

Bayesian Hyperparameter Optimization

Scaling AI Applications in Python with the Ray
Framework

Getting the Materials

<https://sr.ht/~hyphaebeast/ray-live-training/>

Introduction to Tuning Models

Introduction to Tuning Models

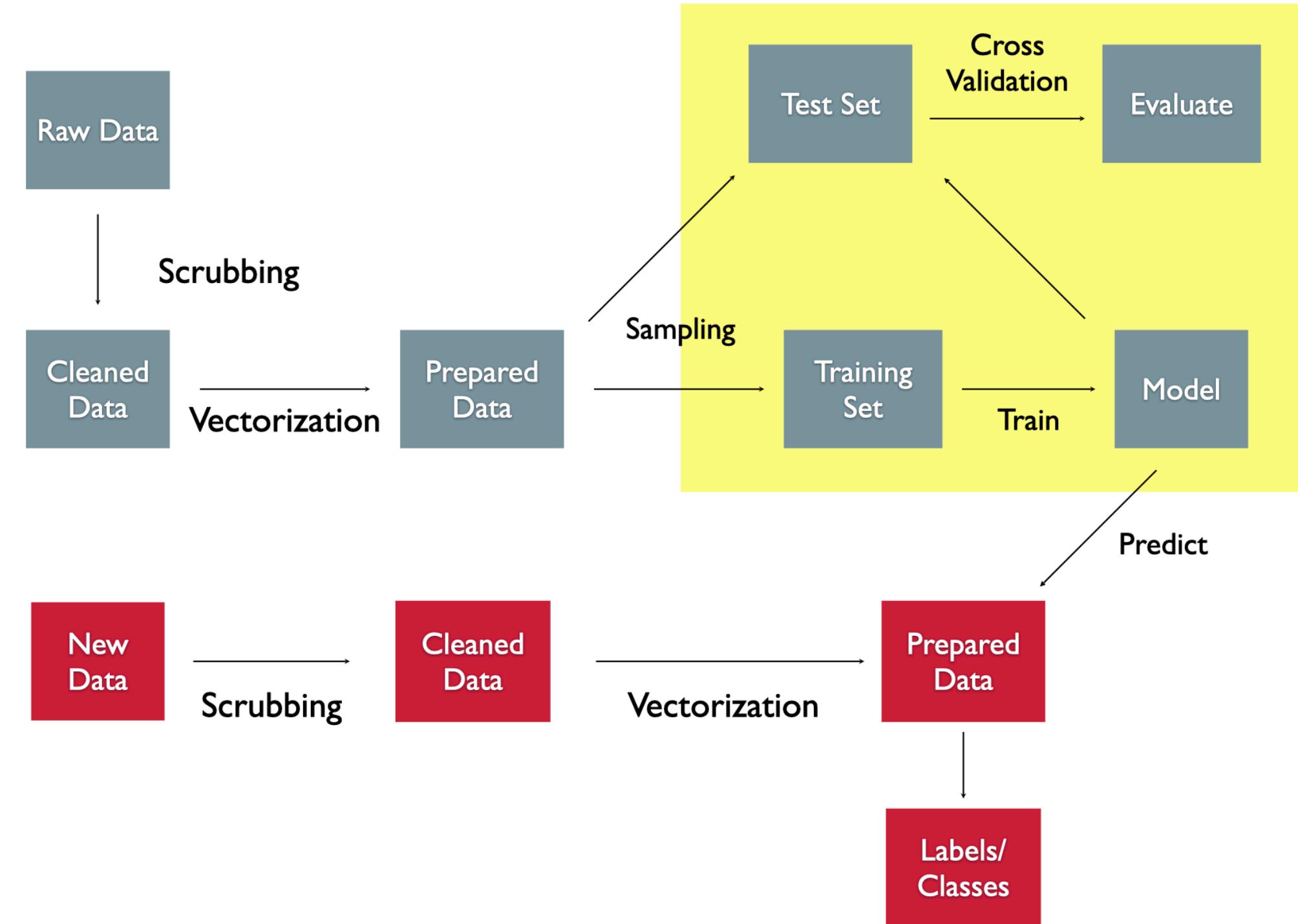
- Understanding the bias-variance trade-off

