Machine Learning I
MATH60629

Machine Learning fundamentals — **Summary**
— Week #2

Three main components

1. Task (T). The question you are answering.

   - Model. How to parametrize?

2. Performance measure (P). How good is the model?

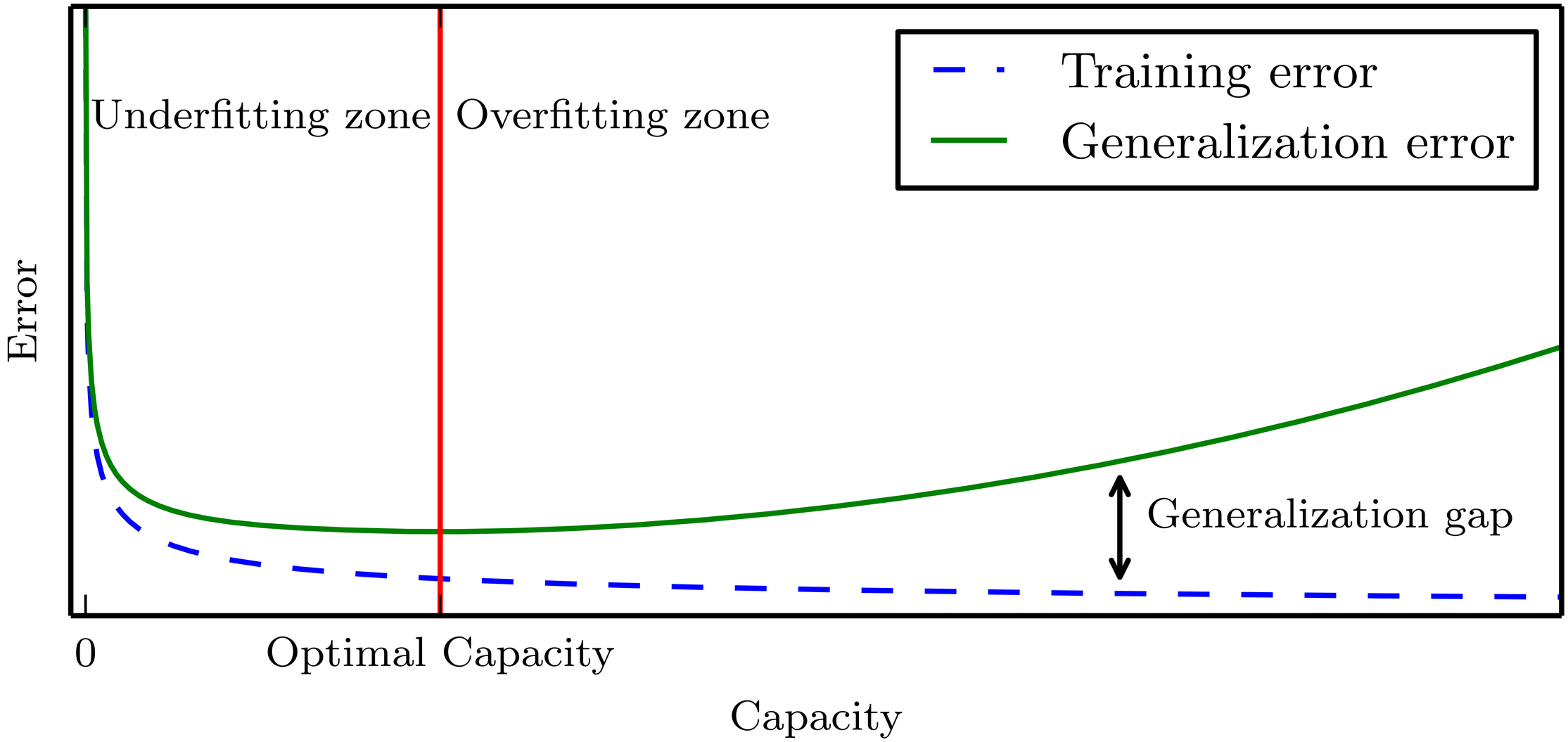
3. Experience (E). What type of data do you have access to?

Les types d’expériences

- Supervised {(x,y)}. e.g., regression, classification. f: X -> Y •
- Unsupervised {(x)}. e.g., clustering, dim. reduction, density estimation
- Reinforcement learning. Agent takes actions in an environment.

