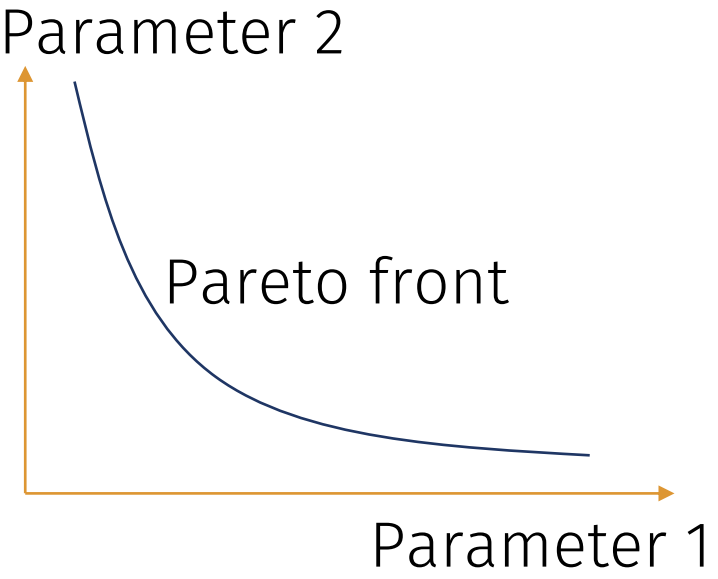
Assessing Models

Loss function

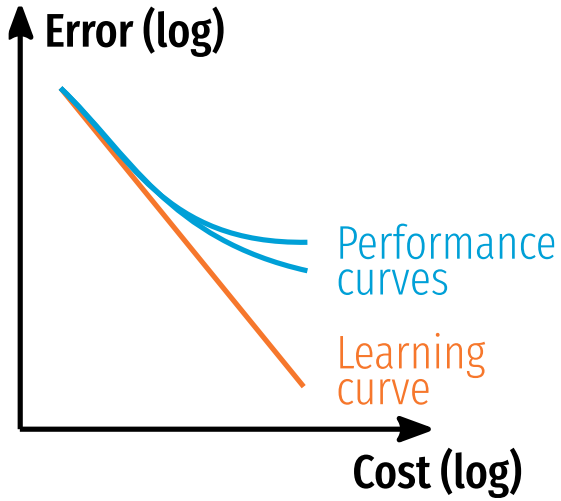
- Need a single variable to optimize
- Generally: $f(\theta) : \mathbb{R}^n \rightarrow \mathbb{R}$ (θ = parameter vector)

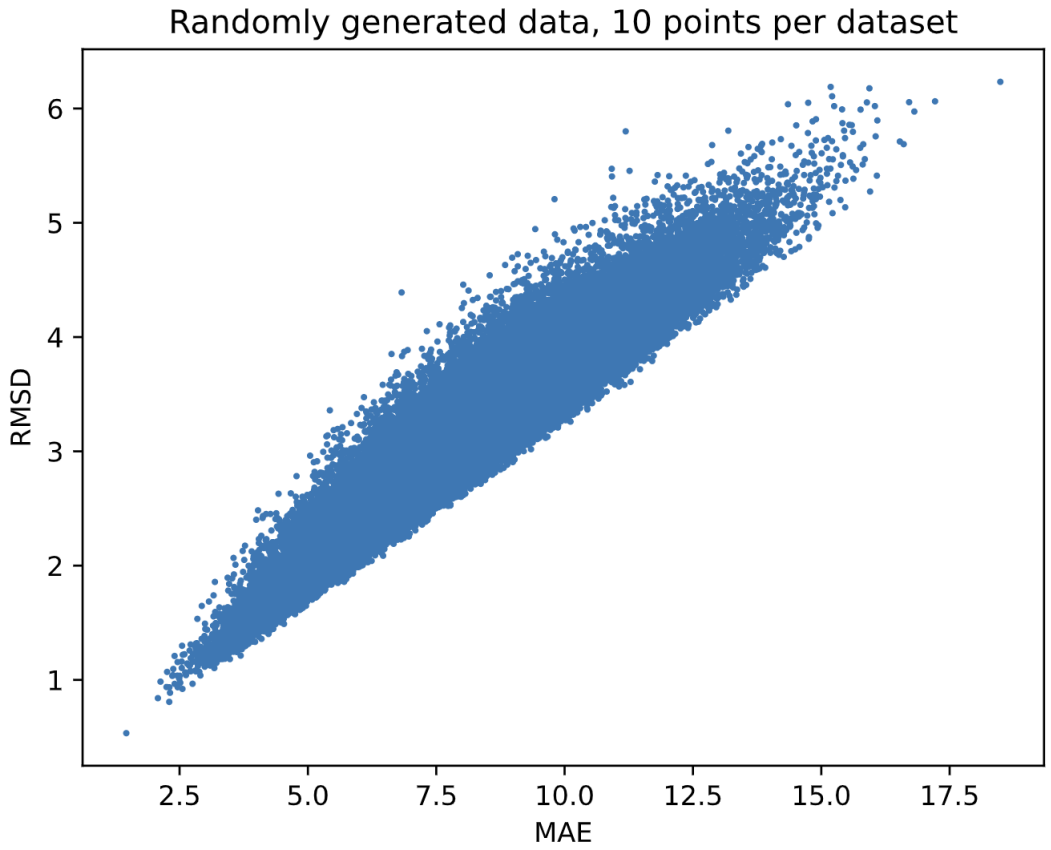
- Other names
  - Cost function
  - Error function
  - Fitness function
  - …
- Convention: minimize to improve
- Ideally: differentiable w.r.t. θ
- Defined by problem in most cases



