Generalization

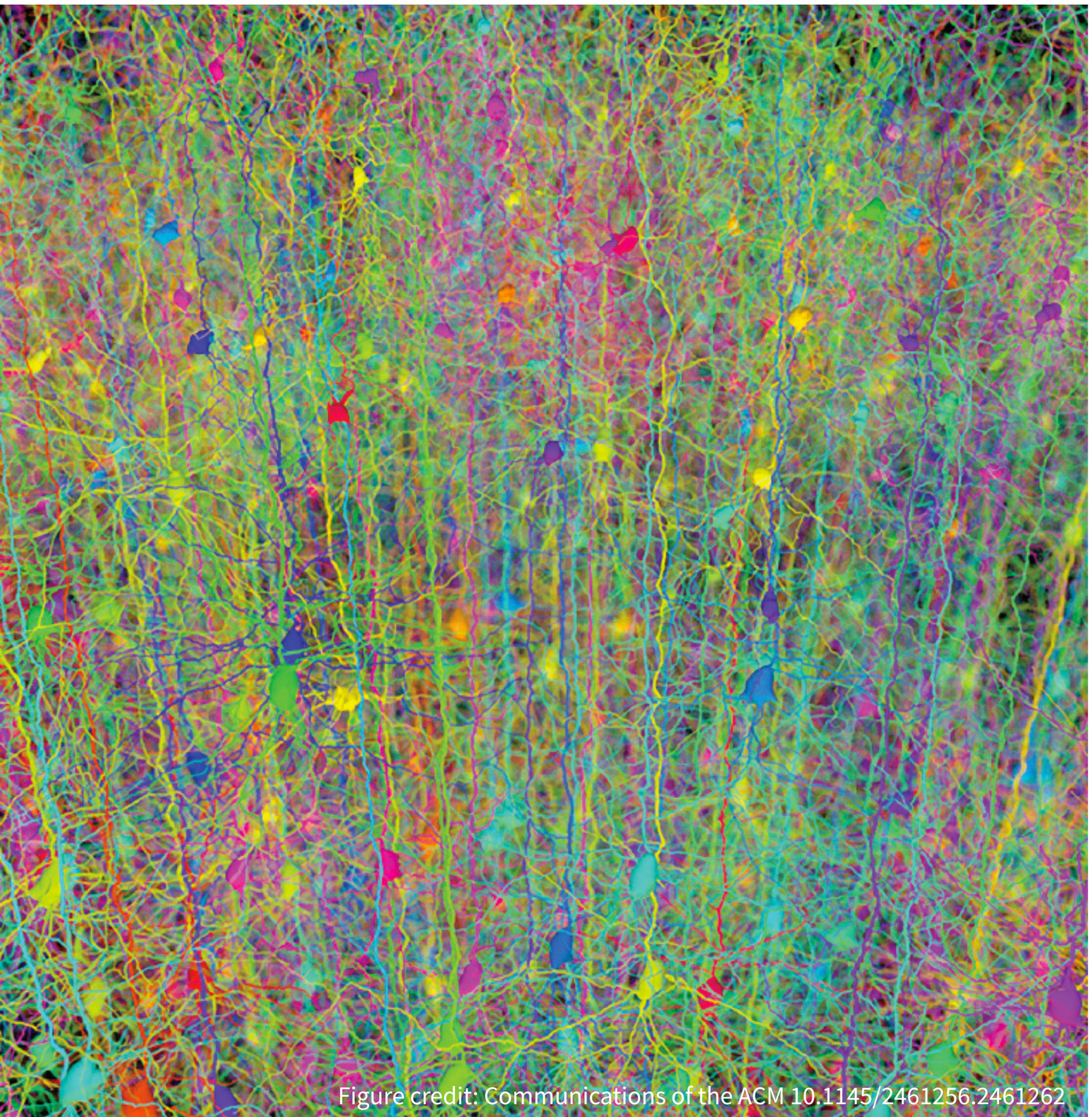


Figure credit: Communications of the ACM 10.1145/2461256.2461262

GRAHAM TAYLOR

VECTOR INSTITUTE

SCHOOL OF ENGINEERING
UNIVERSITY OF GUELPH

CANADIAN INSTITUTE
FOR ADVANCED RESEARCH

UNIVERSITY of GUELPH
CHANGING LIVES
IMPROVING LIFE

CIFAR
CANADIAN
INSTITUTE
FOR
ADVANCED
RESEARCH

Generalization

The central challenge in machine learning is that the algorithm must perform on *previously unseen inputs*

