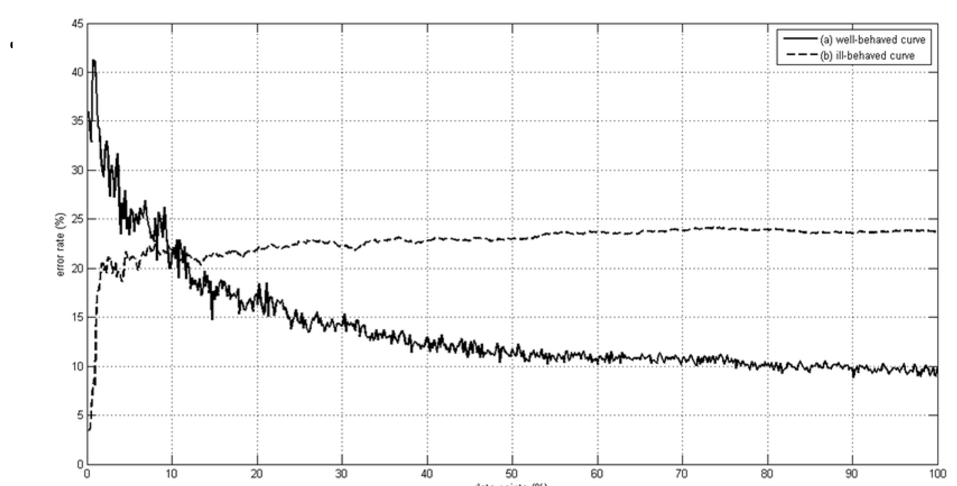
Starting Right: Exploring the impact of random distribution sampling on initial Parameter selection for curve fitting

Author:
Dan Darie
D.Darie@student.tudelft.nl
Responsible professor:
Tom J. Viering
Supervisors:
Cheng Yan, Taylan Turan

1. Introduction

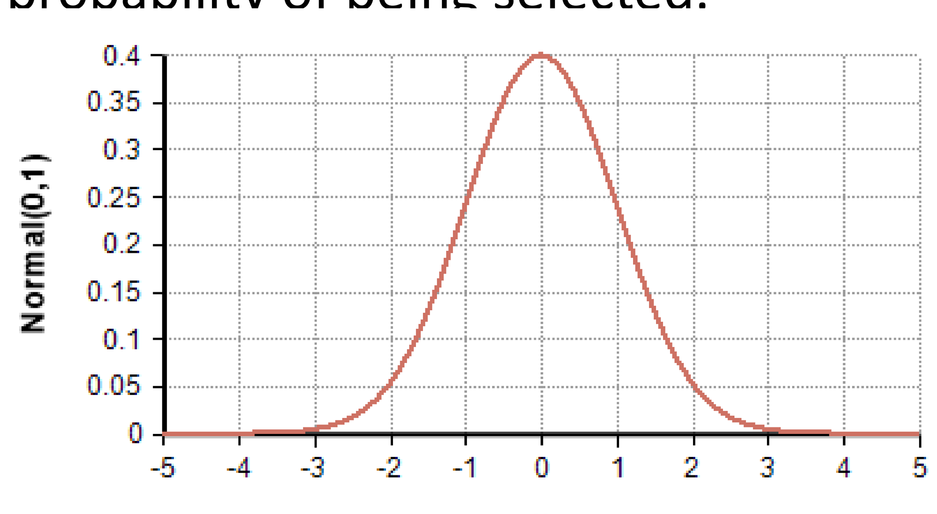
Learning curves are graphical representation of the relationship between the training set sizes and the error rate for a machine learning model. A well-behaved learning curve is a learning curve which reduces its error rate by increasing the size.

Curve fitting is the mathematical process of estimating a learning curve by minimizing the error between the predicted and actual values and arriving at optimal parameters.



2. Background

This study explores the effect of initializing the initial guesses used for curve fitting method **LM** by sampling them from a random distribution. It compares 2 distributions: uniform and normal. **Uniform** distribution samples values with equal probability from an interval described by 2 bounds. **Normal** distribution samples values with regards to the mean and a standard deviation, values closer to the mean having a higher probability of being selected.



Levenberg-Marquardt(LM) is one of the fastest method used for curve fitting. It uses **gradient descent** and **Gauss Newton** for convergence. It is commonly believed that this method is sensitive to bad initial guesses and getting stuck in local minima.

3. Research questions

How does the distribution sampling of the initial parameters of the model function affect the performance of curve fitting?