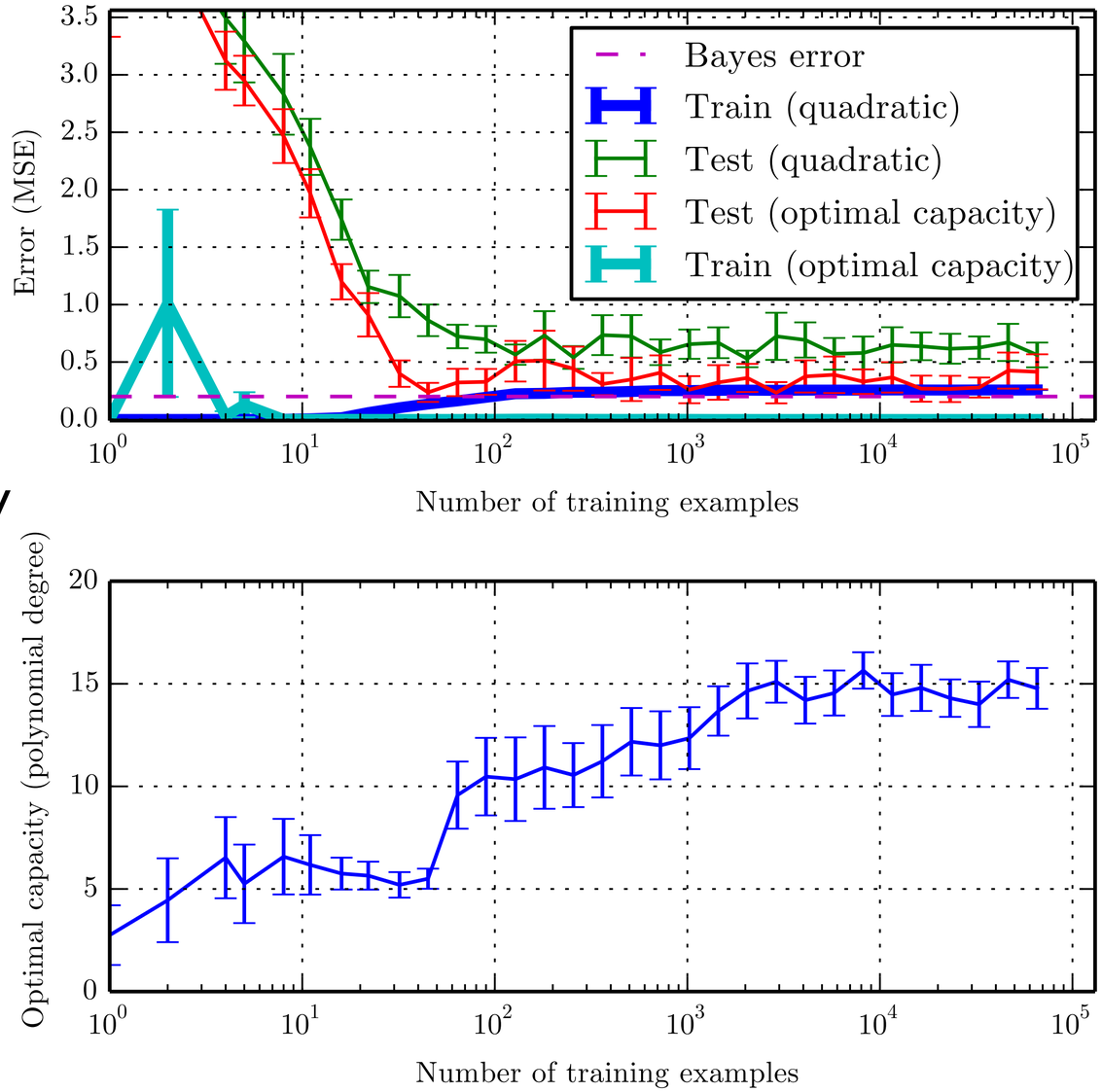
Model Evaluation

- Given:
  - A performance measure
  - A train dataset
  - A model
- Can calculate:
  - Train error: used to learn (to train).
  - Train error cannot be used to evaluate your model
  - Must use a separate dataset for evaluation


Underfitting zone
Overfitting zone
Training error
Generalization error
Error
Generalization gap
0
Optimal Capacity
Capacity

Synthetic data is generated using a degree 5 polynomial $y = w_5x^5 + w_4x^4 + w_3x^3 + w_2x^2 + w_1x^1$

Training set size also plays an important role in a model's capacity to generalize



