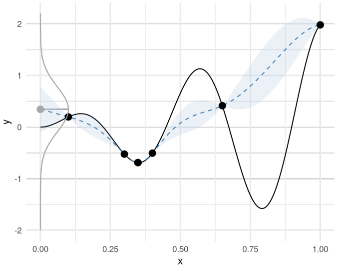
Optimization in Machine Learning

Bayesian Optimization
Posterior Uncertainty and Acquisition Functions I



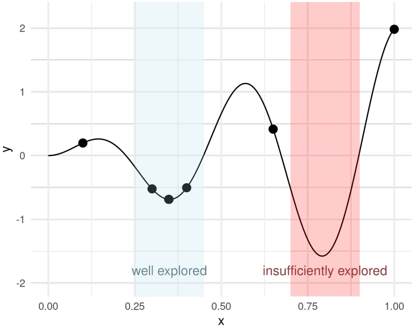
Learning goals

- Bayesian surrogate modeling
- Acquisition functions
- Lower confidence bound

BAYESIAN SURROGATE MODELING

Goal:

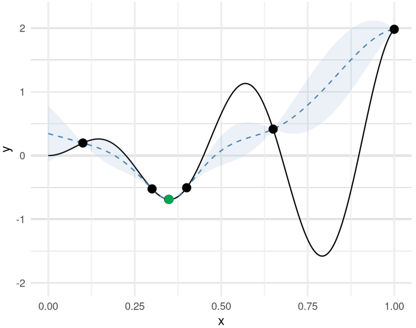
Find trade-off between **exploration** (areas we have not visited yet) and **exploitation** (search around good design points)



ACQUISITION FUNCTIONS

To sequentially propose new points based on the SM, we make use of so-called acquisition functions $a : \mathcal{S} \to \mathbb{R}$

Let $f_{\min} := \min \left\{ f(\mathbf{x}^{[1]}), \ldots, f(\mathbf{x}^{[t]}) \right\}$ denote the best observed value so far (visualized in green - we will need this later!)



In the examples before we simply used the posterior mean $a(\mathbf{x}) = \hat{f}(\mathbf{x})$ as acquisition function - ignoring uncertainty

LOWER CONFIDENCE BOUND

Goal: Find $\mathbf{x}^{[t+1]}$ that minimizes the **Lower Confidence Bound** (LCB):

$$a_{\text{LCB}}(\mathbf{x}) = \hat{f}(\mathbf{x}) - \tau \hat{s}(\mathbf{x})$$

where $\tau > 0$ is a constant that controls the “mean vs. uncertainty” trade-off

The LCB is conceptually very simple and does **not** rely on distributional assumptions of the posterior predictive distribution under a SM