The ability to perform well on previously unseen inputs is called generalization

via Gazette Review

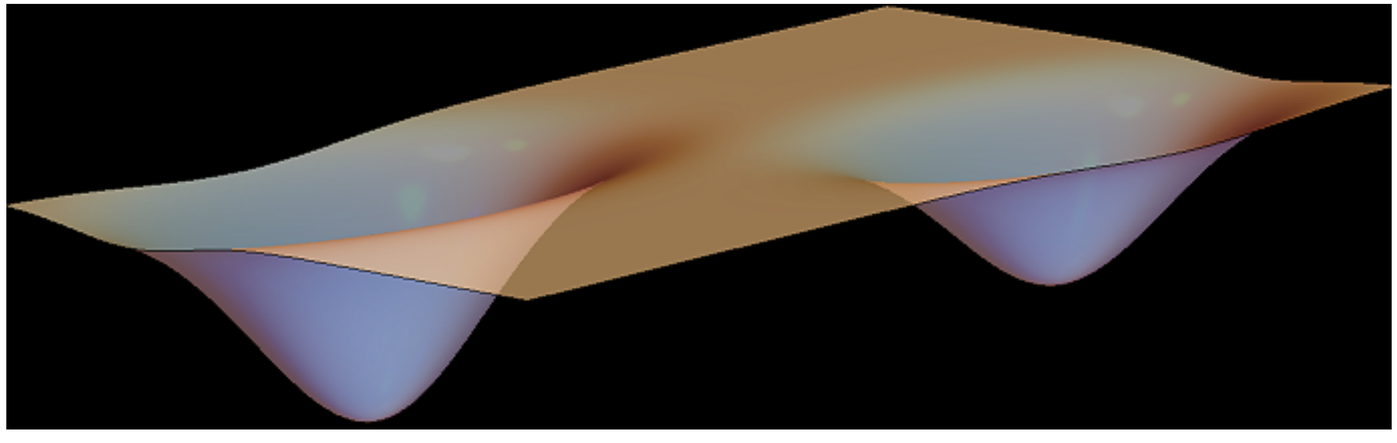
Generalization

The central challenge in machine learning is that the algorithm must perform on *previously unseen inputs*

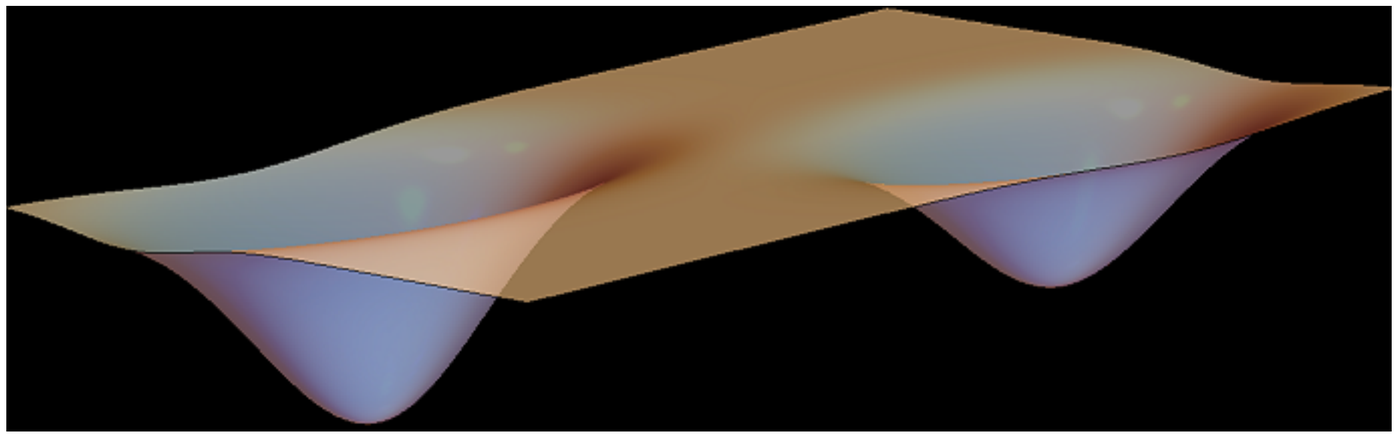
The ability to perform well on previously unseen inputs is called generalization

