

# Group 10, Assignment 1

Bulkens, Hecken, Kisin, Schmidle, Schmidt, Streitberger

May 2, 2023

## **1 Review on the paper MLPerf Tiny Benchmark (Banbury et al. 2021)**

The primary contribution of this paper is presenting a modular and therefore flexible way to benchmark inference of machine learning models in the special case of tiny ML (ML with super low energy consumption). The benchmark offers a open source modular build inference pipeline with reference implementation for each part and enables the submitter to modify every part as needed to be able to show end-to-end concepts as well as single specific adjustments. The benchmark tracks accuracy, latency and energy consumption of the inference task and aims to make different approaches comparable.

Key insight would be that the submissions for the benchmark since its release showed a clear trend in model setting choices e.g. towards 8-bit integers but not towards any specific framework or hardware. The benchmark offers a good basis for comparison but is still limited as e.g. streaming and preprocessing steps are not included as it is hard to define when preprocessing even begins and inference starts.

We personally like the idea of such a modular and therefor flexible pipeline setup with predefined tasks and datasets to compare techniques with each other, since such benchmarks, and their standardization and regularization, are the best source for comparing novel approaches with others. Papers in the ML research area are predestined to tune their numbers and compare themselves with other models in a way that makes the proposition look good. This benchmark, although limited in capability (no preprocessing, no streaming, only 4 tasks, etc.) offers a relatively fair comparison between approaches as well as the capability of only tuning very specific parts of the whole pipeline in general to show improvement independent of the rest. It can still be improved we would accept this paper, as it offers a well regulated and structured benchmark.

## **2 Willingness to present:**

Willing to present both the paper review (1.1) as well as the coding exercises (1.2)

