CONTENTS

1 Introduction 1
1.1 Goals 4
1.2 Outline 6
2 Background 9
2.1 Probabilistic Machine Learning 9
2.2 Variational Inference 10
2.2.1 Evidence Lower Bound 11
2.3 Statistical Divergences and Density-Ratio Estimation . 12
2.3.1 Variational Divergence Estimation 14
2.3.2 Class-Probability Estimation 15
2.4 Gaussian Processes 18
2.4.1 Gaussian Process Regression 18
2.4.2 Sparse Gaussian Processes 23
2.4.3 Random Fourier Features 28
2.5 Bayesian Optimisation 32
2.5.1 Surrogate Models 35
2.5.2 Acquisition Functions 37
Addendum 41
2.A KL Divergence Simplification 41
2.B Optimal Variational Distribution for General Likelihoods 41
2.C Intermediate Lower Bound for Gaussian Likelihoods . 42
2.D Optimal Variational Distribution for Gaussian Likelihoods 42
2.E Collapsed Lower Bound for Gaussian Likelihoods . . . 43
2.F Spectral Density of the Squared Exponential Kernel . . 43
2.G Cosine Difference as Inner Product 44
3 Orthogonally-Decoupled Sparse Gaussian Processes with Spherical Neural Network Activation Features 45
3.1 Introduction 45
3.2 Inter-Domain Inducing Features 46
3.2.1 Spherical Harmonics Inducing Features 47
3.2.2 Spherical Neural Network Inducing Features . 48
3.3 Orthogonally Decoupled Inducing Points 50
3.4 Methodology 53
3.5 Experiments 58
3.5.1 Synthetic 1D Dataset 58
3.5.2 Regression on UCI Repository Datasets 59
3.5.3 Large-scale Regression on Airline Delays Dataset 61
3.6 Summary 63
Addendum 65
3.A Experimental Set-up and Implementation Details . . . 65
3.A.1 Hardware 65
3.A.2 Software 65

3.A.3 Hyperparameters 65
3.B Additional Results 66
3.B.1 Regression on Airline Delays Dataset 66
3.B.2 Extra UCI Repository Datasets 66
4 Cycle-Consistent Generative Adversarial Networks as a Bayesian Approximation 69
4.1 Introduction 69
4.2 Implicit Latent Variable Models 70
4.2.1 Prescribed Likelihood 71
4.2.2 Implicit Prior 71
4.3 Variational Inference 72
4.3.1 Prescribed Variational Posterior 72
4.3.2 Reverse KL Variational Objective 73
4.3.3 Approximate Divergence Minimisation 73
4.4 Symmetric Joint-Matching Variational Inference 75
4.4.1 Variational Joint 75
4.4.2 Forward KL Variational Objective 76
4.5 CycleGAN as a Special Case 77
4.5.1 Basic CycleGAN Framework 78
4.5.2 Cycle-consistency as Conditional Entropy Maximisation 79
4.5.3 Distribution Matching as Approximate Divergence Minimisation 80
4.6 Related Work 83
4.7 Experiments 84
4.8 Summary 85
Addendum 89
4.A Relation to KL Importance Estimation Procedure (KLIEP) 89
4.B Summary of Definitions 90
5 Bayesian Optimisation by Classification with Deep Learning and Beyond 93
5.1 Introduction 93
5.2 Optimisation Policies and Density-Ratio Estimation . . 95
5.2.1 Relative Density-Ratio 95
5.2.2 Improvement-based Acquisition Functions . . . 95
5.2.3 Tree-structured Parzen Estimator 98
5.2.4 Potential Pitfalls 98
5.3 Bayesian Optimisation by Probabilistic Classification . 99
5.3.1 Choice of Proportion γ 102
5.3.2 Choice of Probabilistic Classifier 104
5.3.3 Likelihood-Free BO by Weighted Classification 105
5.4 Experiments 106
5.4.1 Neural Network Tuning (HPOBench) 106
5.4.2 Neural Architecture Search (NASBench201) . . 109
5.4.3 Robot Arm Pushing 109
5.4.4 Racing Line Optimisation. 112

5.4.5 Ablation Studies 112
5.5 Discussion 113
5.6 Summary 116
Addendum 117
5.A Relative Density-Ratio: Unabridged Notation 117
5.B Class-posterior Probability 118
5.C Log Loss 118
5.C.1 Optimum 119
5.C.2 Empirical Risk Minimisation 119
5.D Implementation of Baselines 120
5.E Experimental Set-up and Implementation Details . . . 120
5.E.1 BORE-RF 121
5.E.2 BORE-XGB 121
5.E.3 BORE-MLP 122
5.F Details of Benchmarks 122
5.F.1 HPOBench 122
5.F.2 NASBench201 122
5.F.3 Robot pushing control 123
5.F.4 Racing Line Optimisation 124
5.G Parameters, hyperparameters, and meta-hyperparameters124
5.G.1 Parameters 125
5.G.2 Hyperparameters 126
5.G.3 Meta-hyperparameters 127
6 Conclusion 129
6.1 Summary of Contributions 129
6.2 Future Directions 130
6.3 Final Reflection 131
A Numerical Methods for Improved Decoupled Sampling of Gaussian Processes 133
A.1 Introduction 133
A.2 Decoupled Sampling of Gaussian Processes 134
A.3 Numerical Integration for GP Prior Approximations . 137
A.3.1 Monte Carlo Estimation 138
A.3.2 Quasi-Monte Carlo 139
A.3.3 Quadrature 140
A.3.4 Other Approaches 148
A.4 Experiments 149
A.4.1 Prior Approximation 149
A.4.2 Posterior Sample Approximation 154
A.5 Summary 156
Addendum 159
A.A Product-to-Sum Identity 159
A.B Zero in Expectation 159

Bibliography 161