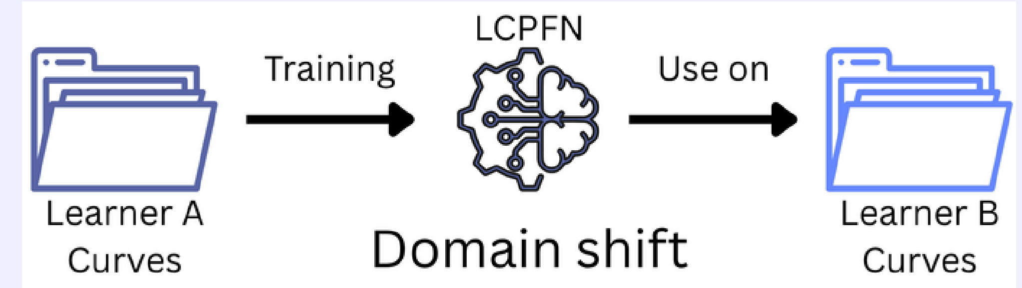
Max Soeters
M.Soeters-2@student.tudelft.nl
TU Delft

The Effect of Domain Shift on Learning Curve Extrapolation

Responsible Professor: Tom Viering
Supervisors: Cheng Yan, Sayak Mukherjee



Introduction

How much data do you need to train a model? This is a question that needs to be answered when you want to train a model to accomplish a machine learning task.

Sample size learning curves compare the performance of the model to the amount of data used for training. [1] This is important as data is expensive. Therefore getting the right amount of data is pivotal to make estimates of financing for projects, research or other projects.

Curve extrapolation is a useful tool for this, as you can find out how much more data you need to acquire by using the first few known points of the curve to predict the rest of the curve. The models used in this experiment are Learning Curve Prior Fitted Networks (LC-PFN) [2], which are trained on real curves and can then extrapolate parts of curves. Previous method show problems when estimating required sample size without taking a correct model into account. [3] Therefore it is a relevant question if taking the correct machine learning model into account is also important when extrapolating learning curves using LCPFNs.

Domain shift is a challenge in machine learning where the distribution of data is not the same between training and testing. This is a big problem for the generalizability of machine learning models. [4] In this project it would be when you train a model on one curves of one type of machine learning algorithm, learner A, and evaluate it on another curves from another, learner B. This is shown in the diagram in the top right. We will investigate how this shifting impacts the performance on curve extrapolation by the LCPFNs.

Research Questions

The research question in this project is What is the effect of domain transfer on learning curve extrapolation?

This can be broken down into the following sub questions:

1. Is there a trend between subtypes of learners in the effect of the domain transfer?
2. Does domain transfer over single learners impact the accuracy of PFNs?

Method

Experiment 1: Groups of learners

The first experiment will look at the domain shift trend among groups of learners. The groups were made by looking at the mechanisms for learning and the SciKitLearn documentation. This will give us an answer to the first research question. Discriminant Analysis and Trees stand out as groups with ill and well-behaved curves respectively.

Experiment 2: Discriminant Analysis Learners

I will evaluate the Linear Discriminant Analysis and Quadratic Discriminant Analysis learners and compare them to the trees group, to see if there is an effect of domain shift when using single learners. The previous experiment clearly indicated that the DA group was of interest to investigate.

Experiment 3: Well-Behaved Curves vs Ill-Behaved Curves.

Experiment 3 will explore if the well-behavedness of curves is a part of the effect of domain shift on extrapolation. It will do this by comparing some well-behaved learners (Gradient and ExtraTree) and some ill-behaved learners (LDA and Sigmoid). Finding out if this is actually the case will help answer research question 1 and 2 as it looks for a trend while also looking at domain shift along
individual learners. The other experiments hinted this could be the pattern, but this will make it clearer if it is a true part of the pattern.

