Report

of

Exercise sheet 3: Machine Learning

for

269002 VO Computational Concepts in Physics I (2021W)

Student name:	Clemens Wager, BSc
Matriculation nr:	01635477
Degree program:	Master of Computational Science
Supervisors:	Univ.-Prof. Dr. techn. Cesare Franchini,
	Dott. mag. Dr. Michele Reticcioli

Vienna, 30.01.2022

Introduction

The task was to optimize an Artificial Neural Network (ANN) model to fit the provided data points. The given datasets were theoretical measurements of a damped oscillation affected by noise. The main goal was to predict certain datapoints that were missing within the initial datasets (interpolation) and to predict the further course of the damped oscillation function (extrapolation). The performance of the ANN model was assessed with two given test datasets. Afterwards, a general conclusion about the applicability of ANN models and their limits was drawn.

Method

Theory

Supervised learning infers a function (in this case a damped oscillation function) from labeled training data. All datapoints are labeled as either input or output and the algorithm connects input-output pairs. The process of supervised learning constitutes of three different phases each with their own dataset: Training, Validation, Test.

During training, the algorithm processes the training dataset (N=715) and tries to fit a regression model into the data. In this phase various hyper-parameters are tuned. In this experiment an artificial neural network was developed, and the task was to tune the number of hidden layers, the number of neurons per layer, the number of epochs, and the learning rate. Further, the tolerance[1] and the alpha value[2] were adjusted.

In the validation phase, the goodness of fit is assessed on a different set (N=209). The model should predict the data in this phase correctly without over- or underfitting. If this is not the case, that training starts again to refine the model. Lastly, the model should be able to predict the two test datasets (N=96, N=212) without further adjustments.

ANN ARCHITECTURE: The model consists of one input layer with 60 nodes, two hidden layers with 40 nodes and an output layer with 40 nodes. Each node uses the activation function ReLU (Rectified Linear Unit) to process its input. Network weights are updated with the solver “Adam”, which is an optimization algorithm. The tolerance was adjusted to 1e-6 and the alpha value to 1e-5. Finally, the learning rate remained constant with the standard value 1e-3 (scikit).

The model’s performance is assessed with two error measurements. The MAE (Mean Absolute Error) is the error given by

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}|$$,

where N is the number of datapoints y_i the true value and $\hat{y}_i$ the predicted value. The MSE (Mean Squared Error) is the squared value of the MAE. The *score* is the model’s coefficient of determination (R^2) and measures the goodness of fit.

[1] tolerance → stops training when the loss or score is not improving by at least the tol (sklearn, 2022)
[2] alpha value → L2 penalty (regularization term) parameter (sklearn, 2022)

Implementation

The ANN was implemented in Python 3.7 and uses the libraries numpy and matplotlib.pyplot, pandas, time and sklearn.neural_network. The ANN model is based on the MLPRegressor (Multi-layer Perceptron regressor) of the scikit library.

When executing the program, the training validation datasets are read in and stored in variables. Then the model is created with the already optimized hyper-parameters and the training phase starts. The resulting prediction model is checked against the validation data and a plot shows the results of the validation phase. Also, information on the quality of the model is displayed in the terminal.

Afterwards, the model is tested on the "data-test.1.dat" dataset and again a corresponding plot and information is displayed. Lastly, the model is checked on the "data-test.2.dat" dataset analogously.

The information displayed in the terminal covers MAE, MSE, model score and the number of epochs.

Results and Discussion

The ANN model went through the training and the validation phase to optimize its hyper-parameters. Finally, it achieved a score of 0.59 with reasonable MAE/MSE values (see table 1 in appendix). This score is not very high, but the resulting regression function (red line in figure 1) fits the data sufficiently as seen in figure 1 without over- or underfitting. Considering the theoretical situation of a damped oscillation the course of the prediction function makes sense.

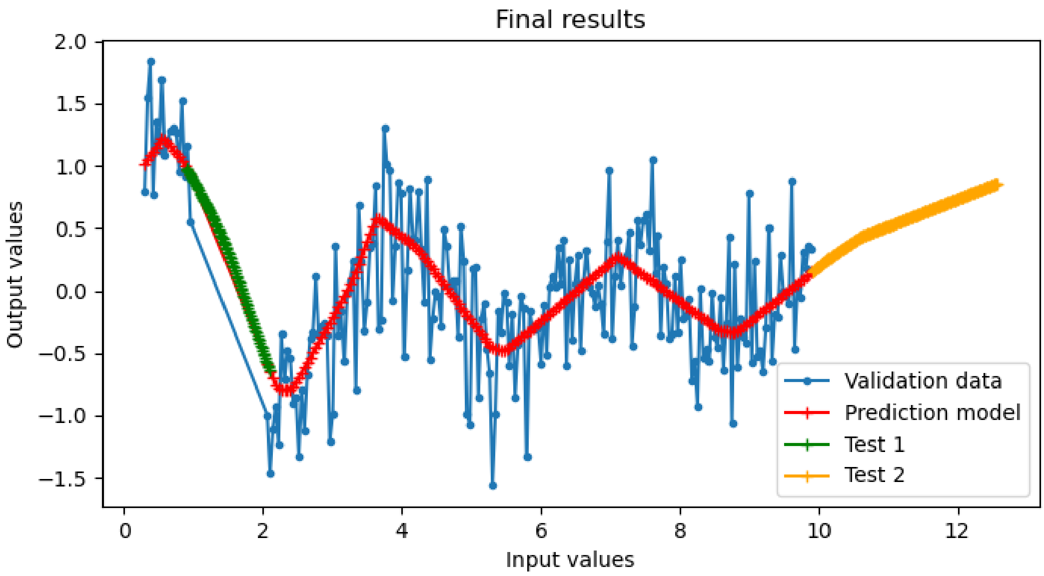

Figure 1 shows the final results of the model development. The blue graph represents the validation data. The red line illustrates the inferred function of the prediction model. The green line is the result of test one and the orange line is the result of test two.