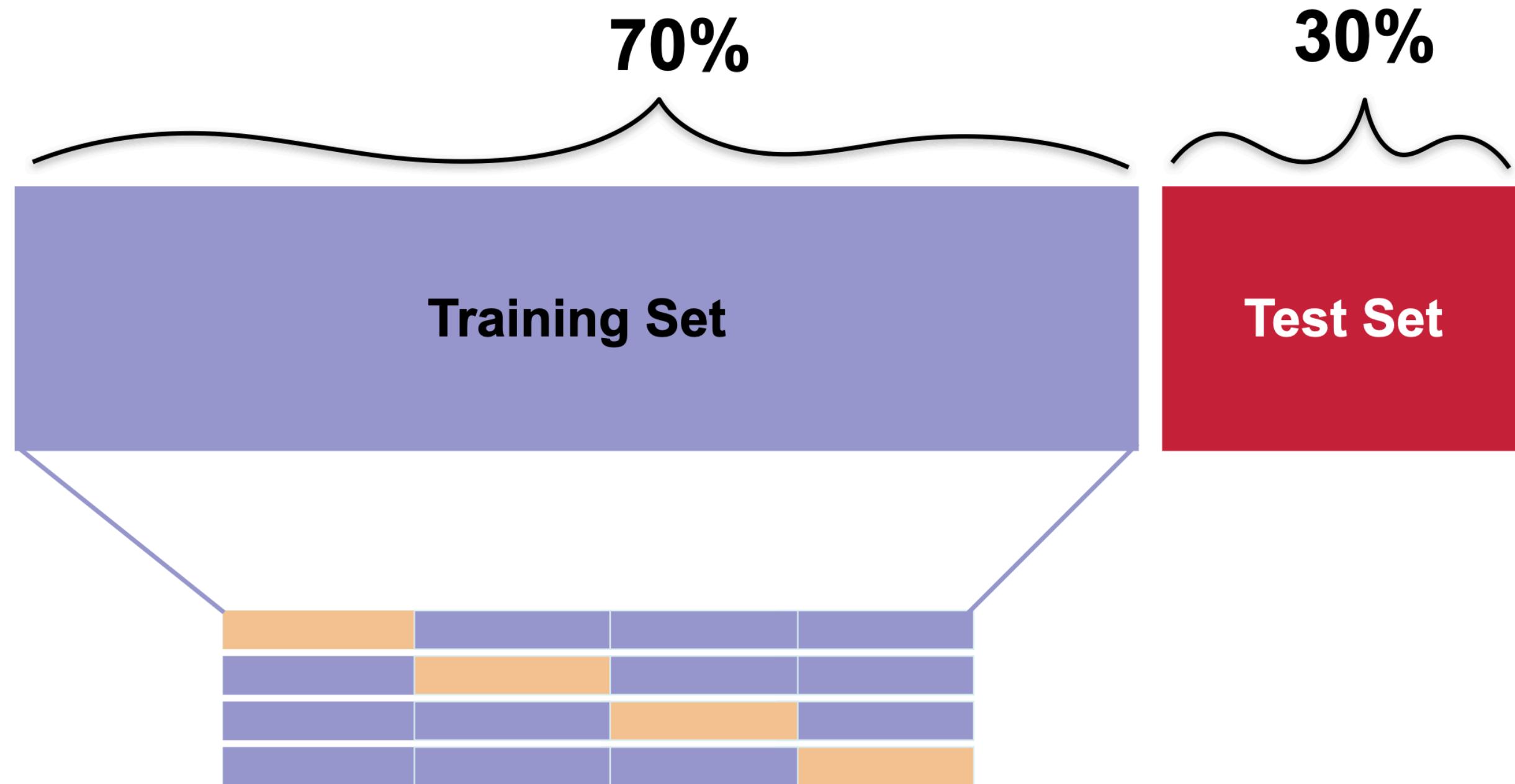
Introduction to Tuning Models

- Understanding the bias-variance trade-off
- Model Regularization

Introduction to Tuning Models

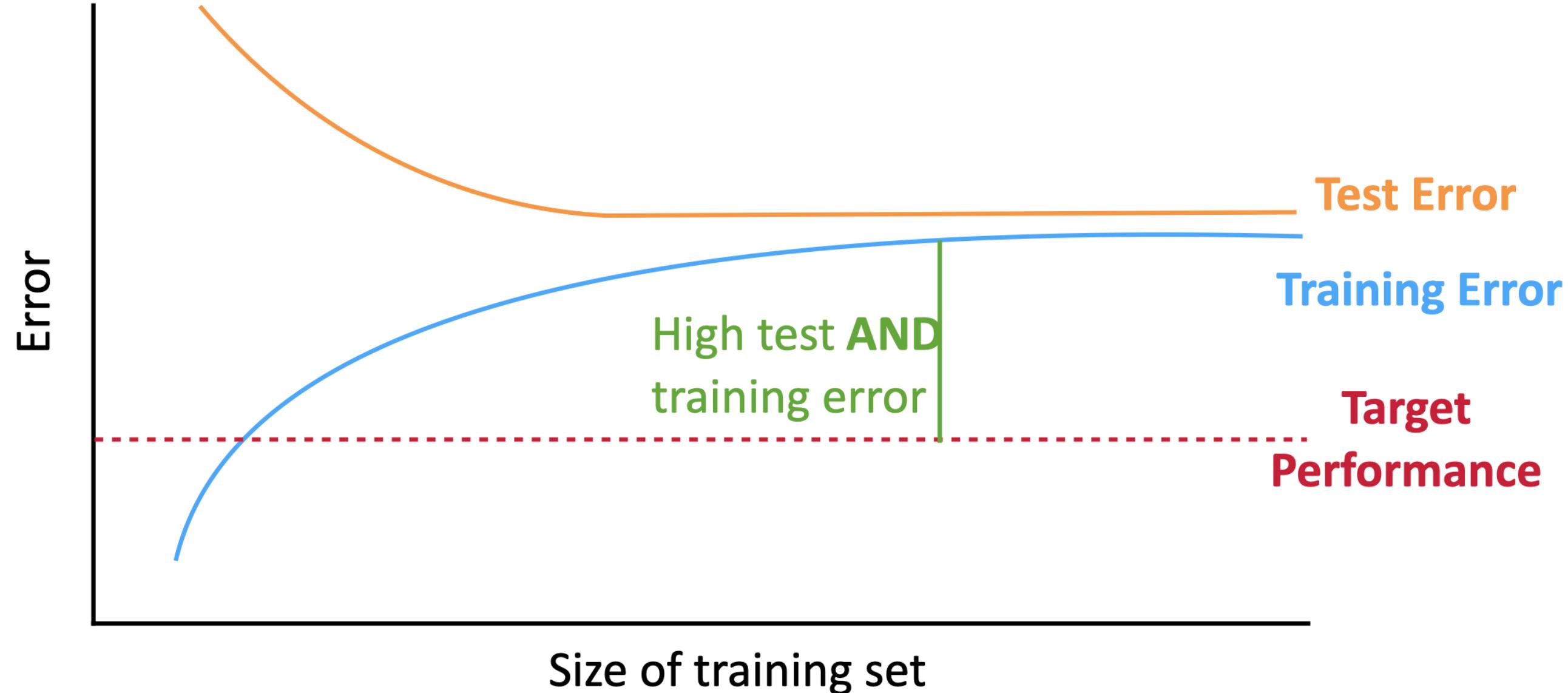
- Understanding the bias-variance trade-off
- Model Regularization
- Bruteforce Hyperparameter Optimization