via Gazette Review

Machine Learning vs. Optimization

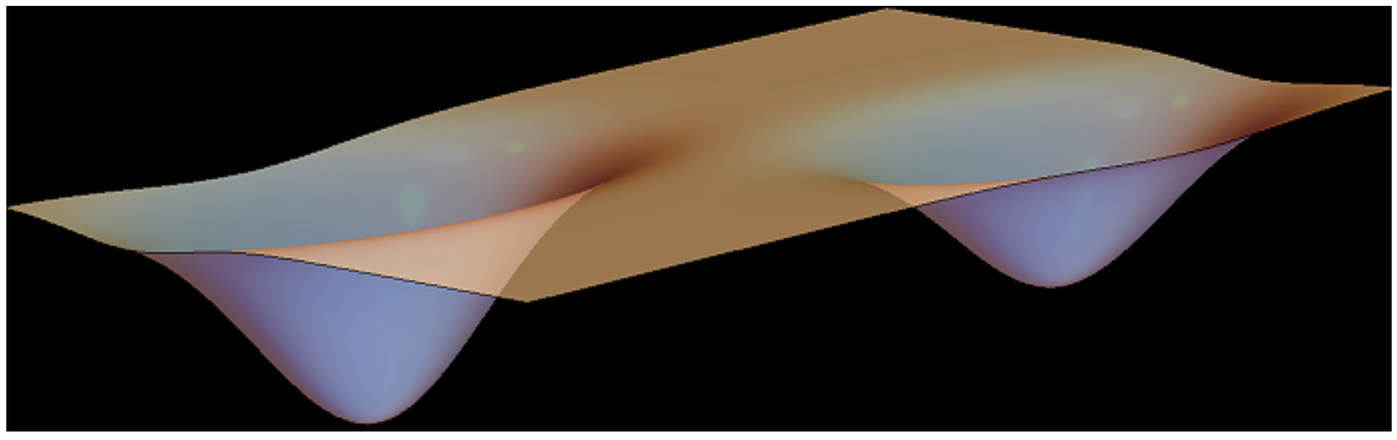


Machine Learning vs. Optimization



- Minimizing error on the training set is simply optimization

Machine Learning vs. Optimization



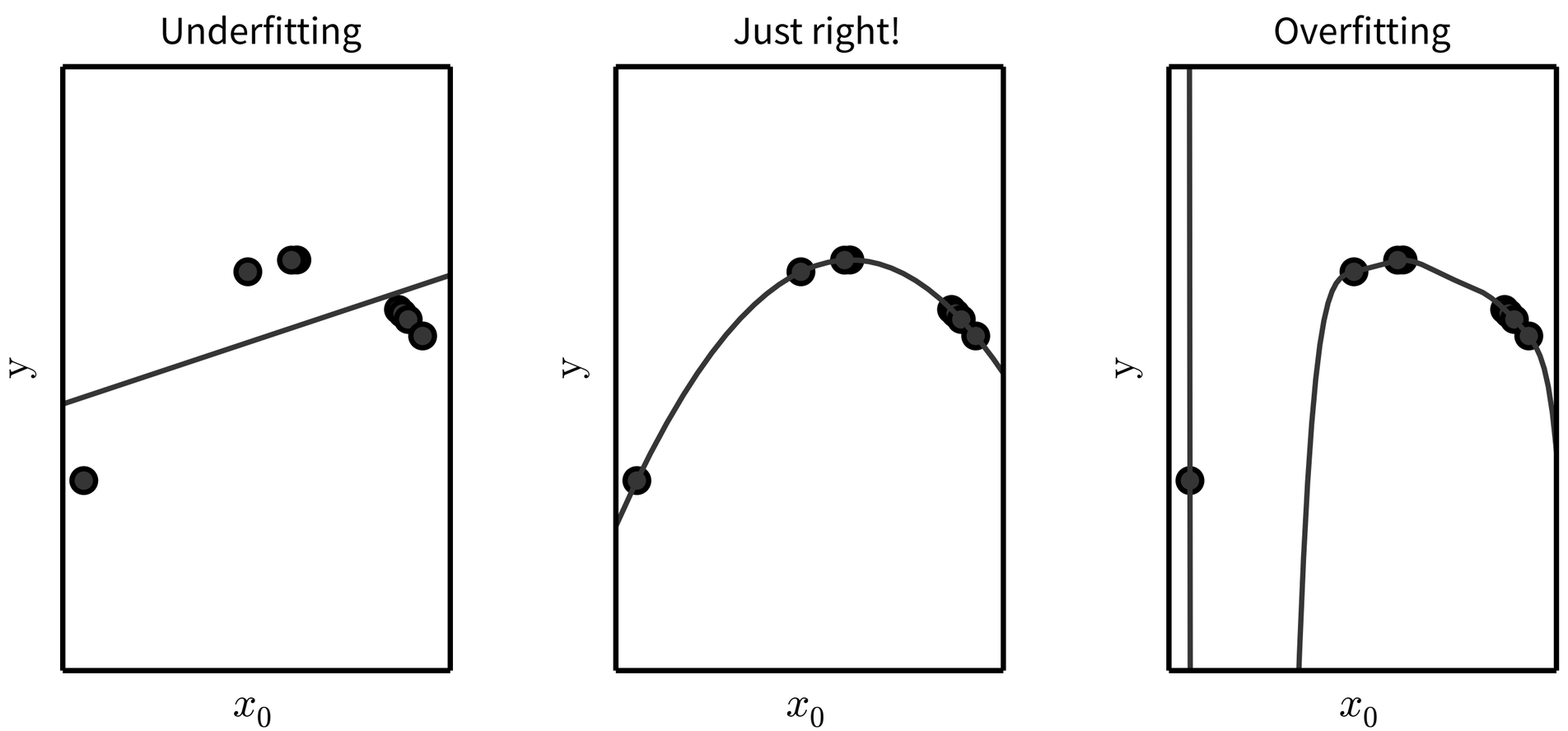
- Minimizing error on the training set is simply optimization