- We tackle this question by looking at well-behaved curves and using LM
- We compare the 2 methods of sampling and then try to improve them analytically

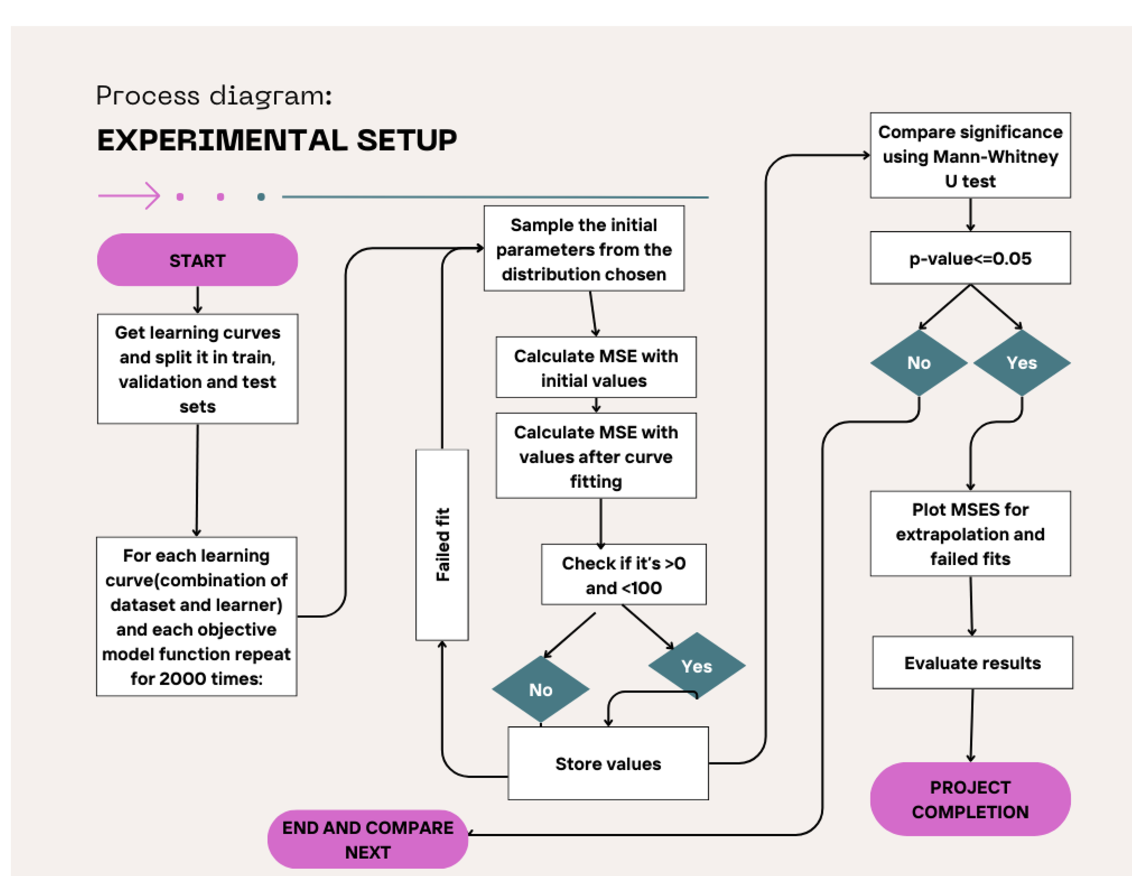
4. Methodology

We compare the distributions in terms of **average number of failed fits** and **average mean squared errors for interpolation and extrapolation for exp and pow functions.**

Initial assumptions:

- Standard normal distribution sampling might yield higher MSE due to negative sampling resulting in failed fits
- Sampling uniform data with negative lower bound similar results to normal distribution
- Increasing bounds and mean would reduce the number of failed fits
- Suboptimal fits resulting in stuck in local minima

