknüpfen zu können, werden in der Lerneinheit ML-Methoden am Beispiel prädiktiver Textsysteme, insbesondere der Wortvorhersage betrachtet. Diese gibt Nutzer*innen beim Schreiben einer Nachricht Vorschläge für das nächste Wort. Im Workshop wird das N-Gramm-Konzept verwendet und die Strategie des überwachten Lernens genutzt: Für den Lernprozess des Modells sind sowohl Eingangsdaten als auch Ausgangsdaten in Form von Beispieltextrn gegeben.

Digitales interaktives Unterrichtsmaterial

Das Lernmaterial ist für Gymnasialschüler*innen ab der zehnten Klasse geeignet. Der Workshop setzt Vorkenntnisse zu relativen und absoluten Häufigkeiten voraus. Darüber hinaus sollten die Lernenden Wahrscheinlichkeiten berechnen und einfache mehrstufige Zufallsexperimente durchführen und auswerten können. Für die Kombination und Bewertung der Modelle werden Kenntnisse über das arithmetische Mittel und den Logarithmus einer Zahl vorausgesetzt. Programmierkenntnisse sind nicht erforderlich.

Das Lernmaterial kann in einem kompakten ein- bis zweitägigen Workshop oder aufgeteilt in mehrere Lektionen einer Unterrichtseinheit (4 - 6 Doppelstunden) sowohl im Präsenz- als auch im Online-Unterricht eingesetzt werden. Besonders eignet sich das Material zum Einstieg oder als Wiederholung des Themas bedingte Wahrscheinlichkeiten. Die Lerneinheit ist in Form von digitalen Arbeitsblättern, sogenannten Jupyter Notebooks, umgesetzt, die über eine vom Karlsruher Institut für Technologie gehostete Cloud-Plattform abgerufen werden können. Der Login-Prozess zur Plattform ist unter www.cammp.online/214.php beschrieben. Jupyter Notebooks ermöglichen die übersichtliche Anordnung von Arbeitsaufträgen, Formeln, Abbildungen, aber auch Codefeldern in einem Dokument sowie eine interaktive Bearbeitung durch die Lernenden. Durch digitales Differenzierungsmaterial wie gestufte Hilfen, Vertiefungsaufgaben und weiterführende Informationsblätter ist das Lernmaterial auch für heterogene Lerngruppen sehr gut geeignet.

Didaktische Umsetzung

Zu Beginn der Unterrichtseinheit erarbeiten die Lernenden verschiedene N-Gramm-Modelle. N-Gramm-Modelle schätzen die Wahrscheinlichkeit, dass ein Wort w_N auf eine gegebene Wortsequenz w_0, \dots, w_{N-1} folgt, aus der relativen Häufigkeit der Wortsequenz w_0, \dots, w_N in einem Datensatz:

$$P(w_N | w_0, \dots, w_{N-1}) = \frac{N(w_0, \dots, w_N)}{\sum_x N(w_0, \dots, w_{N-1}, w_x)}$$

Das Wort w_N , das mit der größten Wahrscheinlichkeit auf die gegebene Wortsequenz folgt und dessen sogenannte Übergangswahrscheinlichkeit somit am größten ist, kann anschließend als Vorschlag ausgegeben werden.

In der Lerneinheit entwickeln die Schüler*innen zunächst eine Strategie zur Schätzung der Übergangswahrscheinlichkeiten aus einem Datensatz unter Verwendung der Häufigkeit des Auftretens von Zweier-Wort-Sequenzen (Bi-Grammen). Bedingte Wahrscheinlichkeiten werden problemorientiert eingeführt. So taucht in einem Beispieltext in zwei von drei Fällen nach dem Wort „Ich“ das Wort „mag“ auf. Die Wahrscheinlichkeit, dass das Wort

„mag“ nach dem Wort „Ich“ auftritt wird somit zu 2/3 geschätzt. Für die Lernenden wird dabei ersichtlich, dass es sich bei der Verwendung der relativen Häufigkeiten nur um eine Schätzung der bedingten Wahrscheinlichkeit handelt, die umso besser wird, je mehr Daten betrachtet werden. Der im Lernmaterial verwendete Datensatz mit ca. 350.000 Wörtern besteht aus den deutschsprachigen Texten des Korpus „What's up, Switzerland“ (Stark et al., 2014-2020) und den Texten der Kategorie „Belletristik“ des Korpus „LIMAS“ (Forschungsgruppe LIMAS, 1970-1971).

Beim Testen des Bi-Gramm-Modells mit unterschiedlichen Wortsequenzen haben die Lernenden anschließend die Aufgabe, verschiedene Probleme des Modells zu identifizieren und Verbesserungen vorzuschlagen. Das stärker kontextbezogene Tri-Gramm-Modell (Modell basierend auf der Betrachtung von Dreier-Wort-Sequenzen) sowie das Uni-Gramm-Modell (Modell basierend auf der Betrachtung von Einer-Wort-Sequenzen) werden daraufhin als Modellverbesserungen erarbeitet. Bei der Diskussion von Vor- und Nachteile der drei N-Gramm-Modelle, sollen die Lernenden erkennen, dass alle Modelle unterschiedliche Stärken und Schwächen haben und die N-Gramm-Modelle für eine verlässliche und zugleich kontextbasierte Vorhersage des nächsten Wortes kombiniert werden müssen. Eine typische Kombinationsmöglichkeit ist die Interpolation der einzelnen Modelle (Wendemuth et al., 2004, S. 30-31). Der neue Schätzer für die interpolierte Übergangswahrscheinlichkeit des Wortes w_2 mit der Wortvorgeschichte w_0, w_1 ergibt sich als gewichteter Mittelwert der einzelnen N-Gramm-Wahrscheinlichkeiten zu

$$\tilde{P}(w_2|w_0, w_1) = g_1 \cdot P(w_2) + g_2 \cdot P(w_2|w_1) + g_3 \cdot P(w_2|w_0, w_1).$$