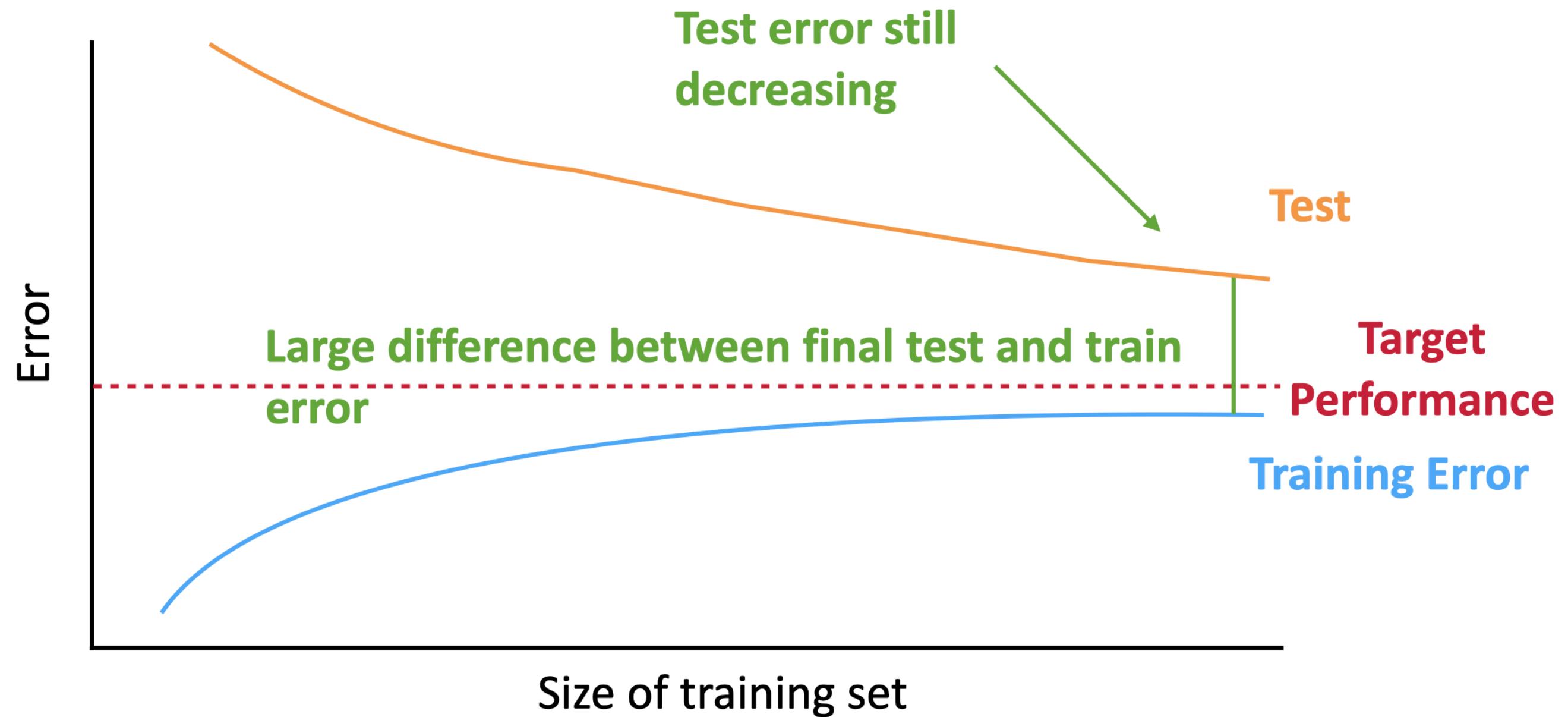


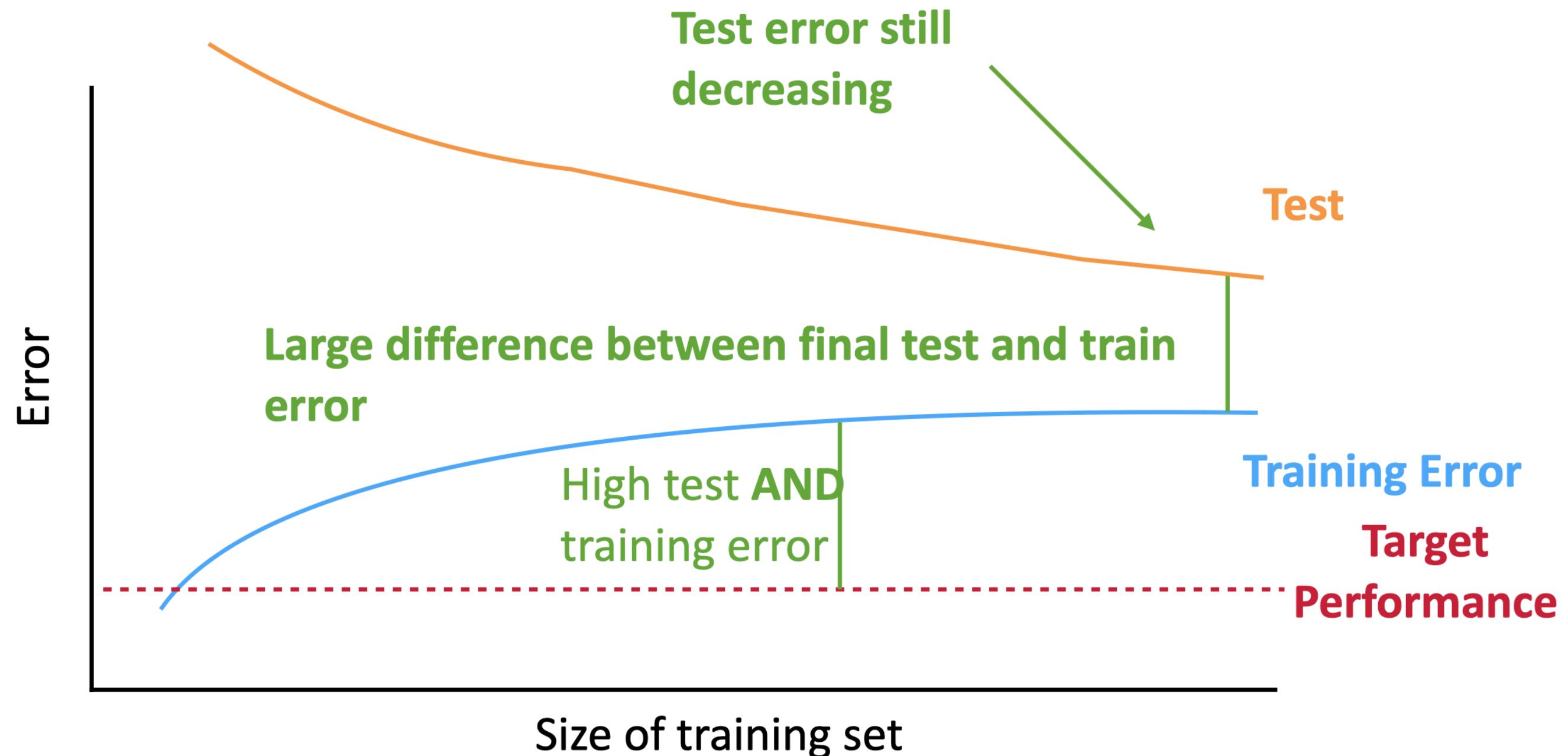


Turns







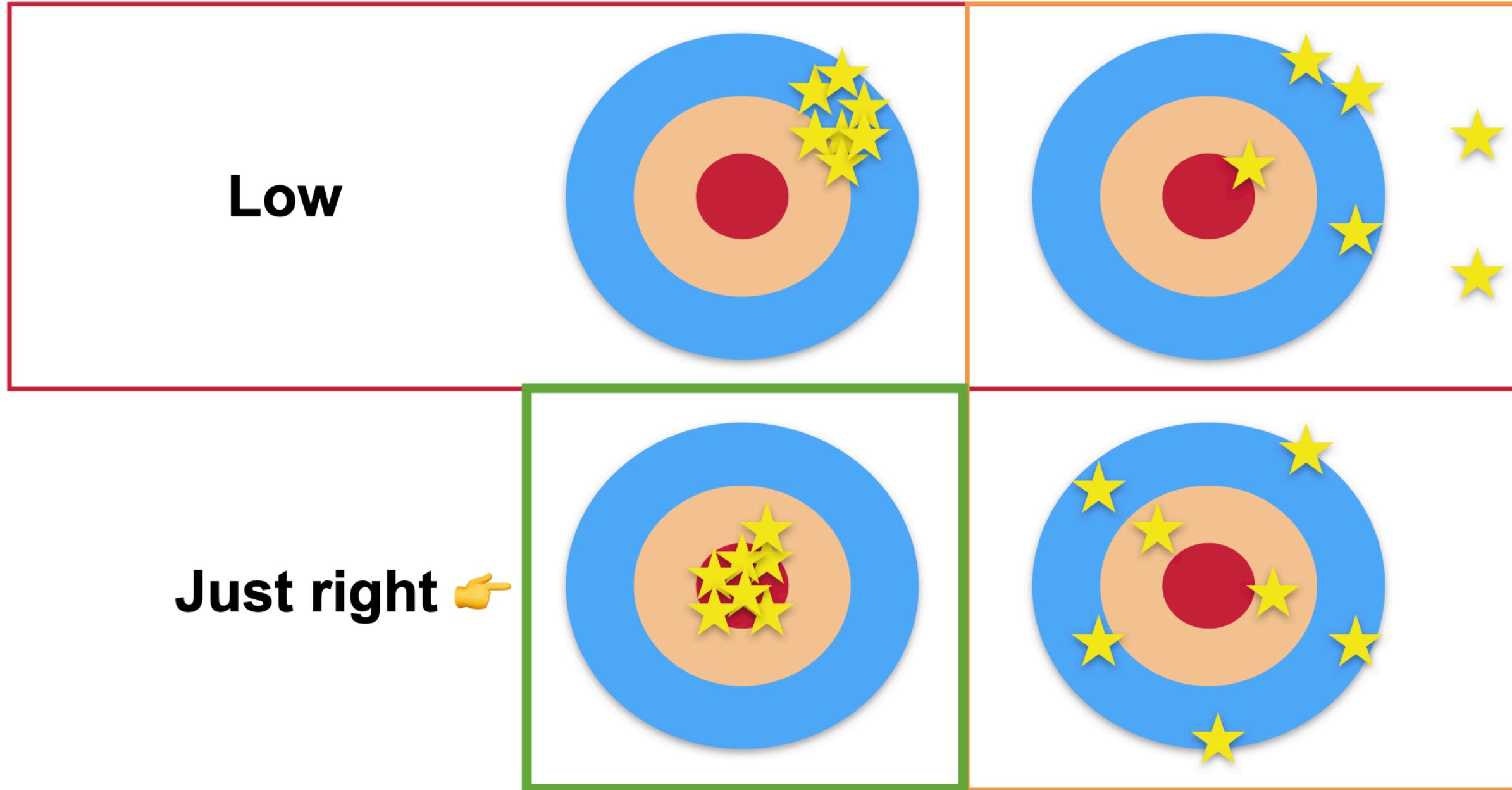


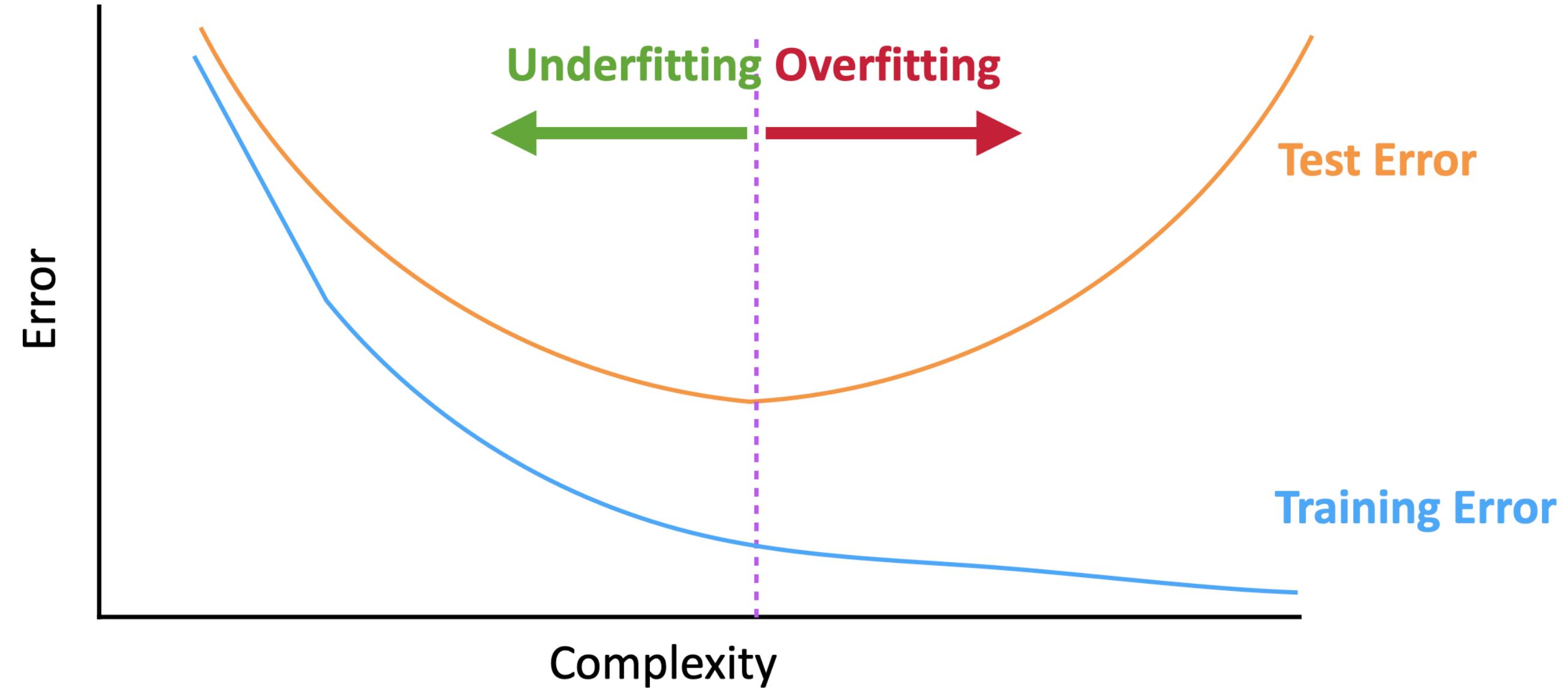
love code

High

Low

Just right👉





love code

Bayesian Hyperparameter Optimization

Bayesian Hyperparameter Optimization

- Gaussian Processes Primer

Bayesian Hyperparameter Optimization

- Gaussian Processes Primer
- Introduction to Bayesian Optimization

Bayesian Hyperparameter Optimization

- Gaussian Processes Primer
- Introduction to Bayesian Optimization
- Sequential Optimization with Ax

Why Bayesian Optimization?

Why Bayesian Optimization?

- **Function** costly to evaluate.

Why Bayesian Optimization?

- **Function** costly to evaluate.
- No closed form but (noisy) sampling is possible.

Why Bayesian Optimization?

- Function costly to evaluate.
- No closed form but (noisy) sampling is possible.
- Non-convex problem.

Why Bayesian Hyperparameter optimization

Why Bayesian Hyperparameter optimization

- Model expensive to evaluate (time, money, etc.)

Why Bayesian Hyperparameter optimization

- Model expensive to evaluate (time, money, etc.)
- No closed form but (noisy) sampling is possible.

Why Bayesian Hyperparameter optimization

- Model expensive to evaluate (time, money, etc.)
- No closed form but (noisy) sampling is possible.
- Non-convex problem.

2 sides of general optimization

2 sides of general optimization

- Objective Function (what are you optimizing)

2 sides of general optimization

- Objective Function (what are you optimizing)
- Search function (in our case Bayesian Optimization)

2 (additional) sides of Bayesian Optimization

2 (additional) sides of Bayesian Optimization

- Model of objective (approx. of response surface)

2 (additional) sides of Bayesian Optimization

- Model of objective (approx. of response surface)
- Aquisition function (sampling strategy)