Loss Function vs Scoring Metrics

Loss functions: training aims to improve them
Scores: improve accidently





Common loss functions

Regression

- Mean absolute error	MAE	
- Root mean squared error	RMSE	Least squares, used by KRR

Classification

- Information gain	IG	see Decision Trees
- Log loss	LL	also: Neural Networks

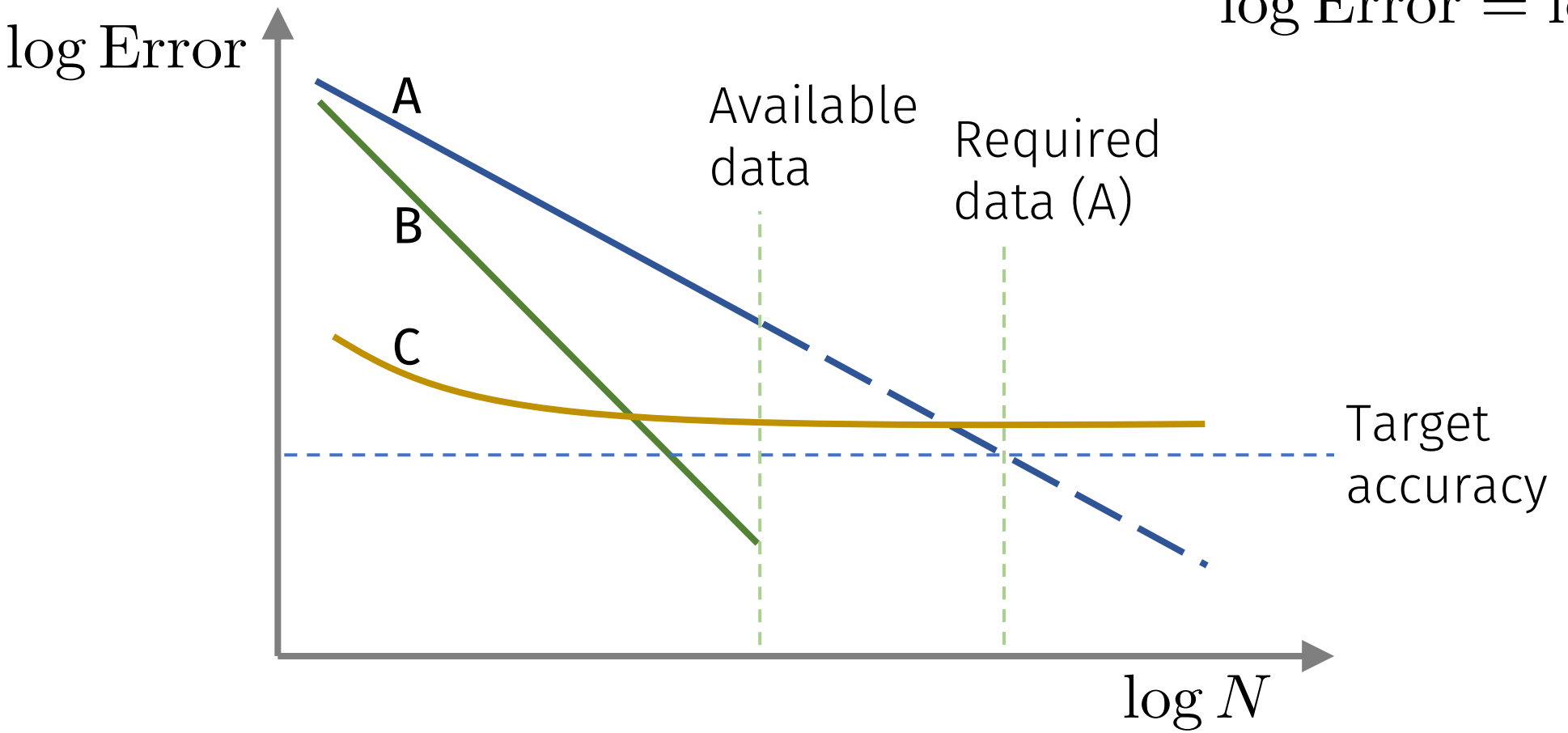
$$LL \equiv -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{i,k} \log p_{i,k}$$