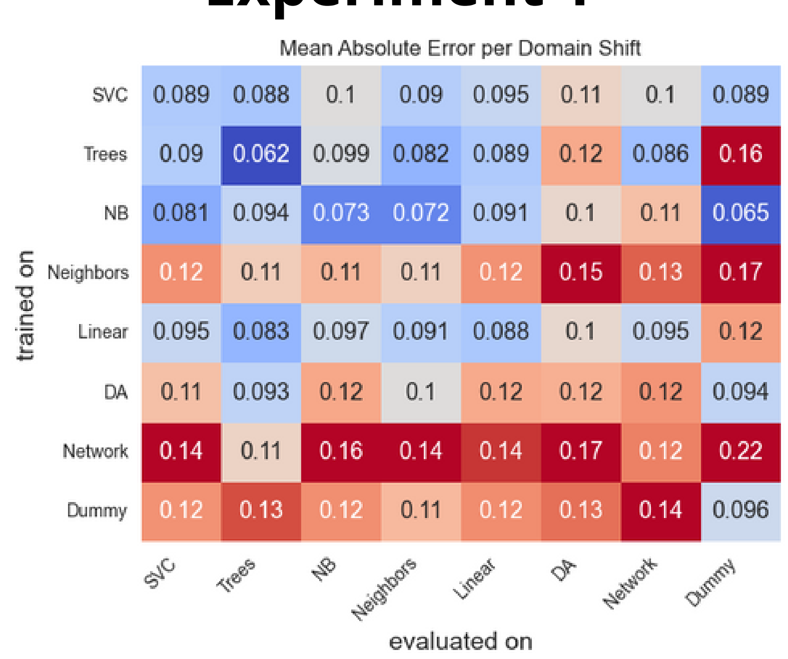
Conclusion

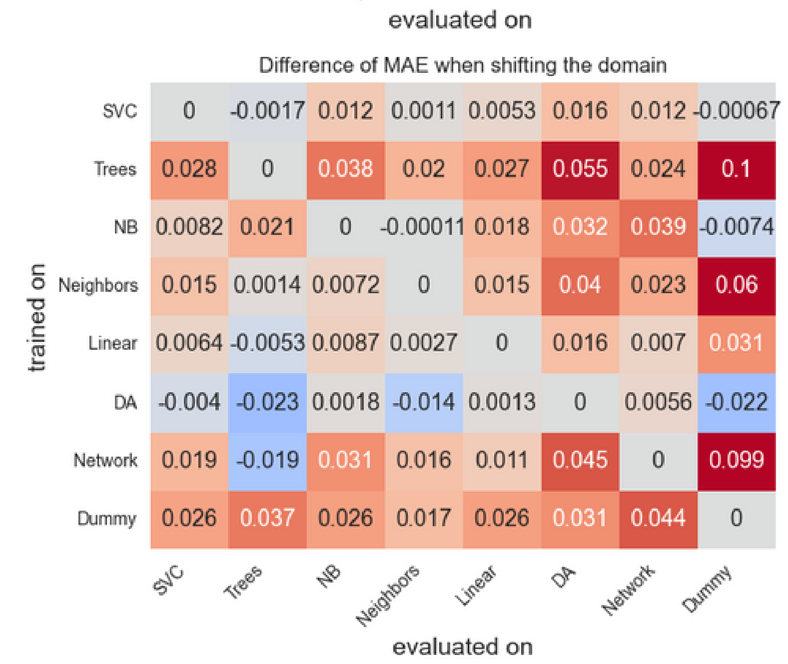
- Research Question 1: There is a trend in the effect of domain shift between groups of learners present. Experiment 1 concluded that it seems to be partly dependant of how well-behaved the group or learners are, because some domain shifts resulted in models improving performance. Experiment 2 and 3 also indicated a similar pattern. However, we have definitely not uncovered the full pattern behind the effect of domain shift.

- Research Question 2: The domain transfer definitely impacts the accuracy of the LCPFNs between single learners. With 88% of the effects of domain transfers being significant with single learners, we can see that most domain shifts have a significant effect, but not all.