## **3 Plots from the code**

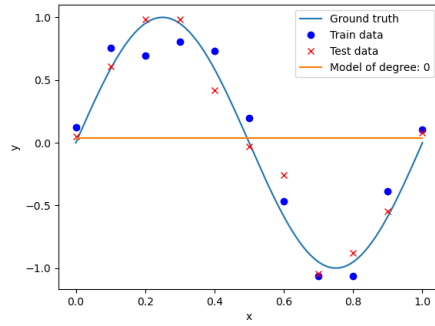


Figure 1: Polynomial Fit Degree 0

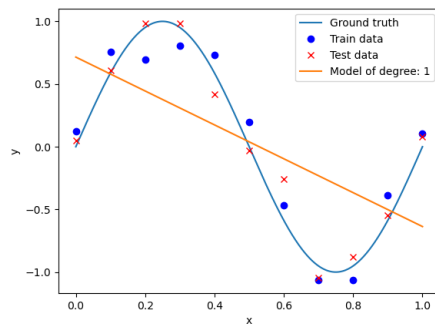


Figure 2: Polynomial Fit Degree 1

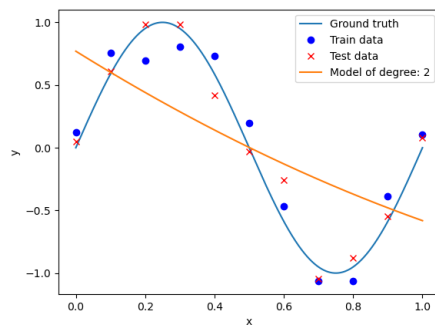


Figure 3: Polynomial Fit Degree 2

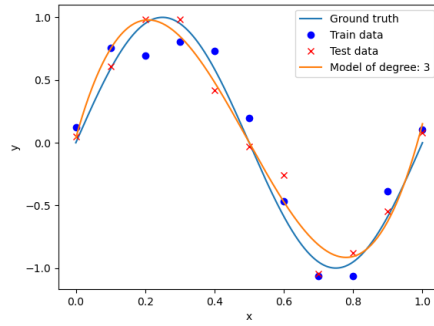


Figure 4: Polynomial Fit Degree 3

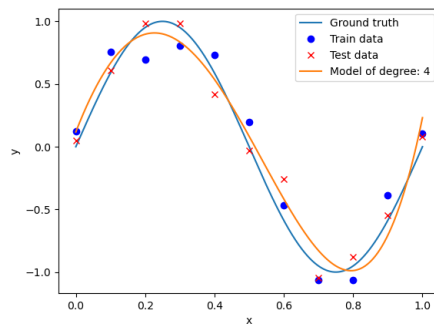


Figure 5: Polynomial Fit Degree 4

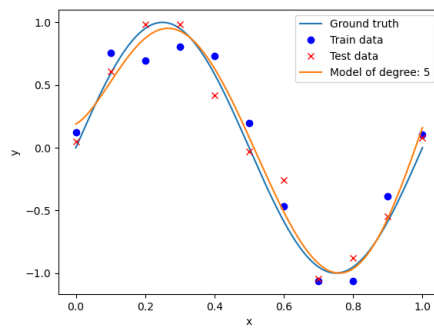


Figure 6: Polynomial Fit Degree 5

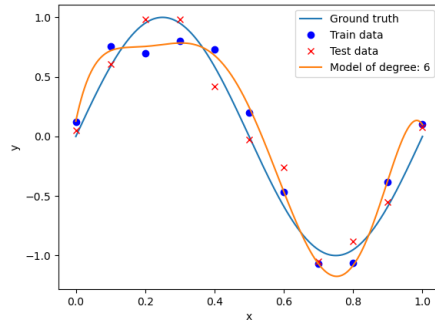


Figure 7: Polynomial Fit Degree 6

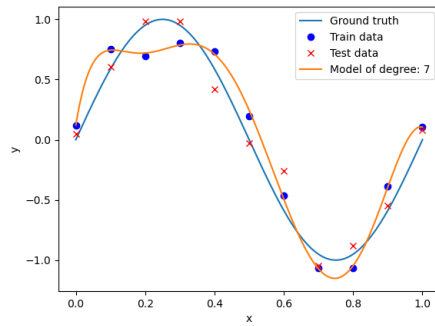


Figure 8: Polynomial Fit Degree 7

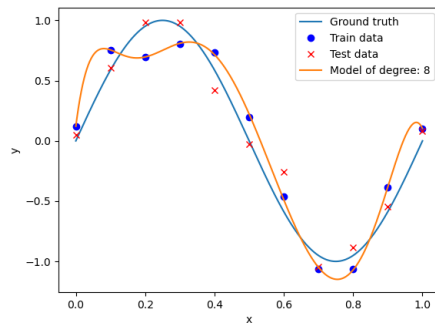


Figure 9: Polynomial Fit Degree 8

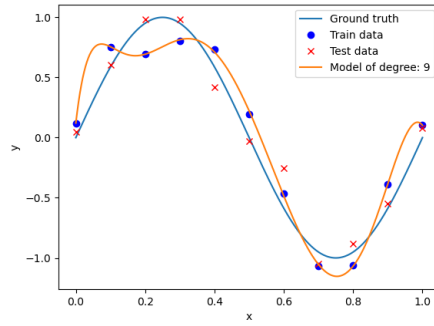


Figure 10: Polynomial Fit Degree 9

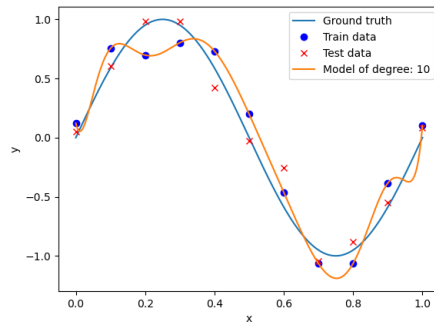


Figure 11: Polynomial Fit Degree 10

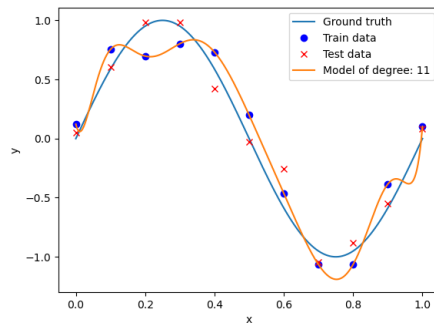


Figure 12: Polynomial Fit Degree 11

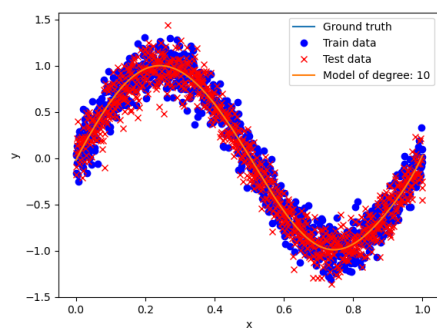


Figure 13: Higher Data Population