Die Frage nach der optimalen Wahl der Interpolationsgewichte g_1, g_2 und g_3 motiviert das Definieren eines geeigneten Fehlermaßes. Im Unterrichtsgespräch werden Ideen der Schüler*innen zur Bewertung von Sprachmodellen gesammelt. Wichtig in diesem Zusammenhang ist die Unterscheidung zwischen Trainingsdaten, die zur Schätzung der Übergangswahrscheinlichkeiten genutzt werden, und unbekanntem, d. h. nicht zur Erzeugung des Modells verwendeten Testdaten, die zur Bestimmung der Güte des Sprachmodells eingesetzt werden. Anschließend haben die Lernenden die Möglichkeit, selbstständig ein Optimierungsverfahren zur Minimierung des Fehlermaßes in Abhängigkeit von den Gewichten zu entwickeln oder, stärker angeleitet, nach dem Aufstellen der Nebenbedingungen und der Zielfunktion des Optimierungsproblems ein Optimierungsverfahren als Black Box einzusetzen.

Auf einem weiteren Arbeitsblatt gibt es die Möglichkeit, einen zweiten Datensatz, bestehend aus Werken von Johann Wolfgang von Goethe (Goethe, 1774a, 1774b, 1796, 1809), für das Training des Sprachmodells zu verwenden und die Güte beider Modelle zu vergleichen. Ziel ist es, die Bedeutung

eines geeigneten Trainingsdatensatzes, der die Anwendungssituation bezüglich des Vokabulars und der Textart passend repräsentiert, herauszuarbeiten.

Erfahrungen und Fazit

Das Lernmaterial zeigt wie reale Probleme, die mit ML-Methoden gelöst werden, für Schüler*innen aufbereitet werden können. Es wurde bereits mit über 50 Lernenden der Klassen 10 bis 13 sowohl in Workshops mit einzelnen interessierten Schüler*innen als auch im Kursverband als Unterrichtsreihe (4 Doppelstunden) eingesetzt. Das Feedback zeigt, dass die Schüler*innen ein großes Interesse an der Problemstellung haben. Dass die Lerneinheit zu einem besseren Verständnis und einem realistischeren Bild des Themas ML führt, unterstreicht das Zitat eines Teilnehmers: „Dieser Workshop hat für mich zur Entmythologisierung von Künstlicher Intelligenz beigetragen“.

Literatur

- Biehler, R. & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, 43(S1). <https://doi.org/10.1111/test.12279>
- Davies, N. & Sheldon, N. (2021). Teaching statistics and data science in England's schools. *Teaching Statistics*, 43(S1). <https://doi.org/10.1111/test.12276>
- Forschungsgruppe LIMAS (Hrsg.) (1970-1971). *LIMAS-KORPUS (Linguistik und Maschinelle Sprachbearbeitung)*. Universität Bonn, Universität Regensburg.
- Goethe, J. W. von. (1774a). *Die Leiden des jungen Werthers*. Bd. 1. Deutsches Textarchiv.
- Goethe, J. W. von. (1774b). *Die Leiden des jungen Werthers*. Bd. 2. Deutsches Textarchiv.
- Goethe, J. W. von. (1796). *Wilhelm Meisters Lehrjahre*. Bd. 4. Deutsches Textarchiv.
- Goethe, J. W. von. (1809). *Die Wahlverwandtschaften*. Bd. 1. Deutsches Textarchiv.
- Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, 43(S1). <https://doi.org/10.1111/test.12267>
- Opel, S., Schlichtig, M., Schulte, C., Biehler, R., Frischemeier, D., Podworny, S. & Wassong, T. (2019). Entwicklung und Reflexion einer Unterrichtssequenz zum Maschinellen Lernen als Aspekt von Data Science in der Sekundarstufe II. In A. Pasternak (Hrsg.), *Informatik für alle* (S. 285–294). Gesellschaft für Informatik. <https://doi.org/10.18420/infos2019-c14>
- Schönbrodt, S., Camminady, T. & Frank, M. (2022). Mathematische Grundlagen der Künstlichen Intelligenz im Schulunterricht. *Mathematische Semesterberichte*, 69(1), 73–101. <https://doi.org/10.1007/s00591-021-00310-x>
- Stark, E., Ueberwasser, S. & Göhring, A. (2014-2020). *Korpus "What's up, Switzerland?"*. Universität Zürich. www.whatsup-switzerland.ch
- Wendemuth, A., Andelic, E., Barth, S., Dobler, S., Katz, M., Krüger, S., Maiwald, M., Mamsch, M. & Schafföner, M. (2004). *Grundlagen der stochastischen Sprachverarbeitung*. Oldenbourg Wissenschaftsverlag GmbH. <https://doi.org/10.1524/978>