BO pseudocode

for $t = 1, 2, 3, \dots, T$ do

$$x_t = \arg \max_x u(x|D_{1:t-1})$$

$$y_t = f(x_t) + \epsilon$$

$$D_{1:t} = \{D_{1:t-1}, (x_t, y_t)\}$$

$$M_{t+1} = \arg \max_\theta M_t(\theta|D_{1:t})$$

end for

T steps of BO

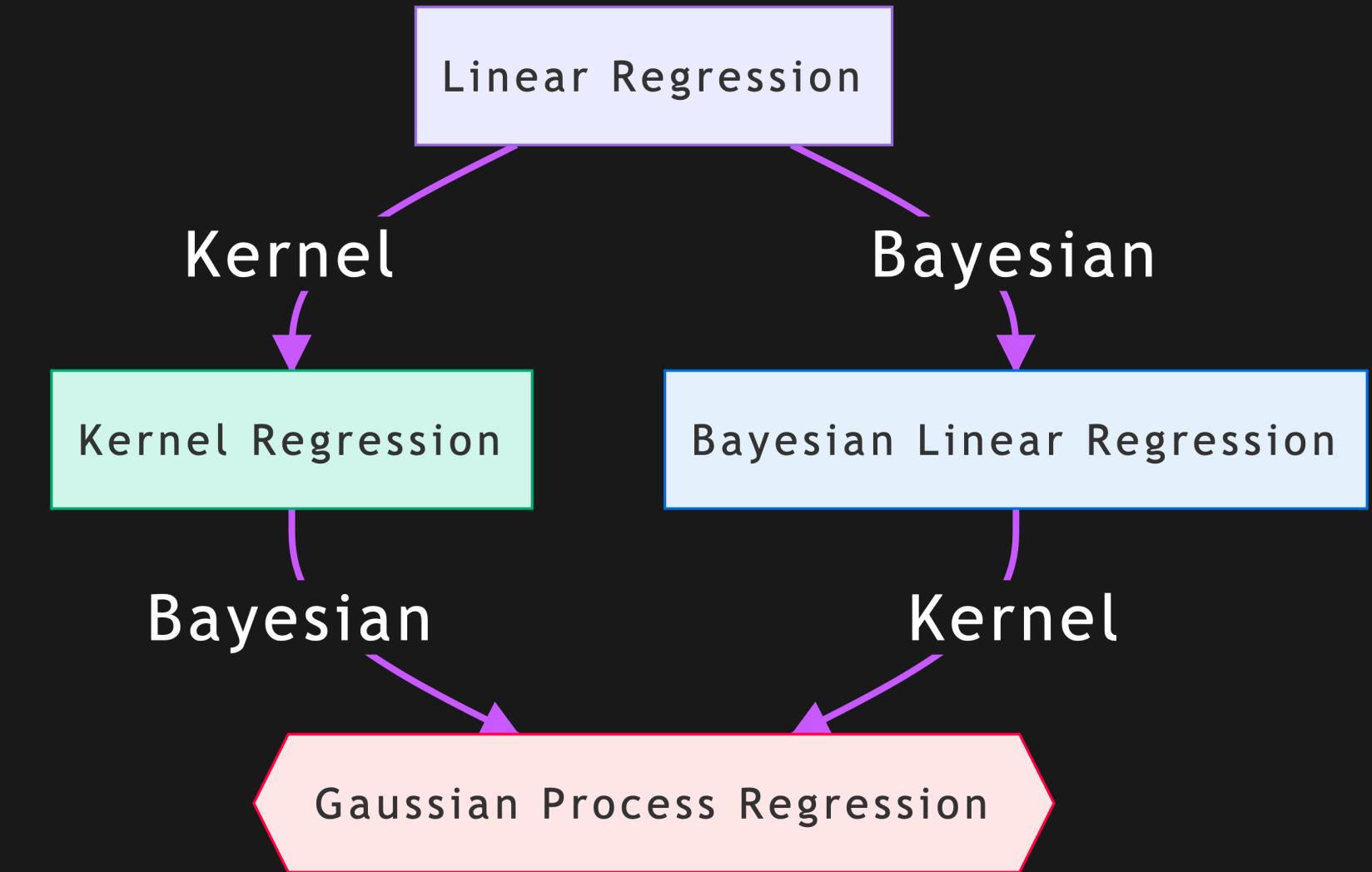
optimize over aquis. fn

evaluate objective fn

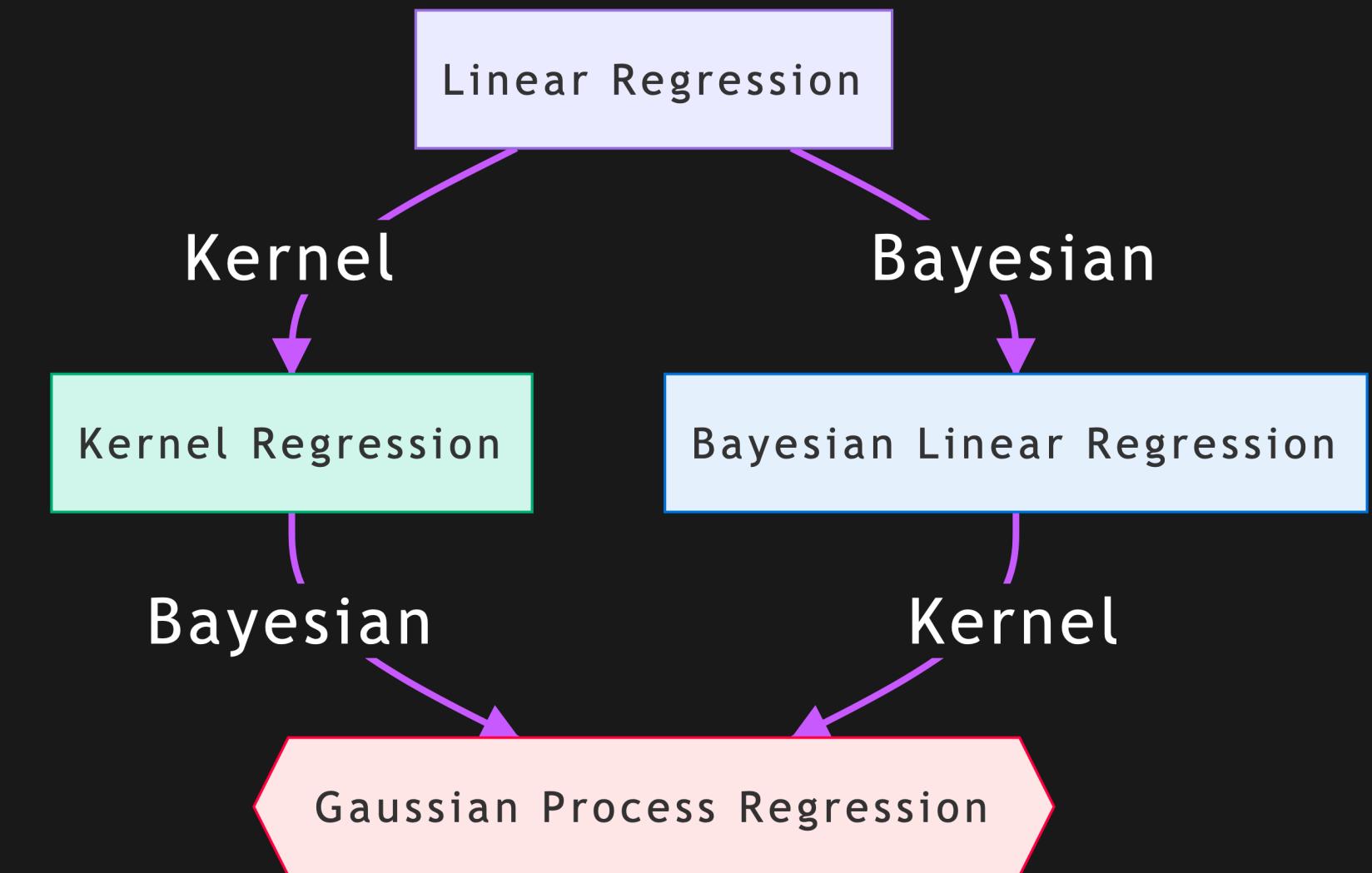
update model

love code

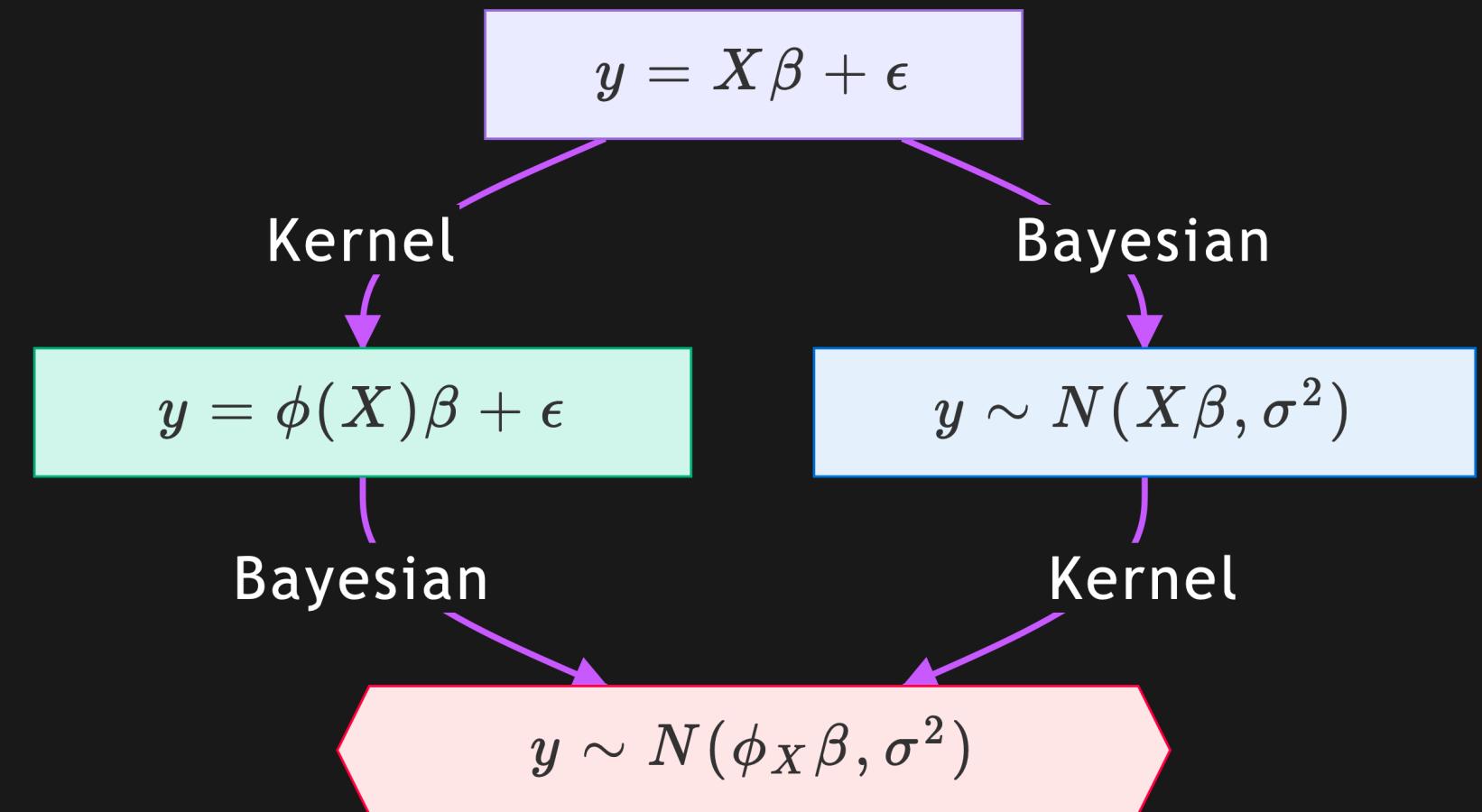
From OLS to GP



