Module 12: Bayesian Optimisation

Learning outcomes

1. Analyse the effects of hyperparameters on surrogate models.
2. Analyse the role of acquisition functions in Bayesian optimisation.
3. Apply surrogate models to support decision-making in Bayesian optimisation.
4. Evaluate the feasibility and applications of Bayesian optimisation in real-world contexts.
5. Apply ML techniques to optimise black-box functions.

Bayesian optimisation

An optimisation method for black-box functions where querying the function is costly or time-consuming. It builds a surrogate model, usually a Gaussian process (GP), to approximate the function and uses an acquisition function to decide where to evaluate next.

Surrogate models

Approximate the true objective to make optimisation more efficient. Common surrogate models include GPs and regression trees.

- GP
  - It is a non-parametric probabilistic model.
  - It is defined by a mean function $m(x)$, representing expected values, and a covariance (kernel) function $K(x,x')$, capturing relationships between points.
  - Mathematically, it is written as $f(x) \sim GP(m(x), K(x,x'))$.
- Regression tree
  - It is a decision-tree-based model for predicting continuous values.
  - Parameters include `max_depth`, `min_samples_split`, criterion, etc.

Acquisition functions

Acquisition functions guide the optimiser in selecting the next point to evaluate. They are affordable for computing and are optimised instead of the actual objective.
The common acquisition functions are:

- Expected improvement (EI)
  - For maximisation: $EI(x) = E(max[0, f(x) - f_{best}])$
  - For minimisation: $EI(x) = E(max[0, f_{best} - f(x)])$

  In both cases, f_{best} is the best function value seen so far. The direction of improvement (higher vs lower) changes based on your goal.
- Probability of improvement (PI)
  - For maximisation: $PI(x) = P(f(x) > f_{best} + \xi)$

Imperial College Business School | Executive Education

  - For minimisation: $PI(x) = P(f(x) > f_{best} - \xi)$

  The small positive value ξ is used to encourage exploration. Again, the inequality flips based on the optimisation direction.
- Upper confidence bound (UCB) and lower confidence bound (LCB)
  - UCB is typically used for maximisation: $UCB(x) = \mu(x) + \kappa.\sigma(x)$
  - For minimisation, use LCB: $LCB(x) = \mu(x) - \kappa.\sigma(x)$
  - The parameter κ controls the balance between exploration (large κ) and exploitation (small κ).

Exploration vs exploitation

- Exploitation: sampling points with high predicted performance
- Exploration: sampling points with high uncertainty
- Acquisition functions combine to guide the optimiser towards promising yet underexplored areas.

When to stop tuning

- Validation performance plateaus or becomes inconsistent.
- Cross-validation scores show increasing variance.
- Model becomes overly complex or hard to interpret.
- A pre-set resource or iteration limit is reached.