Anabel Yong

ucabyon@ucl.ac.uk — +447845636286 — Personal Website: https://anabelyong.github.io/

Summary

MSc in Computational Statistics & Machine Learning from UCL with expertise in Gaussian Processes, Bayesian Optimisation, and kernel-based methods. Interested in machine learning for life sciences discovery and learning biological data.

Education

MSc Computational Statistics & Machine Learning, UCL (Distinction, Dean's List) Sept 2023 - Sept 2024
Modules: Probabilistic & Unsupervised Learning (Gatsby PhD module) (91%), Machine Learning Seminar (Marc Deisenroth- DeepMind)(76%), Numerical Optimization (80%), Statistical Natural Language Processing

BSc Biochemistry (with Mathematics), University of Edinburgh October 2019 - June 2023
Modules: Bioinformatics (95%), Machine Learning (80%), Numerical Methods for Differential Equations (85%), Linear Programming, Modelling & Solutions (76%), Mathematics for Natural Sciences (80%)

Work Experience

- **Visiting Machine Learning Research Scholar, National University of Singapore Jonathan Scarlett** May 2024 - Present
  - Implemented biological sequence kernels, Inverse Quadratic Hamming kernels for amino acid sequences.
  - Designing an active learning algorithm for mutation data.
- **Machine Learning Research Intern, IgnotaLabs.AI, Dr.Layla-Hosseini-Gerami, Dr.Austin Tripp, Cambridge MLG Group & Brooks Paige, UCL** June 2024- Sept 2024
  - Developed a novel multi-objective Bayesian optimization algorithm (GP-MOBO) using Tanimoto/MinMax Kernels for small molecule optimization. (Paper pending submission.)
  - Deployed a Python package (KERN-GP), enabling hardware-optimized Gaussian processes for molecular property prediction, with scalability tested against GPU-based PyTorch models.
- **Research Assistant, Kustatscher Lab, University of Edinburgh x Technical University of Munich** Sept 2022 - July 2023
  - Benchmarked a data pre-processing pipeline for proteomics, pgFDR), to enhance protein detection accuracy.
  - Aided in database curation of human proteins and microproteins of ProteomeHD2
- **Statistical Genetics Intern, ZiHeng Yang Lab, UCL Centre of Computational Biology, UCL** March 2022 - May 2022
  - Developed Bayesian MCMC algorithms to model ABO blood type frequencies, contributing to population genetics and personalised medicine approaches.
- **Research Intern, Grima Lab, Department of Systems & Synthetic Biology, University of Edinburgh** June 2020 - Jan 2022
  - Implemented stochastic and deterministic models to predict mRNA decay kinetics, aiding in the understanding of gene regulation mechanisms and supporting drug efficacy studies in preclinical research.

Personal Projects (on Portfolio)

- **DoLa-based Decoding for Instruction-Following LLMs**
  - Evaluated and adapted DoLa contrastive decoding in T5 and FLAN-T5, enhancing instruction accuracy and reducing hallucinations for large language models (LLMs).

Skills

Programming: Python, C++, R, MATLAB, git, Linux
Machine Learning: PyTorch, JAX, BoTorch, GPyTorch
Bioinformatics/Cheminformatics: RDKit, PyMOL, UniProt, MaxQuant
Cloud Computing & HPC: Google Cloud Platform (GCP), AWServices (AWS)