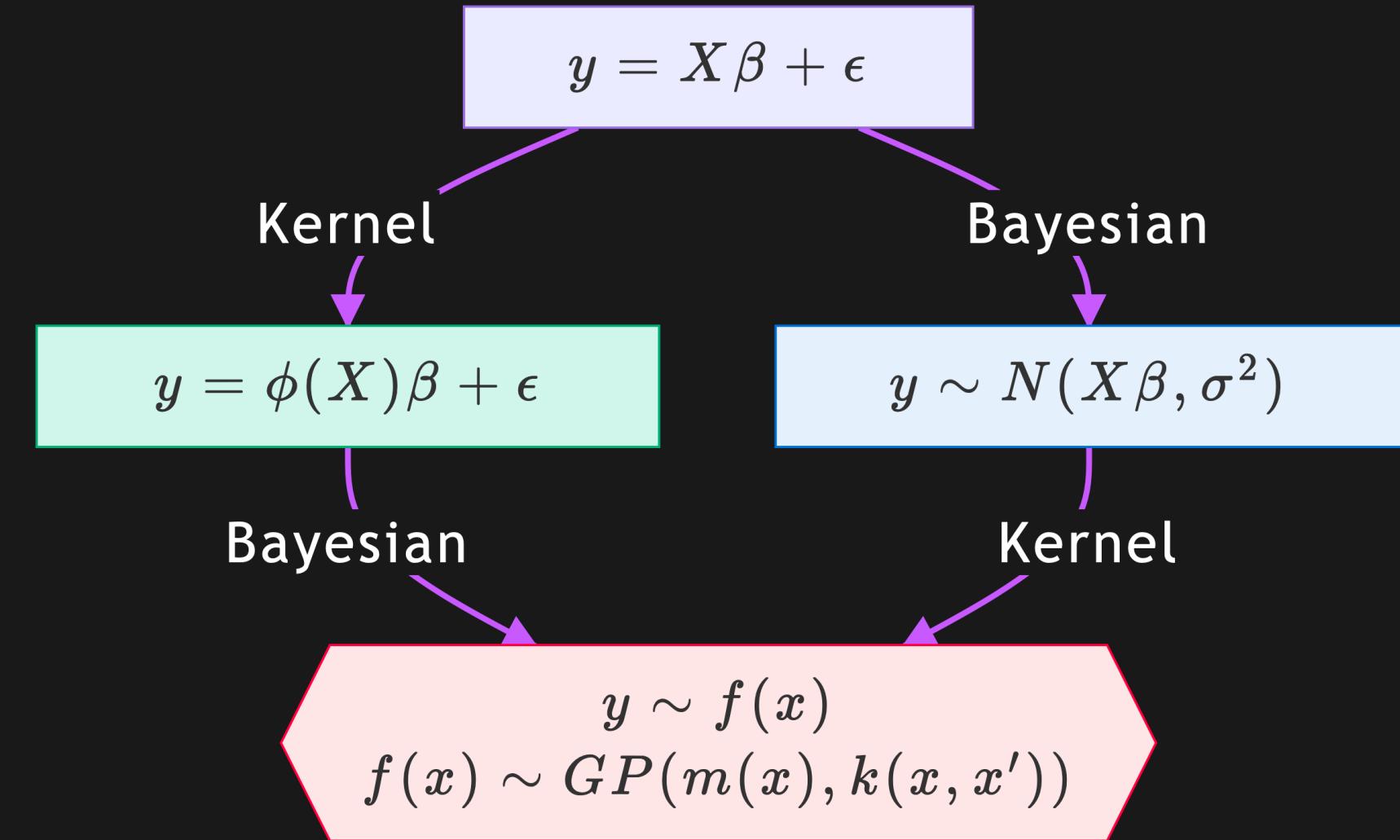
Bayesian Kernelized Linear Regression

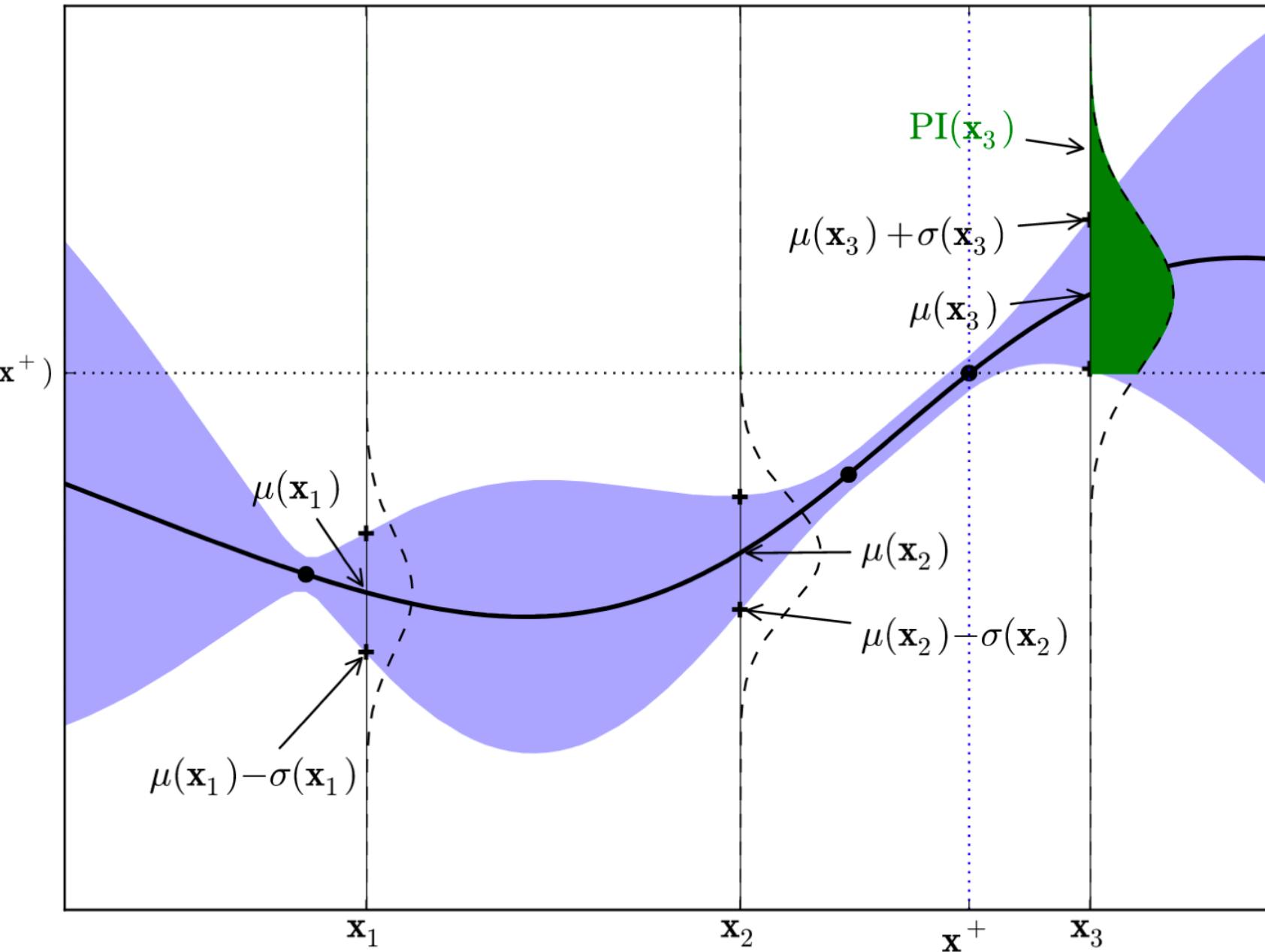


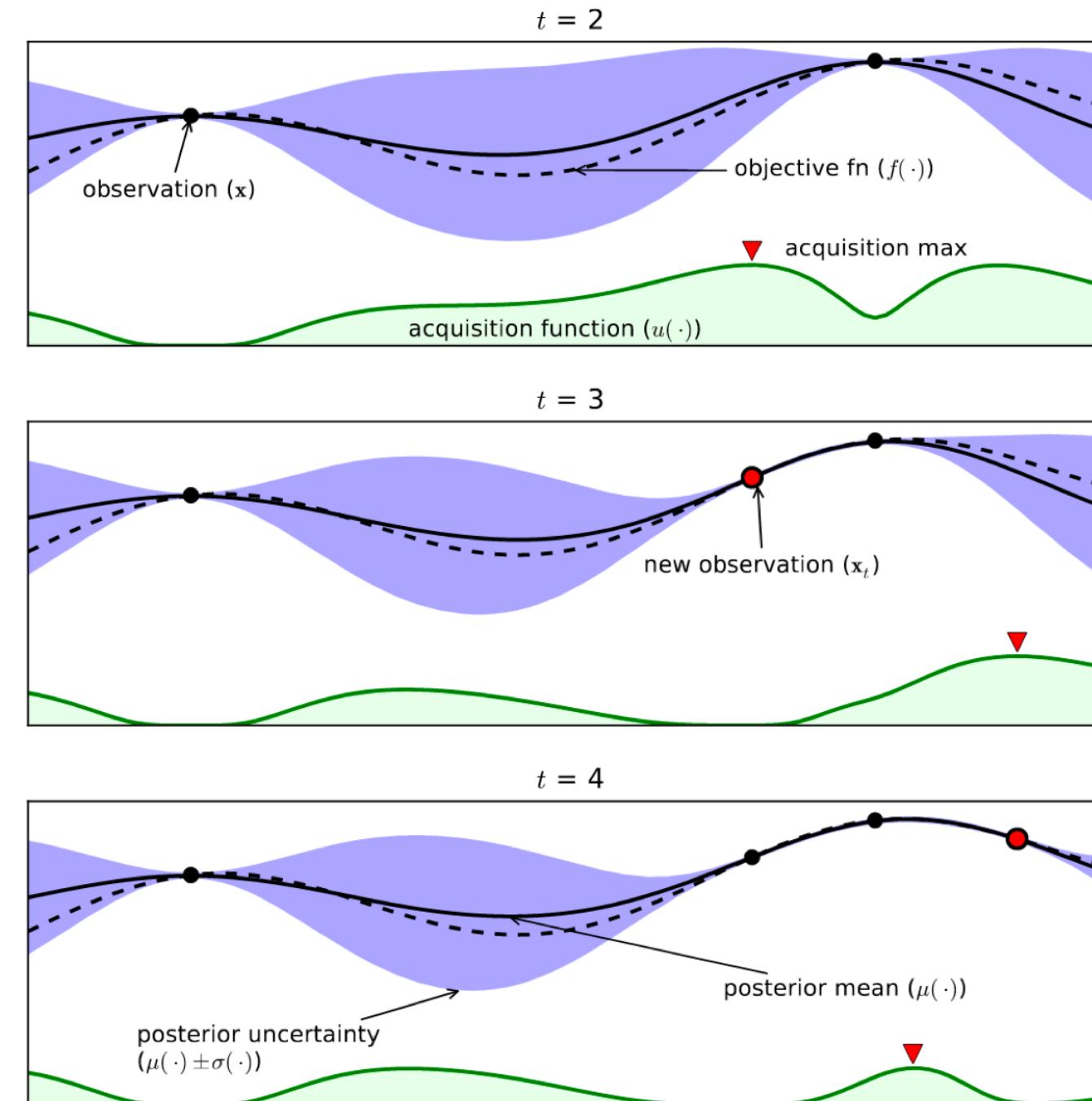
Bayesian Regression with kernelized features