- What separates ML from optimization is that we want the generalization error (also called the test error) to be low as well

Underfitting vs. Overfitting

The factors determining how well a ML algorithm will perform are:

1. Make the training error small
2. Make the gap between training and test error small

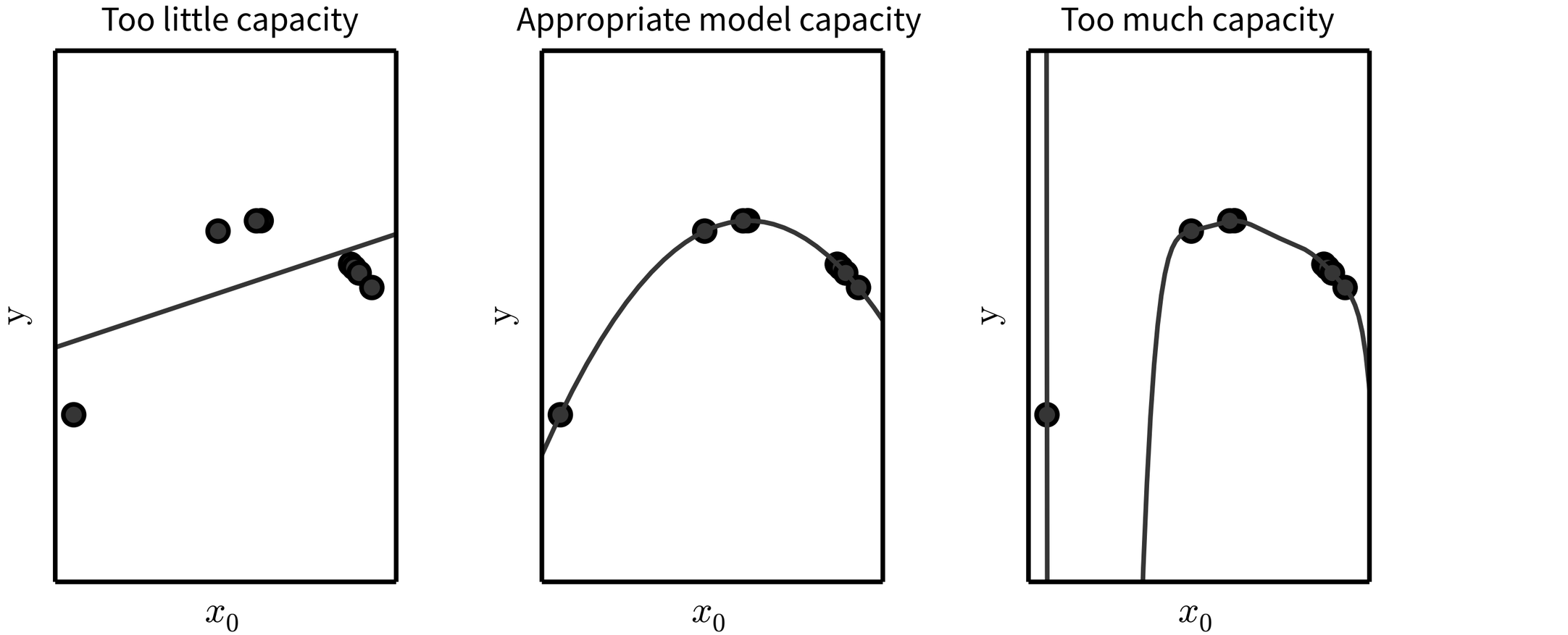


via deeplearningbook.org

Capacity

We can control whether a model is more likely to overfit or underfit by altering its capacity