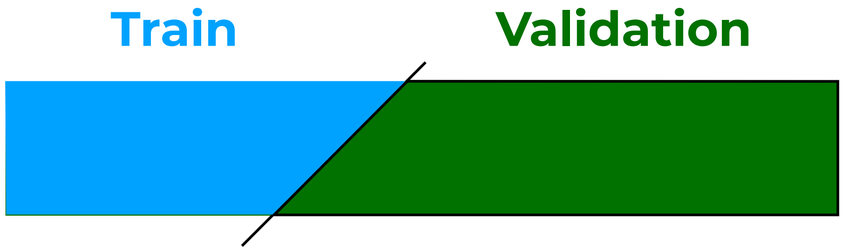
[Figure 5.4, Chapter 5, Deep Learning]

Regularization

- Can be thought of as way to limit a model's capacity
- Loss $:= \text{MSE}^{\text{train}} + \lambda \mathbf{w}^\top \mathbf{w}$

Validation set

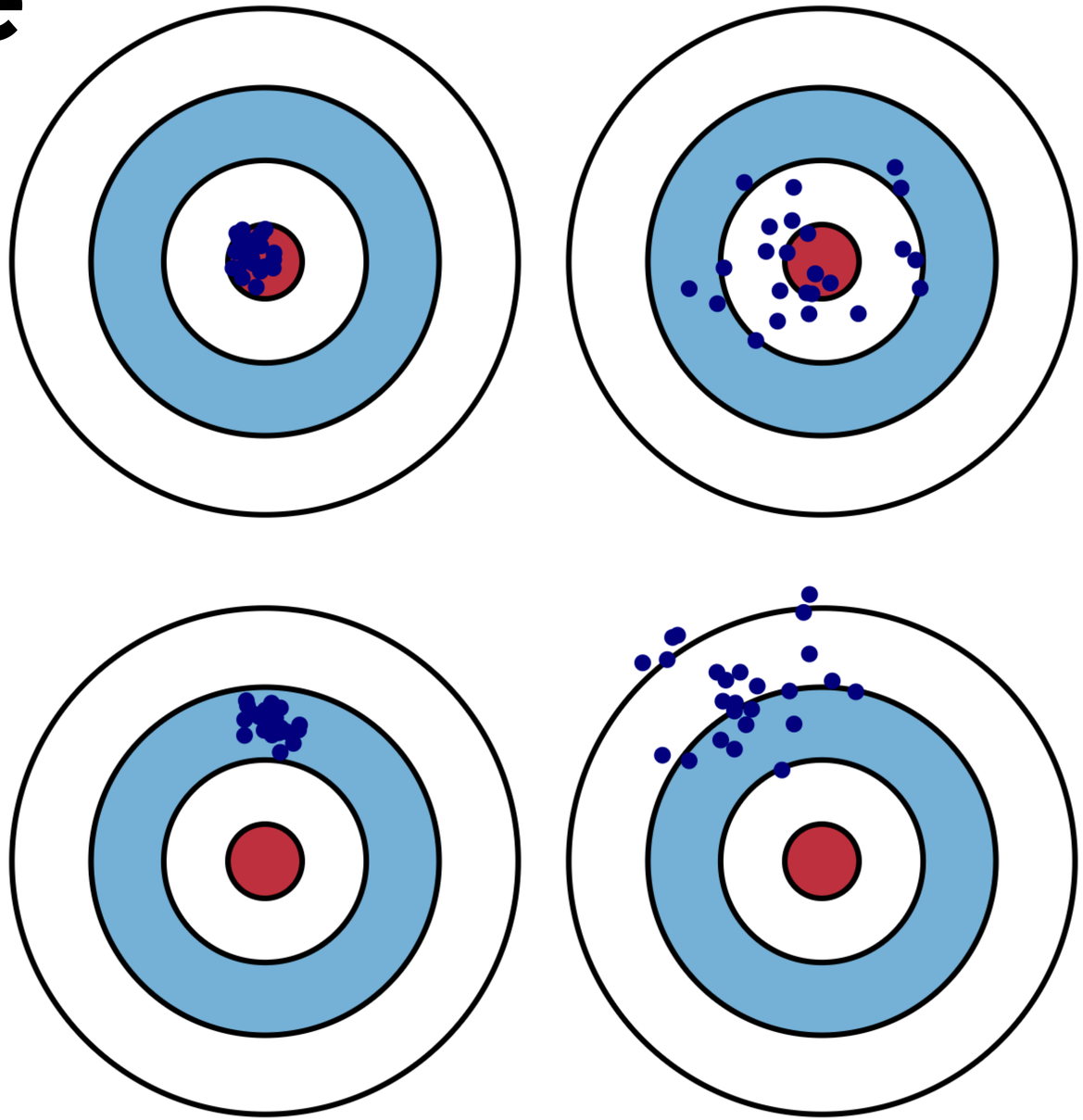
- How do we choose the right model and set its hyper parameters (e.g. λ)?
  - Use a validation set
    - Split the original data into two:
      1. Train set
      2. Validation set
        - Proxy to the test set

