



V17 – Datenbasierte Modelle für komplexe physikalische Zusammenhänge (Grundlagen maschinellen Lernens)

Ort: MAIN-Labor C50.239 (Simulation von Nanomembranen, Hybridmaterialien und Grenzflächen)

Betreuer: Dr. Florian Fuchs

Die Beschreibung physikalischer Zusammenhänge, entweder über Experimente oder numerischen Simulationen, ist oft mit hohem Kosten und Zeitaufwand verbunden. Eine vielversprechende Alternative stellen datenbasierte Ersatzmodelle dar. Hierbei handelt es sich um vergleichsweise recheneffiziente Modelle, welche wesentliche Charakteristika eines (physikalischen) Zusammenhanges abbilden können und im Vorfeld auf Basis eines anderweitig bestimmten Datensatzes trainiert werden. Ersatzmodelle kommen zunehmend in den Materialwissenschaften und der Industrie zum Einsatz und stellen daher ein vielseitiges Werkzeug sowohl für Experimentatoren und auch für Theoretiker dar.

Im vorliegenden Versuch sollen die grundlegenden Fähigkeiten zur Arbeit mit datenbasierten Ersatzmodellen vermittelt werden. Dafür werden Datensätze zur Verfügung gestellt, die anschließend unter Verwendung verschiedener Ansätze als Ersatzmodell abgebildet werden. Es kommen Ansätze aus freien Bibliotheken zum Einsatz. Diese Bibliotheken werden über die Programmiersprache Python angesprochen, jedoch sind rudimentäre Programmierkenntnisse für den Versuch ausreichend.

Datenbasiertes Ersatzmodell, Maschinelles Lernen, Regression, Neuronale Netzwerke, Über- und Unteranpassung, Kernel-Trick, Hauptkomponentenanalyse

1. Analyse von Datensätzen

- 1.1 Beurteilen Sie die zur Verfügung gestellten Datensätze hinsichtlich ihrer Qualität. Betrachten Sie hierzu zum Beispiel das Rauschen der Daten, die Dimensionalität und die Verteilung der Daten im Parameterraum. Verwenden Sie die Hauptkomponentenanalyse (engl. principal component analysis, PCA) um die wesentlichen Parameter der jeweiligen Datensätze zu finden.
- 1.2 Entscheiden Sie welche Art von datenbasiertem Modell für die jeweiligen Datensätze geeignet sind. Entscheiden Sie zwischen Lasso/Ridge Regression, Kernel Regression, Support Vector Machines, Decision Tree und Neuronales Netzwerk. Begründen Sie Ihre Entscheidung.



2. Regressionsmethoden

- 2.1 Teilen Sie einen zur Verfügung gestellten Datensatz in einen Trainings- einen Validierungs- und eines Testdatensatz auf. Welche Aufgabe haben diese Teildatensätze?
- 2.2 Verwenden Sie skikit-learn [1] um ein datenbasiertes Ersatzmodell zu realisieren. Vergleichen Sie verschiedene Algorithmen. Stellen Sie den Lernfortschritt im Laufe des Trainings dar.
- 2.3 Beurteilen Sie die Ersatzmodelle. Betrachten Sie hierzu auch die Metriken absoluter Fehler (engl. mean absolute error, MSE), mittlerer quadratischer Fehler (engl. mean squared error, MAE), (engl. root mean squared error, RMSE) und das Bestimmtheitsmaß (R^2).
- 2.4 Diskutieren Sie die Grenzen der verwendeten Modelle.

3. Neuronales Netzwerk

- 3.1 Beschreiben Sie die verschiedenen Bestandteile neuronaler Netze. Gehen Sie dabei auch auf das sog. Residuale Neuronale Netz (engl. residual neural network, ResNet) ein.
- 3.2 Verwenden Sie z.B. Keras [2] um ein neuronales Netzwerk zur Bilderkennung zu realisieren und wenden Sie dieses für einen zur Verfügung gestellten Datensatz an.
- 3.3 Optimieren Sie die Topologie des neuronalen Netzwerks um eine möglichst hohe Erfolgsquote des Modells zu erzielen
- 3.4 Diskutieren Sie die Nachteile bei der Verwendung von neuronalen Netzwerken.

Machen Sie sich mit den oben genannten Stichworten anhand der unten angegebenen Quellen [8] und [4] vertraut.

- [1] Scikit-Learn: <https://scikit-learn.org/>
- [2] Keras: <https://keras.io/>
- [3] Kaggle: <https://www.kaggle.com/>
- [4] James, Gareth, et al. An introduction to statistical learning. New York: Springer, 2013. Kostenloses E-Book: <https://www.statlearning.com/>
- [5] Goodfellow, Ian, Yoshua Bengio und Aaron Courville. Deep learning. MIT press, 2016. Kostenloses E-Book: <https://www.deeplearningbook.org/>
- [6] Raschka, Sebastian und Vahid Mirjalili. *Machine Learning mit Python und Scikit-Learn und TensorFlow*. mitp-Verlag, 2017. Kostenloses E-Book über die Uni-Bibliothek: <https://katalog.bibliothek.tu-chemnitz.de/Record/0-1668498855>
- [7] Towards data science: <https://towardsdatascience.com/>
- [8] Tom Rothe, Masterarbeit, 2021, Kap. 3, <https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-756307>