1. What are Learning Curves?

Plots generalization error \( C \) versus training set size \( S \). Use them to:

1. Estimate the value of collecting more training. By extrapolating a learning curve.
2. Speed up training. If the curve doesn’t improve with \( S \) anymore, stop adding more data to save time.
3. Faster model selection. By extrapolating the curve of learners, we can rule out bad learners early [1].

2. Why study Learning Curves?

Learning curves can have surprising shapes in artificial settings, such as curves with local minima/maxima. How widespread is this in practice?

No consensus on the shape of learning curves [2]. Best modelled by a power law, exponential, …?

3. Database Highlights

We publish a large database of learning curves:

- 20 learners on 246 datasets
- Getting our data: pip install lcdb
- Precomputed error rate, F1, AUC ROC, log loss
- Provide all predictions (can compute any metric)
- Bootstrapping: 25 train / validation / test splits
- Training set sizes \( S_i = \left\lfloor 2^{(i+1)/2} \right\rfloor \), \( S = \{16, 23, 32, \ldots\} \)

4. Preliminary Findings

A. Are error rate curves monotone?

Define \( \epsilon_{\text{mon}} = \max\{0, \hat{C}(S_{i+1}) - \hat{C}(S_i)\} \)

- Empirical Cumulative Density of Compliance with Monotonicity

Averaged over the whole database most curves seem monotone according to \( \epsilon_{\text{mon}} \)

B. Local maxima in error rate?

- Only some learners peak (local maximum)
- Peaking lessens when \( S \) is larger

C. Do curves cross?

- Yes curves cross 10-20% on average

5. Discussion

Error rate: seemingly monotone, without too many local maxima, but do cross.

Extrapolation: mmf4, wbl4 perform best. Often excluded in prior studies. Other results agree with Brumen [3].

2% discarded fits: need for robust fitting!

6. Version 2.0?

- Support for pipelines (to implement feature scaling)
- Hyperparameter tuned models (we use defaults)
- Use monotonic training sets so that \( S_1 \subset S_2 \subset S_3 \ldots \)
  (currently all training sets are sampled independently)