N data points, K classes

Learning Curves

$$\log \text{Error} \propto \frac{a}{N^b}$$

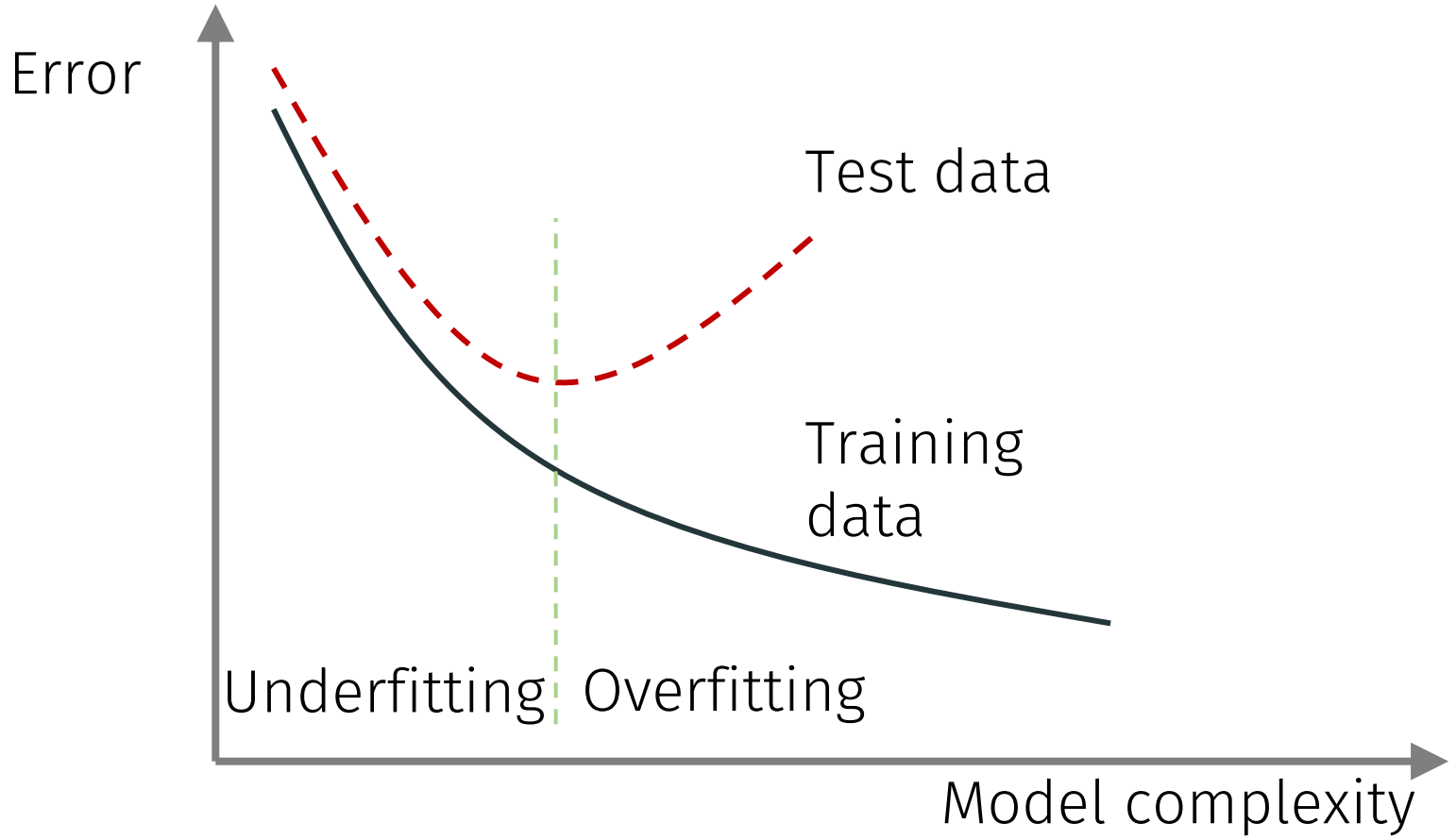
$$\log \text{Error} = \log a - b \log N$$



