CPSC 488 – Assignment 1 Report

Supervised Learning with HPC

Ahmad Aldasouqi

09-21-2025

1 Hardware Environment

The experiments were executed on the Nautilus Kubernetes cluster. The node used had the following specifications:

- **CPU:** AMD EPYC 7551P, 32 cores / 64 threads, 2.0 GHz
- **System Memory:** 252 GB DDR4 RAM
- **GPUs:** NVIDIA GeForce RTX 2080 Ti (11 GB GDDR6, 4352 CUDA cores, 420.2 GFLOP/s FP64, 616 GB/s memory bandwidth)
- **Operating System:** Ubuntu 22.04.5 LTS
- **CUDA:** CUDA 12.8

2 Least Squares

2.1 Results

Order	Train RMSE	Train R^2	Train Time (s)	Test RMSE	Test R^2
1	1.272e-01	0.746	0.0183	2.204e-01	0.235
2	1.050e-01	0.826	0.0693	2.992e-01	-0.409
3	8.810e-02	0.878	0.2493	1.114e+00	-18.54
4	7.977e-02	0.900	0.7507	3.311e+00	-171.62
5	7.776e-02	0.905	2.0804	1.510e+00	-34.92

2.2 Best Model

The best model under Least Squares is order 1, since higher orders overfit and lead to negative R^2 on test data.

$$Idx = 66.791555 + 18.736199 \cdot T - 1.560921 \cdot P - 14305.229820 \cdot TC - 27.284839 \cdot SV$$

2.3 Operation counts and theoretical runtime estimates

We can estimate the number of operations to compute the closed form least squares solution for a polynomial of order k with the following.

Let d = number of expanded features and n = the size of the training set (336,000).

1. compute $X^T X$ (dot-products): $\frac{d(d+1)}{2} \cdot n$ multiplications and $\frac{d(d+1)}{2} \cdot (n-1)$ additions,
2. invert / solve the $d \times d$ normal equations: $\approx \frac{2}{3}d^3$ multiplications and similar additions,
3. compute $X^T y$: $d \cdot n$ multiplications and $d \cdot (n-1)$ additions,
4. final multiply $(X^T X)^{-1}(X^T y)$: d^2 multiplications and $d(d-1)$ additions.

For our problem ($m = 4$ original features) the number of expanded features is:

$$d = 1 + \sum_{r=1}^{k} \binom{m+r-1}{r}$$

so $d = 5$ for order 1 and $d = 126$ for order 5.

Below are example numeric totals (multiplications, additions, total ops) for order 1 and 5.

- **Order 1 ($d = 5$):** multiplications $\approx 6.72 \times 10^5$, additions $\approx 6.72 \times 10^5$, **total ops** $\approx 1.34 \times 10^6$.
- **Order 5 ($d = 126$):** multiplications $\approx 2.74 \times 10^8$, additions $\approx 2.74 \times 10^8$, **total ops** $\approx 5.49 \times 10^8$.

To approximate the runtime, we can divide the total ops by the peak device FLOPs rate (1.024 TFLOP/s on the AMD EPYC 7551P).

- **Order 1 ($\mathbf{1.344 \times 10^6}$ FLOPs)**: 1.31×10^{-6} s (1.024 TFLOP/s).
- **Order 5 ($\mathbf{5.488 \times 10^8}$ FLOPs)**: 5.36×10^{-4} s (1.024 TFLOP/s).

These times vary greatly from the observed times to compute the least squares solution. This is because the estimate assumes continuous usage of resources, when in reality there are many bottlenecks such as memory bandwidth, data movement, and software overhead.

These calculations do not give realistic runtime estimates, but they can be used to quantify the difference in runtime between different orders.

3 Gradient Descent

3.1 Results

Order	Train RMSE	Train R^2	Train Time (s)	Test RMSE	Test R^2
1	1.817e-01	0.479	49.15	2.053e-01	0.336
2	1.241e-01	0.757	63.55	1.421e-01	0.682
3	1.304e-01	0.732	79.05	1.663e-01	0.565
4	1.364e-01	0.707	101.59	1.525e-01	0.634
5	1.416e-01	0.684	128.03	1.438e-01	0.674

The best Gradient Descent model is order 2, with the highest test R^2 (0.682).

4 Multilayer Perceptron (MLP)

4.1 Structure and Hyperparameters

- Architecture: 2 hidden layers, 10 neurons each
- Learning rate: 0.001
- Initialization: Kaiming uniform
- Activation: ReLU
- Optimizer: AdamW

4.2 Results

Order	Train RMSE	Train R^2	Train Time (s)	Test RMSE	Test R^2
1	2.140e-01	0.277	24.86	2.221e-01	0.223
2	1.435e-01	0.675	20.69	1.691e-01	0.550
3	1.295e-01	0.735	22.86	1.394e-01	0.694
4	1.414e-01	0.685	19.72	1.522e-01	0.635
5	1.441e-01	0.673	18.64	2.394e-01	0.098

The best MLP model is order 3, achieving the highest test R^2 (0.694).

5 Summary and Discussion

- Least Squares: severe overfitting beyond order 1. Only the linear model generalizes (test $R^2 = 0.235$).
- Gradient Descent: robust, best at order 2 ($R^2 = 0.682$).

- MLP: competitive, best at order 3 ($R^2 = 0.694$).
- Overall: iterative and neural approaches outperform closed-form LS in terms of generalization.