When looking at the result of test one (see figure 2), it becomes obvious that the prediction was very accurate. The model scored 0.94 in this test with even better MAE/MSE values (see table 1 in appendix) than during validation phase. The dataset of test one was taken from an interval within the range of the training data at about $x = [0.9, 2.0]$. Thus, the model managed to interpolate the range of known datapoints and predict the missing datapoints in the interval (see figure 1 green line).

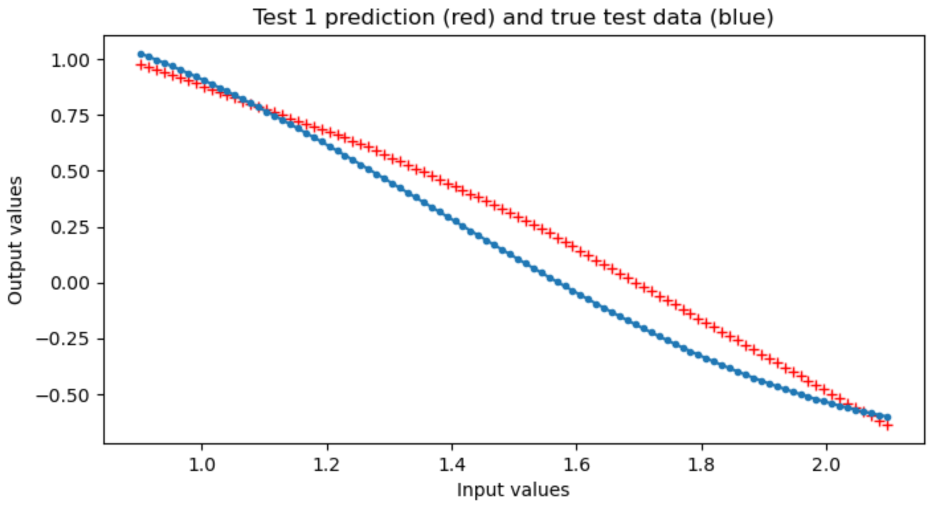

Figure 2 shows the results of test 1. The red line is the prediction, and the blue line is the true test data. The model score was 0.94 indicating a visibly good fit.

When turning to the result of test two (see figure 3) we find that the ANN model failed to predict the further course of the function. The range of the test two dataset about $x = [10, 12.5]$ lied beyond the range of the training set. Hence, the algorithm had to extrapolate the missing datapoints. This test illustrates how the model is not capable of successfully predicting values that were not covered in the training set's range.

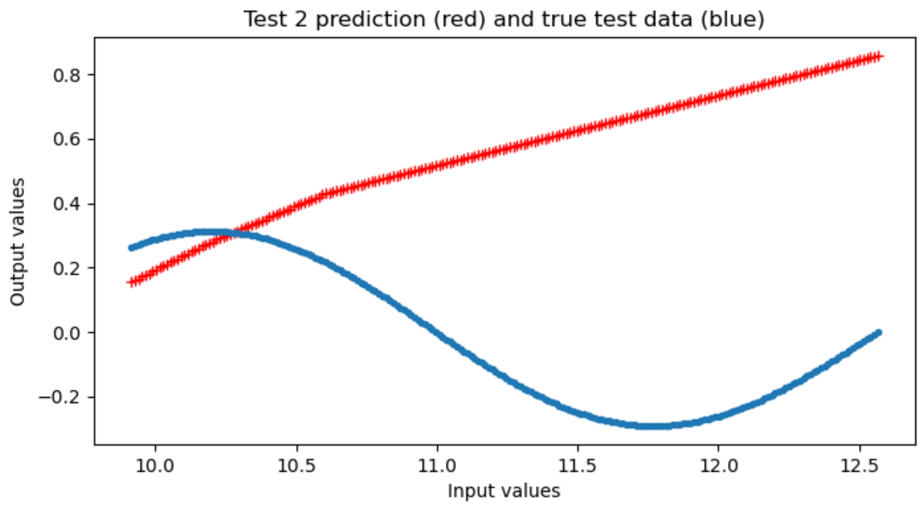

Figure 3 shows the results of test 2. The red line is the prediction, and the blue line is the true test data. The model could not fit the test data as the two diverging lines indicate.

Conclusion

In this computer experiment a machine learning model with an ANN architecture was developed. Two tests were carried out to evaluate its performance in a task of interpolation and an extrapolation. Without any adjustments in the test phase, it was impossible to satisfyingly fit the theoretical model as stored in both test data sets. This behavior leads to the conclusion that the ANN model cannot be applied on types of data which it was not trained on. Extrapolation poses a problem for the ANN model, but it could possibly be implemented by changing the activation function as suggested in a paper by Liu Ziyin et al. (Ziyin, et al., 2020). Overall accuracy of the model could be improved by providing more data or performing cross-validation.

Sources

sklearn, 2022. *scikit-learn.org.* [Online]
Available at: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html
[Accessed 29 January 2022].

Ziyin, L., Hartwig, T. & Ueda, M., 2020. *Neural Networks Fail to Learn Periodic Functions and How to Fix It.* [Online]
Available at: https://arxiv.org/abs/2006.08195

Appendix

	Validation	Test 1	Test 2
MAE	0.3229	0.1077	0.5903
MSE	0.1641	0.0162	0.4835
Score	0.5854	0.9389	-8.9031
Epochs	232	232	232

Table 1 shows the ANN model's performance in validation phase, test one and test two.