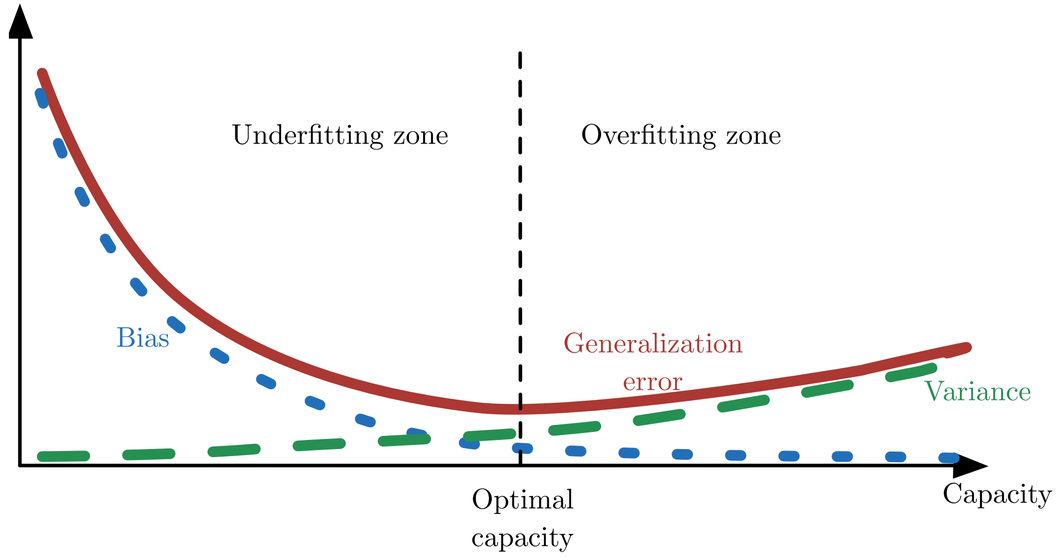


  - Train different models/hyperparameter settings on the train set
  - Pick the best according to their performance on the validation set

Bias / Variance

- The goal is to hit the bull's eye
- Each blue dot represents the "performance" of a fixed model on different data from the same distribution