Results & Discussion

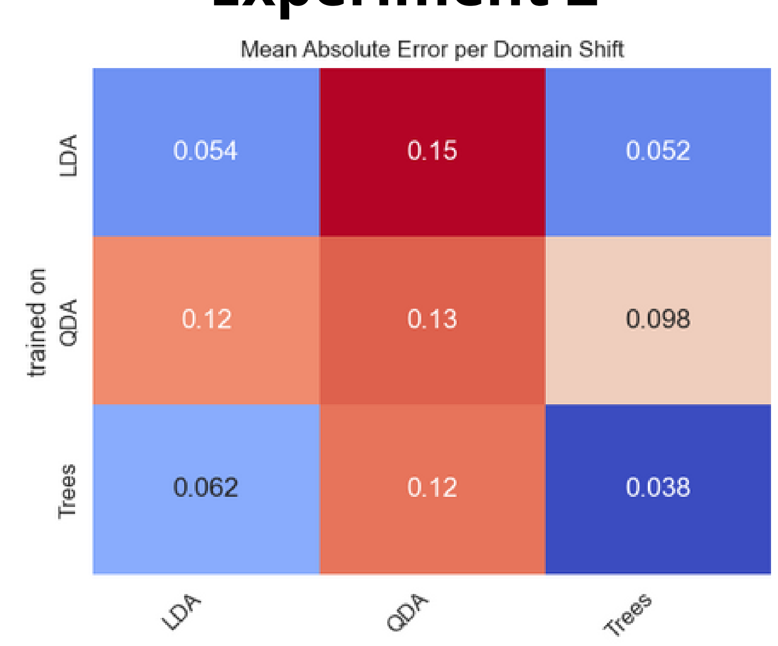
Experiment 1

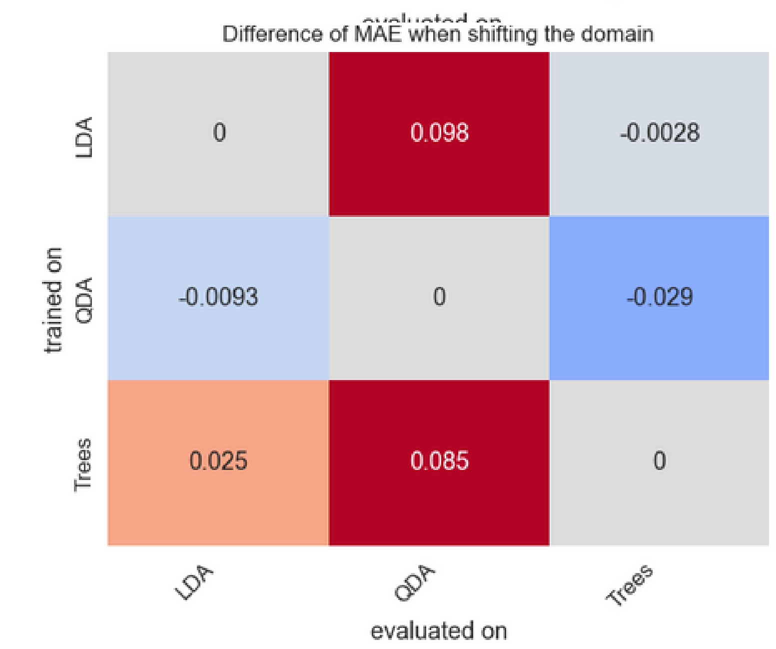




- Discriminant analysis group is very poorly evaluated by all models.
- The model trained on discriminant analysis group does better on other groups
- The trees group is evaluated better on average by other models then those models evaluating their own groups. The trees model performs the best on it's own group out of all models.
- Trees to discriminant analysis is the biggest effect of domain shift