Assessing Model Quality

Classical fitting

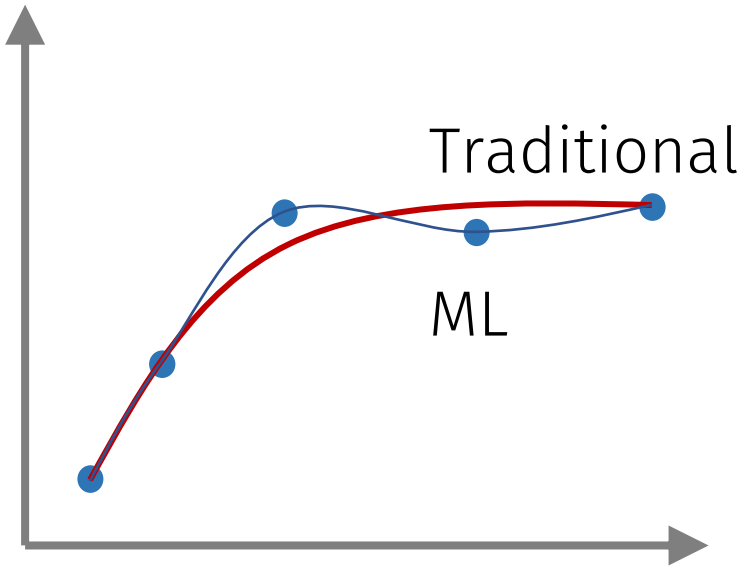
- As few parameters as possible
- Conflict between accuracy and transferability

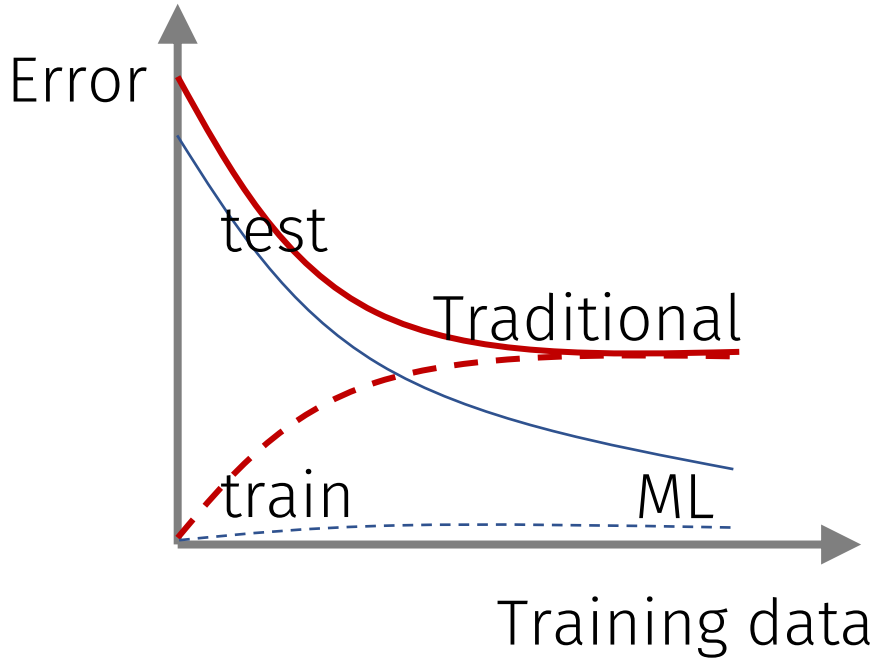


Assessing Model Quality

Comparison fitting approaches

- Number of parameters
- Relative error asymptotic values





Assessing Model Quality

Possible reasons for inefficient learning in models

- No relationship in the aspects covered in the representation
- Ambiguous representation
- Symmetries not covered
- Little transferability in the representation
- To few data points in relation to complexity
- Underfitting / Overfitting

Common scoring functions

Classification

- Accuracy / Zero-One-Loss: which share is correctly labelled?
- Top-k-accuracy: how often is the correct label amongst the top k labels?
- F1 score: harmonic mean of precision and recall

Regression

- Mean absolute error
- Root mean squared error
- Max error

Precision: True positives / (true+false positives)
Recall: True positives / (true positives + false negatives)

Summary Assessing models

- Loss function: single target that is optimized
- Score function: testing how well model does, not necessarily the loss function
- Iff loss function = score function: should stricly improve with training
- Some loss functions are fixed: KRR requires RMSE
- In general: free to choose, some are more useful than others (simplicity of solution, gradients)
- Rule of thumb: choose loss function such that model is cheap to train, choose score function such that model is useful in practise