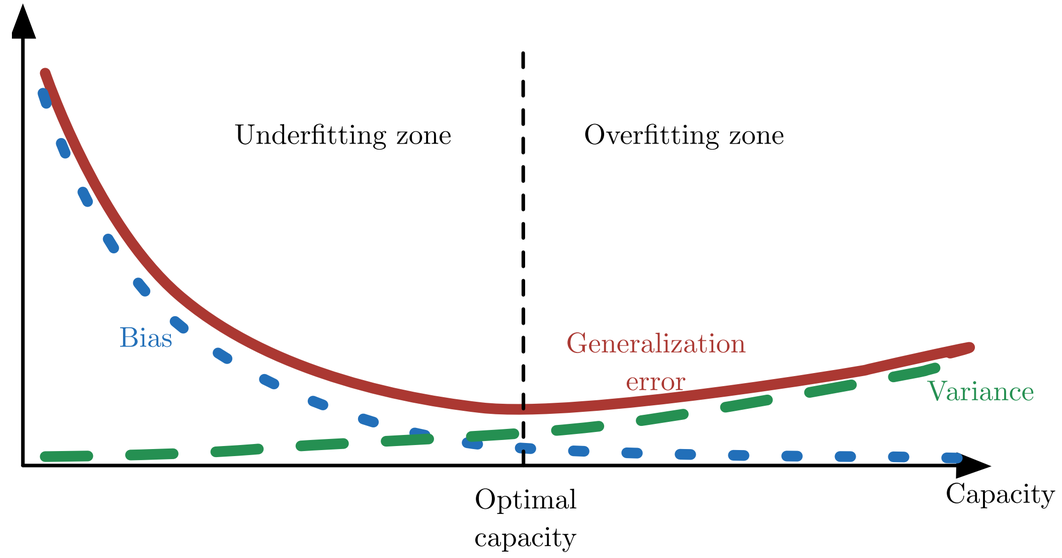


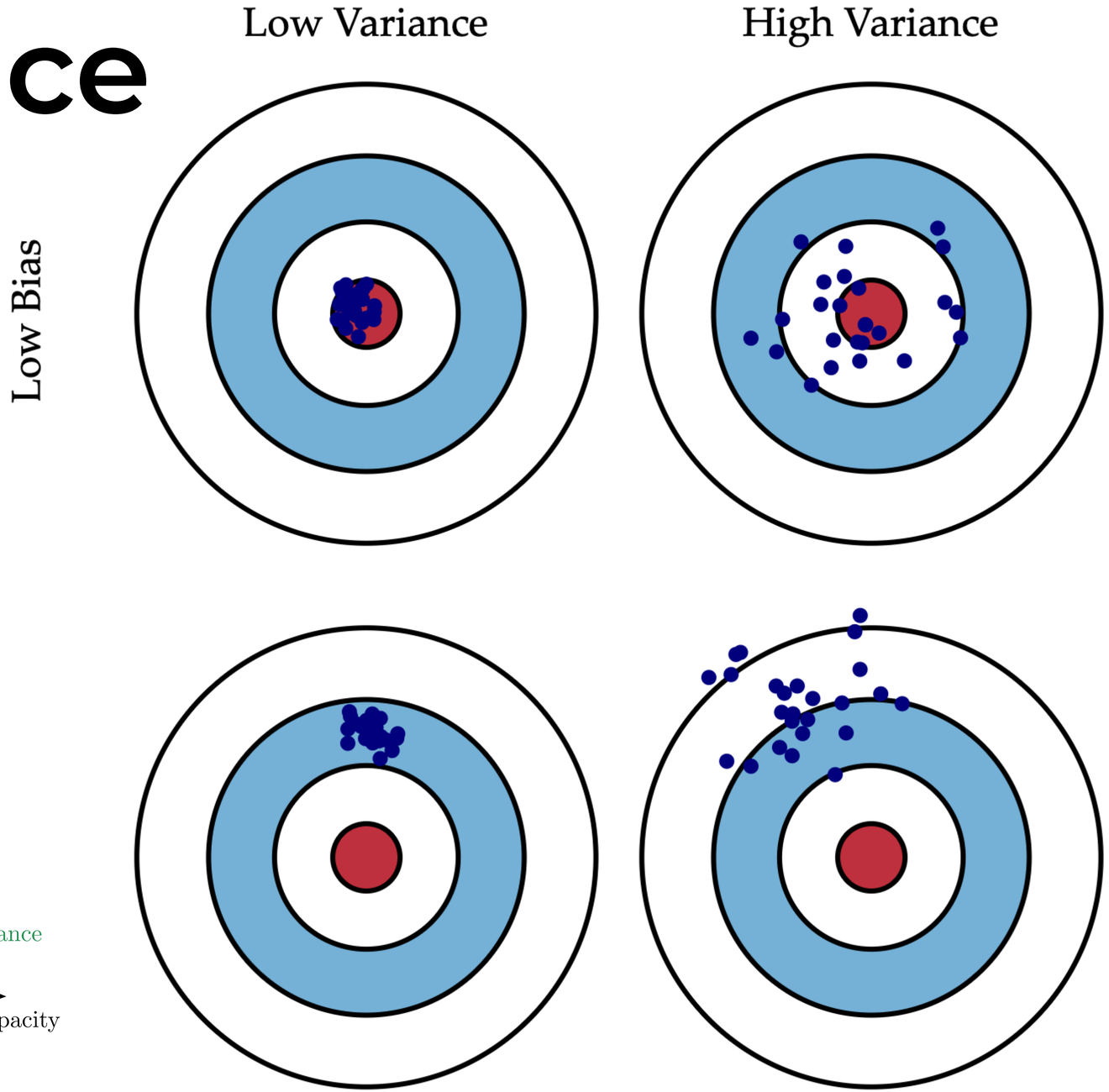


http://scott.fortmann-roe.com/docs/BiasVariance.html

Bias / Variance

- The goal is to hit the bull's eye
- Each blue dot represents the "performance" of a fixed model on different data from the same distribution





http://scott.fortmann-roe.com/docs/BiasVariance.html