One way to control capacity is to choose hypothesis space

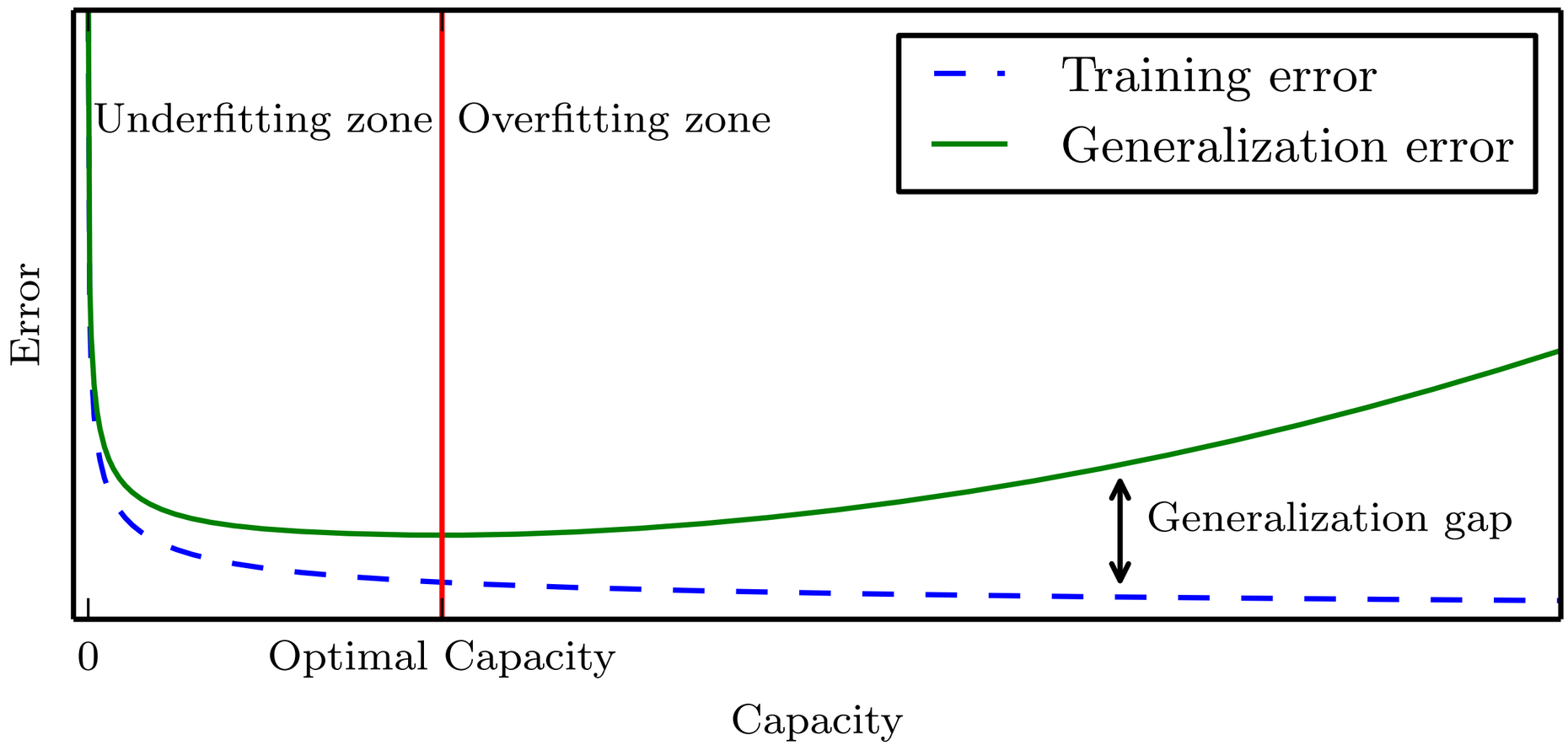


via deeplearningbook.org

Error vs. Capacity

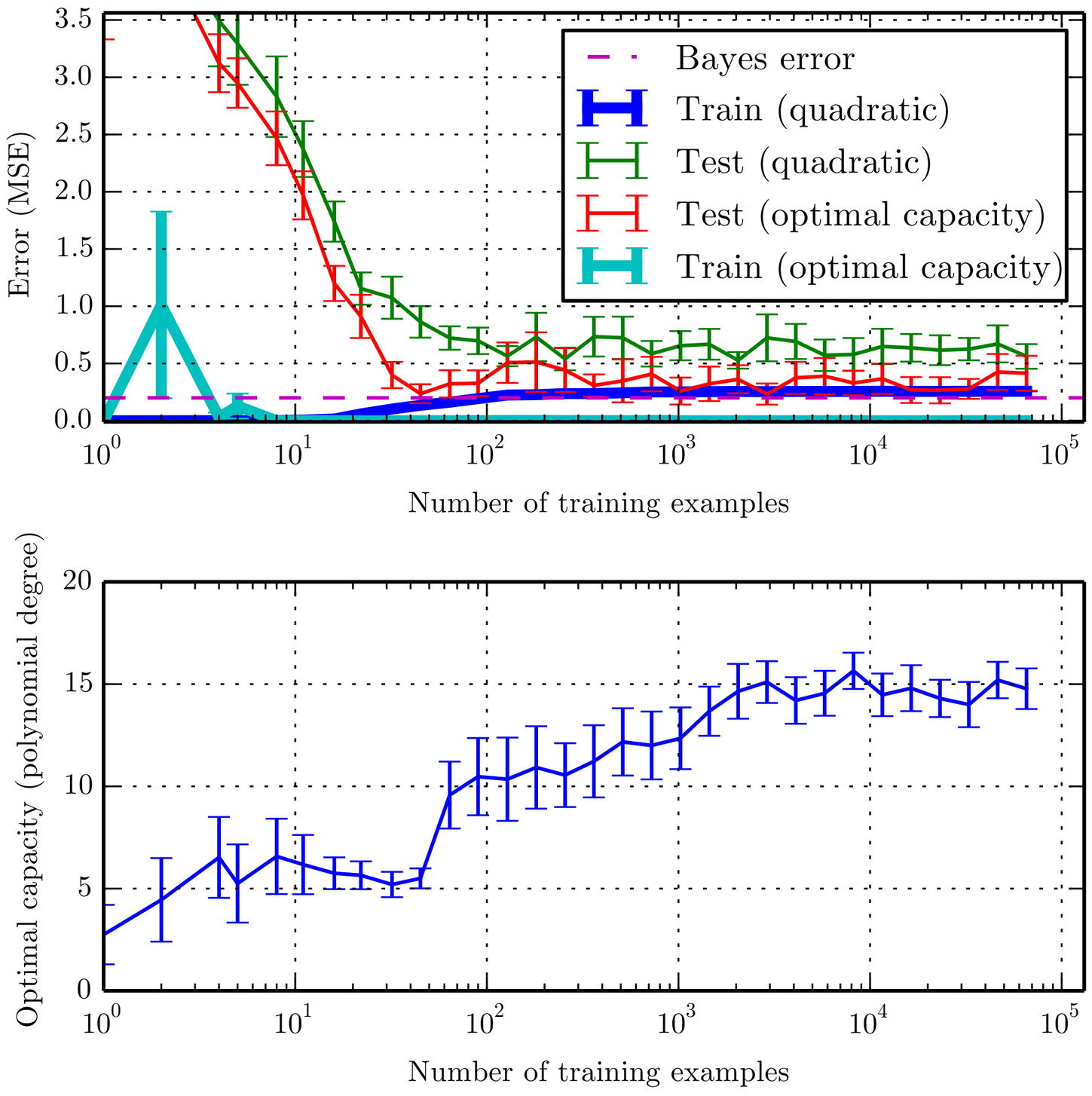
Typically:

- Training error decreases as model capacity increases
- Generalization error has a U-shaped curve as a function of capacity



via deeplearningbook.org

Error vs. Training Set Size



via deeplearningbook.org

No Free Lunch Theorem

Wolpert (1996)

Averaged over all possible data-generating distributions, every classification algorithm has the same error rate when classifying previously unobserved points.