5. Experimental setup



6. Results and discussion

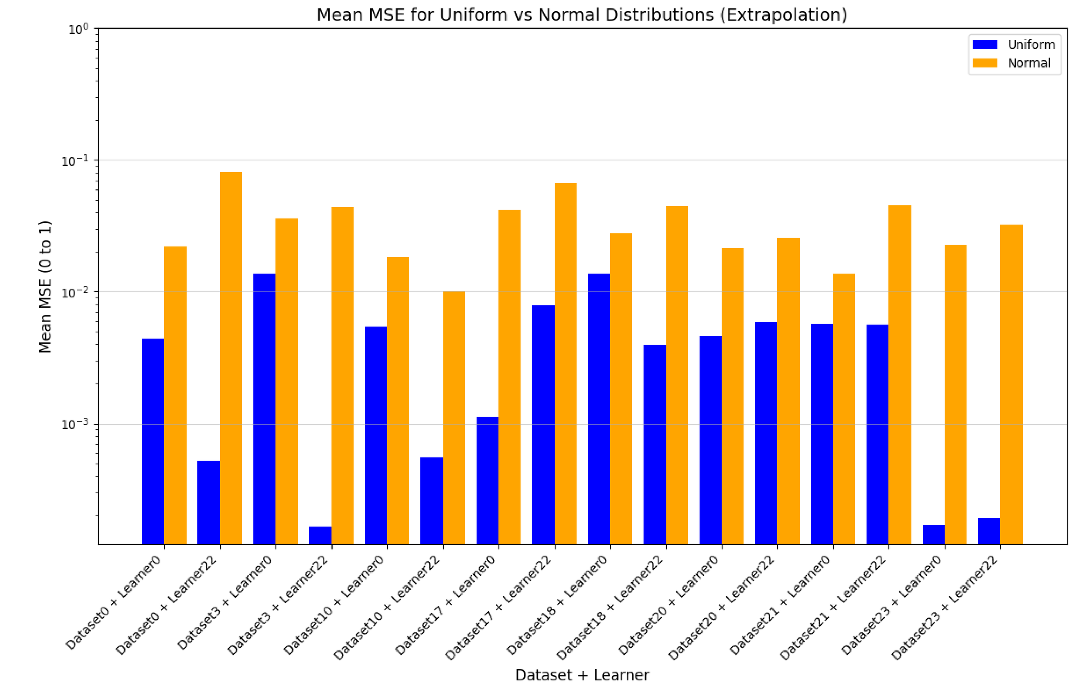


Figure 1 Standard Normal Extrapolation vs default uniform distribution

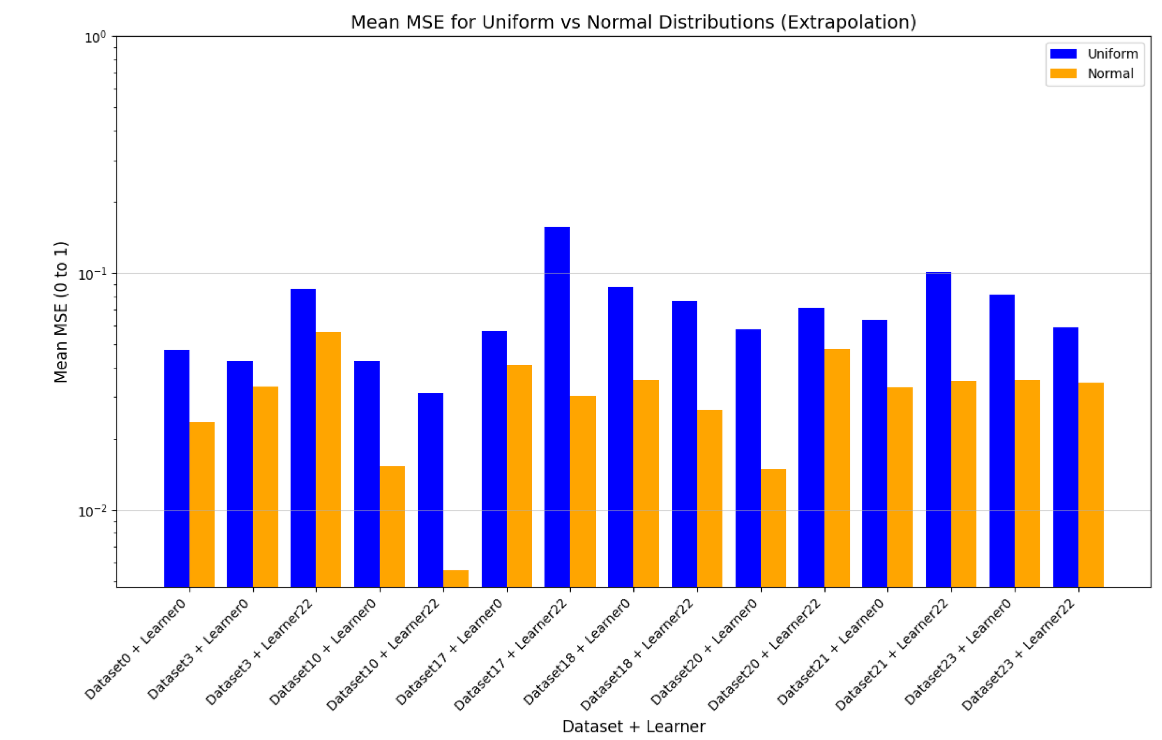
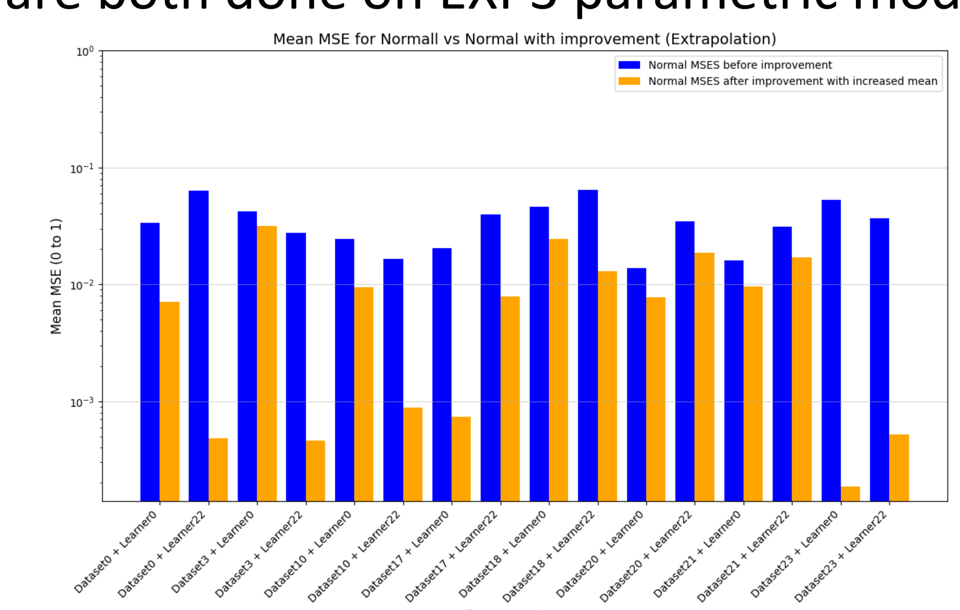