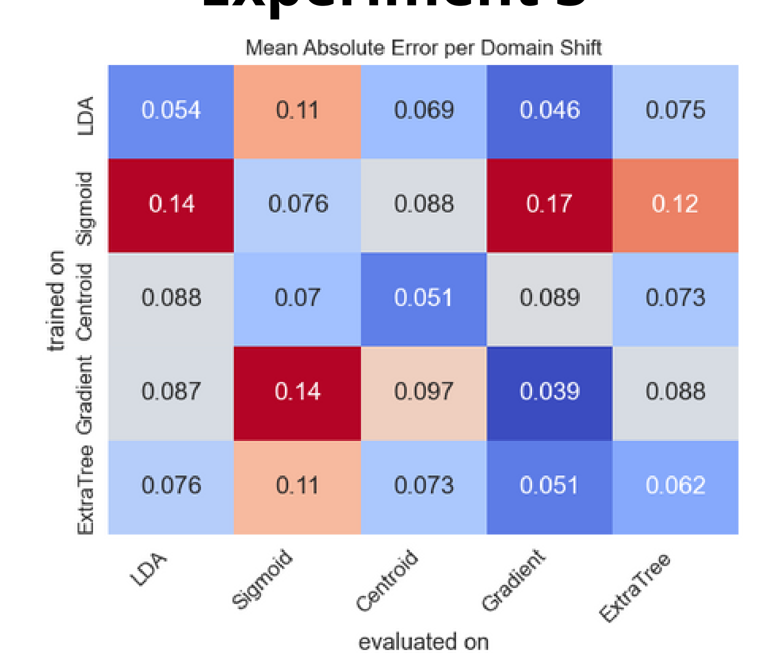
Experiment 2

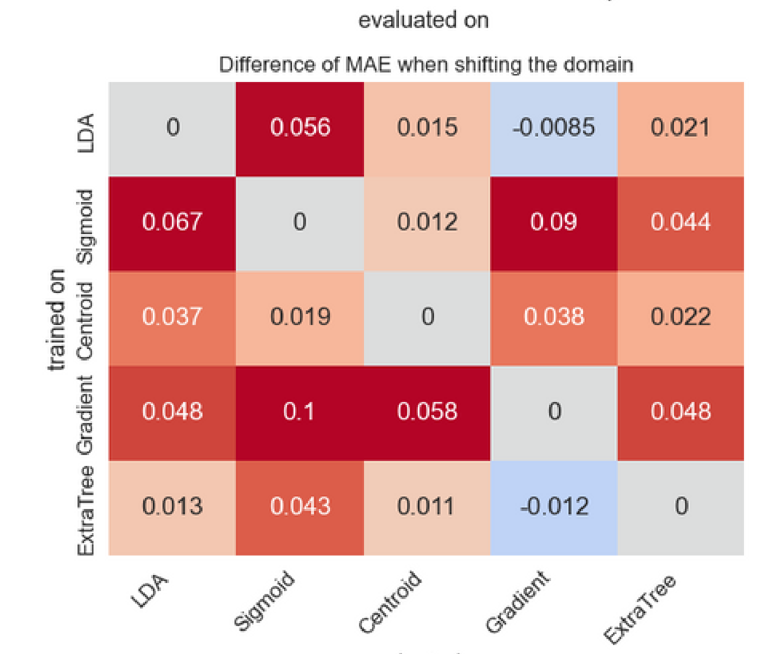




- Used improved training method using augmentation. Verified to see similar patterns.
- QDA is the worst performing model by far.
- QDA model improves after both domain shifts.
- Both domain shifts towards Trees were improvements over the baselines. However, the LDA to trees effect is not significant.

Experiment 3





- Sigmoid classifier was evaluated the worst overall. Both with its own model and other models, it had the biggest performance drop on average.
- LDA to Sigmoid is a big domain shift, while both are statistically ill-behaved curves.
- The two domain shifts that are not seen as significant are domain shifts to the centroid learners, which was neither very well or ill behaved.

Discussion

- Experiment 1 heavily suggests that there is a correlation between the well-behavedness of a group and the effect of domain shift, with the DA model performing very bad on its own group, while actually improving on other groups. This all while the Trees group gets evaluated better by other models than their own group on average. This is very peculiar behaviour that suggests this pattern.
- Experiment 2 reinforces this point by showing that the QDA learner shows the same behaviour on its own, while the LDA learner seems more mild.
- The fact that the domain shift from LDA to sigmoid has a big effect is interesting as the difference in distribution between the groups is mostly in the dipping percentage and flatness, with the percentage of peaking being similar. This means that the dipping and flatness is certainly a factor of the effect of domain transfer.
- The fact that the domain shift towards the centroid learner is not significant twice makes sense as it was included as the middle of the line learner with percentages of peaking and dipping that are not too good and not too bad, making it closer to both well and ill behaved groups.

References:

[1] Viering, T. J., Adriaensen, S., Rakotoarison, H., & Hutter, F. (2024). From Epoch to Sample Size: Developing New Data-driven Priors for Learning Curve Prior-Fitted Networks.
[2] Adriaensen, S., Rakotoarison, H., Müller, S., & Hutter, F. (2023). Efficient Bayesian Learning Curve Extrapolation using Prior-Data Fitted Networks (No. arXiv:2310.20447). arXiv. https://doi.org/10.48550/arXiv.2310.20447
[3] Viering, T., & Loog, M. (2022). The Shape of Learning Curves: A Review (No. arXiv:2103.10948). arXiv. https://doi.org/10.48550/arXiv.2103.10948
[4] Wang, Y., Zhang, Z., Gong, D., & Xue, G. (2025). Mitigating domain shift problems in data-driven risk assessment models. Reliability Engineering & System Safety, 263, 111263. https://doi.org/10.1016/j.ress.2025.111263