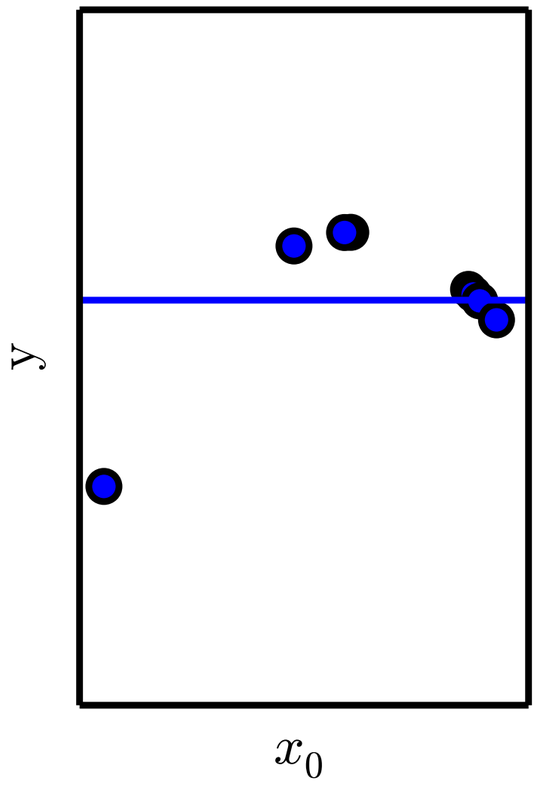
- No ML algorithm is universally better than any other
- However, if we make assumptions about the kinds of probability distributions we encounter in real-world applications, we can design learning algorithms that perform well on these distributions

Preferences in Hypothesis Space

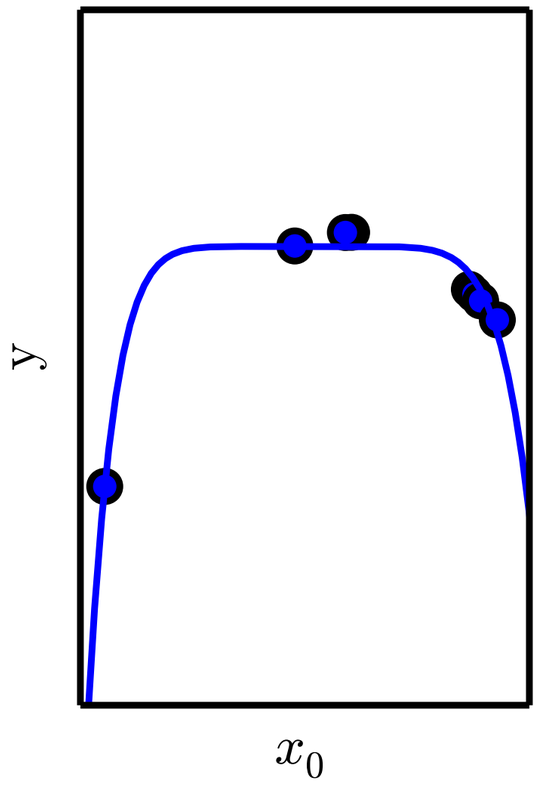
One way to control capacity is by adding or removing functions from the hypothesis space of solutions (e.g. degree of polynomial)

We can also give the learning algorithm a preference for one solution over another in its hypothesis space

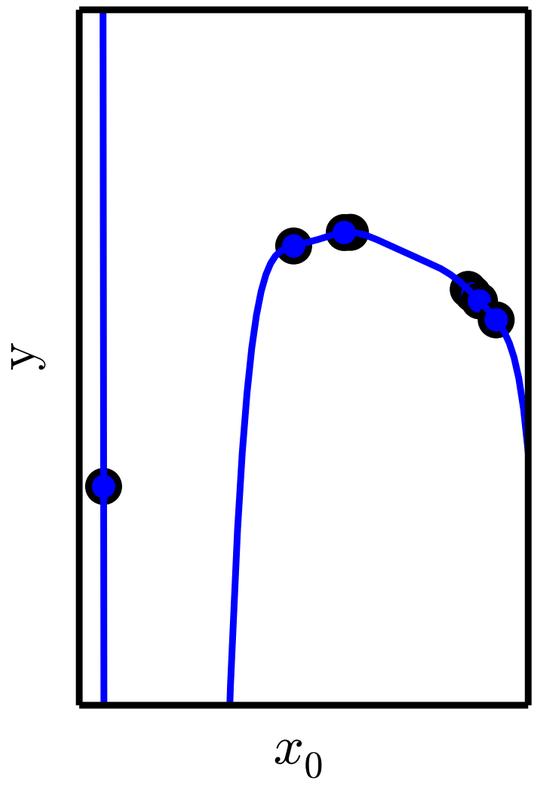
Underfitting (Heavy regularization)