Figure 2 Standard Normal Extrapolation vs uniform distribution with [-3.3]

Despite the **standard normal distribution** sampling giving higher MSE values than the default uniform distribution, we can observe that by changing the lower bound of the uniform, it actually performs worse. Moreover, the number of failed fits is similar 1017 vs 1045. These experiments are both done on EXP3 parametric model.



Picture 3: Standard Normal Distribution vs normal distribution increased mean

Reducing the number of failed fits to 2 from 1000, increasing the bounds with the suggested values does also improve the error. This gives us an insight about potential good values for the parameters of the exponential 3 being around the mean of the distribution.

In figure 4, we can observe that despite many of the random sampling of parameters and the distribution of MSE values, LM tends to constantly converge to a local optima, arriving at the same value all the time.

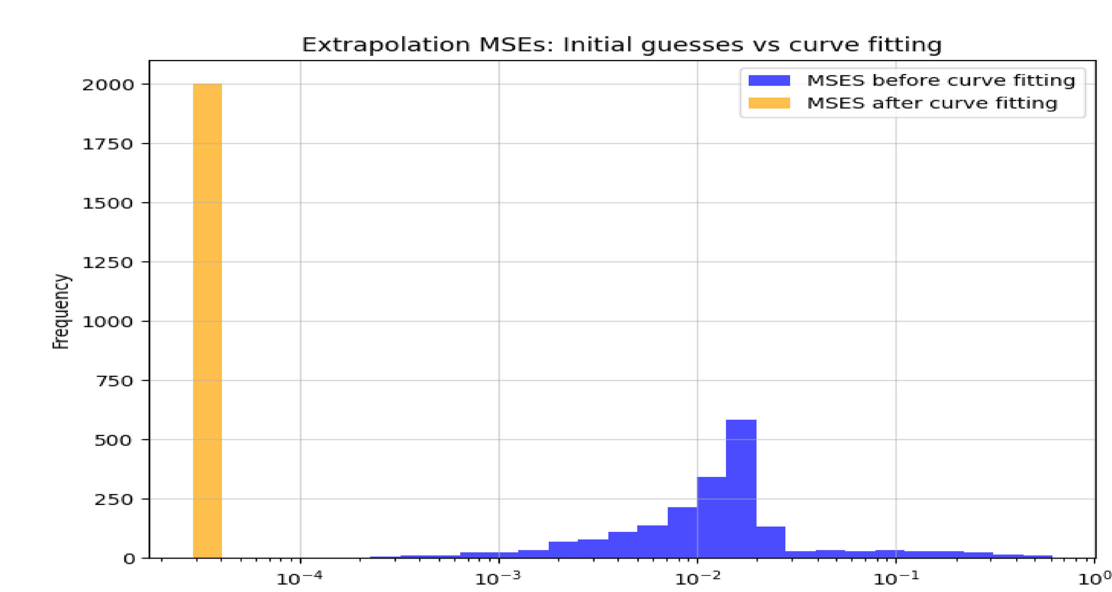


Figure 4: Initial MSES before curve fit vs after curve fit for exp3 dataset 0, linear

8. Conclusion

Uniform performs similar to normal with the right parameters, normal distribution work preciser for higher means, if the mean is closer to the optimal values.

Initial guesses are affected by **negative values**, causing more failed fits on average.

LM is prone to converging to a local minima, obtaining sub-optimal fits and not reaching global optima most of the times.

Future work could explore applications of our pipeline on ill-behaved curves and with more function models that do not fit so well.