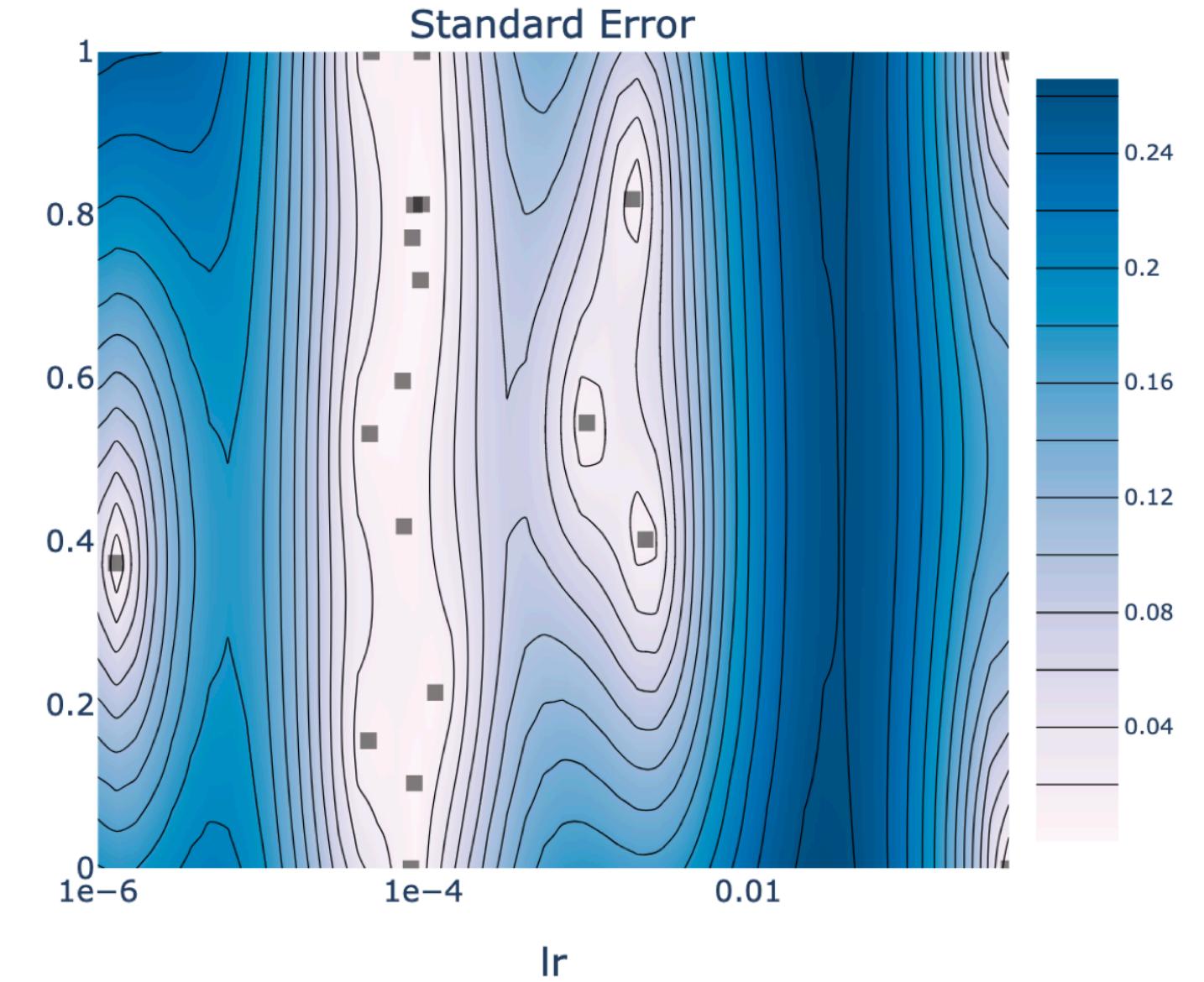
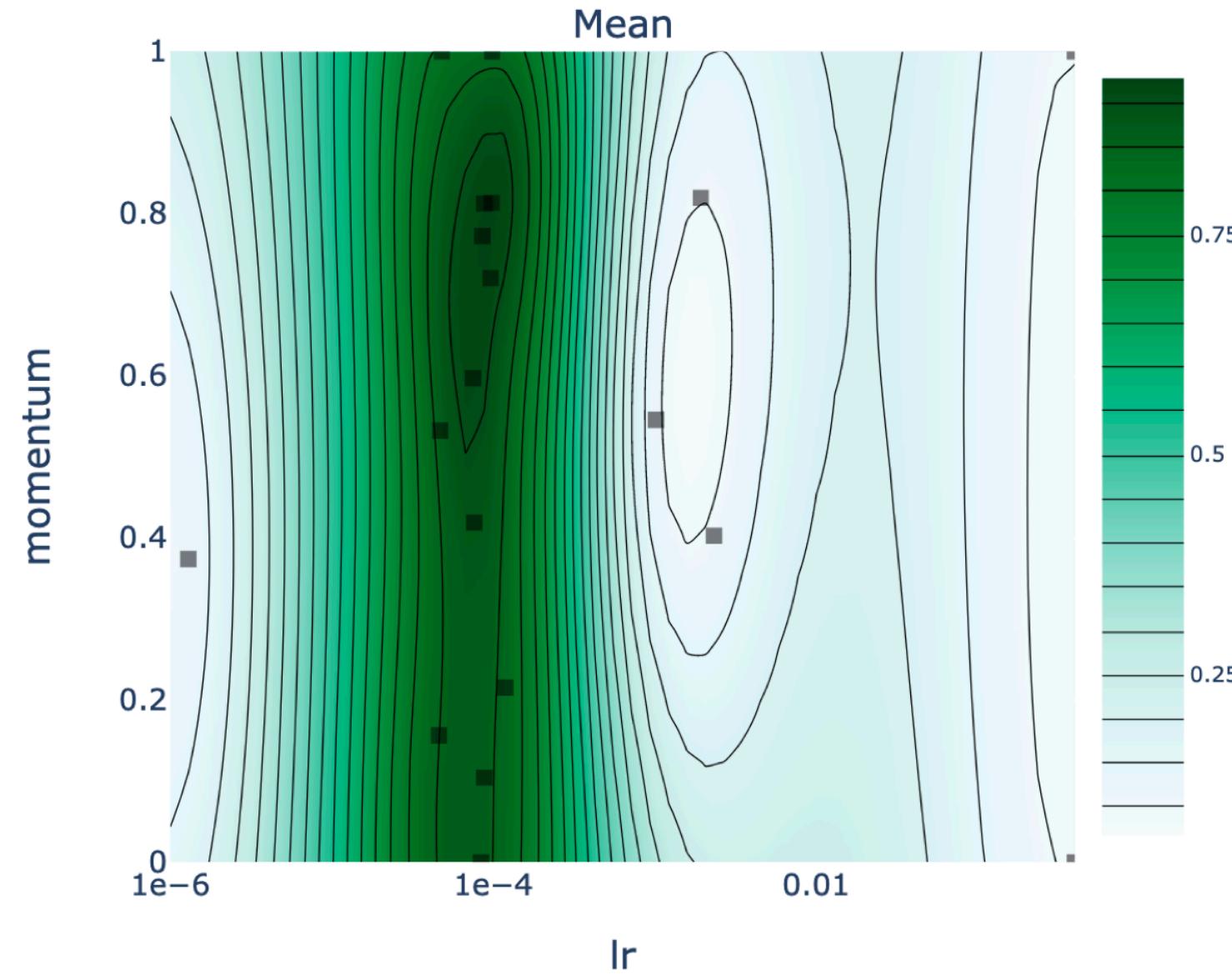


Gaussian Process prior over function space

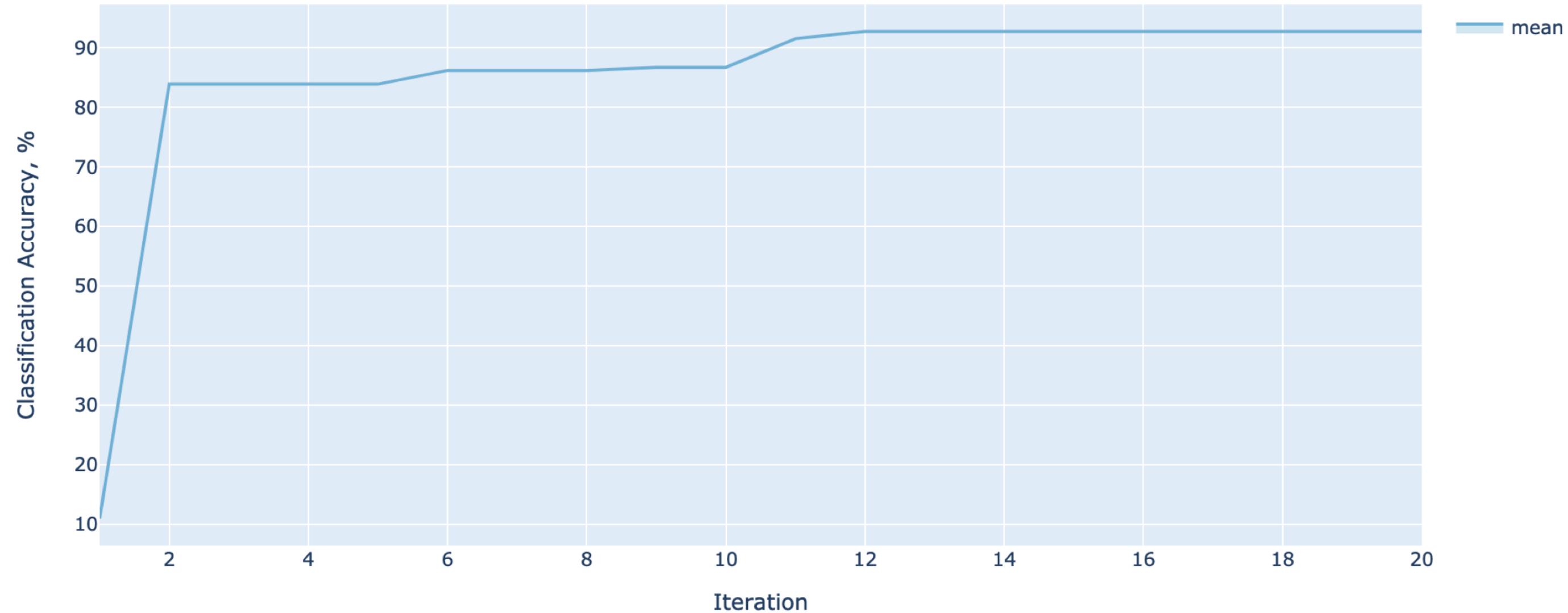








Model performance vs. # of iterations



Distributed Hyperparameter Optimization with Ray Tune

love code

Q & A

Materials, mailing list, questions, feedback 

<https://sr.ht/~hyphaebeast/ray-live-training/>