Appropriate regularization



Overfitting (Weak regularization)

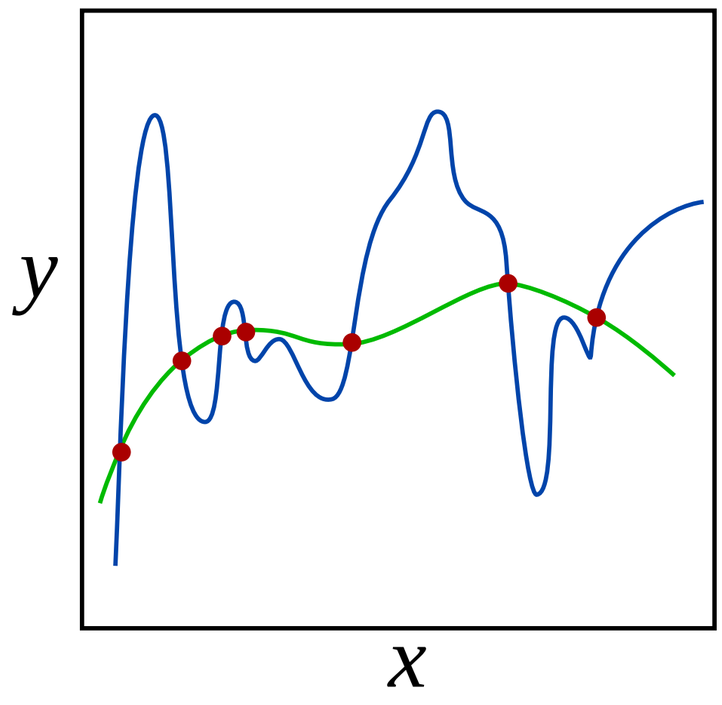


via deeplearningbook.org

Regularization

Regularization is any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error

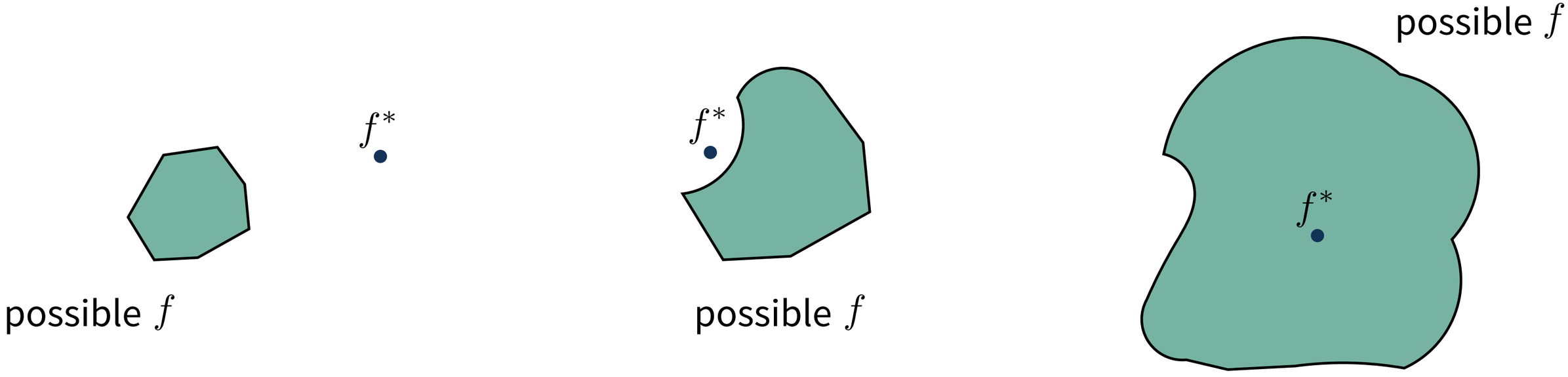
Advice: always regularize your model



via commons.wikimedia.org

Bias-Variance Trade-off

- **Variance:** does the model vary a lot if we change the training set?
- **Bias**: is the average model close to the true solution?
- **Generalization**: can be seen as the sum of the (squared) bias and the variance



reproduced from Larochelle