

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	14
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
2.6	Summary	41
	Addendum	43
2.A	KL Divergence Simplification	43
2.B	Optimal Variational Distribution for General Likelihoods	43
2.C	Intermediate Lower Bound for Gaussian Likelihoods	44
2.D	Optimal Variational Distribution for Gaussian Likelihoods	44
2.E	Collapsed Lower Bound for Gaussian Likelihoods	45
2.F	Spectral Density of the Squared Exponential Kernel	45
2.G	Cosine Difference as Inner Product	46
3	Orthogonally-Decoupled Sparse Gaussian Processes with Spherical Neural Network Activation Features	47
3.1	Introduction	47
3.2	Inter-Domain Inducing Features	48
3.2.1	Spherical Harmonics Inducing Features	49
3.2.2	Spherical Neural Network Inducing Features	51
3.3	Orthogonally Decoupled Inducing Points	52
3.4	Methodology	55
3.5	Related Work	58
3.6	Experiments	60
3.6.1	Synthetic 1D Dataset	60
3.6.2	Regression on UCI Repository Datasets	61
3.6.3	Large-scale Regression on Airline Delays Dataset	62
3.7	Summary	63
	Addendum	69
3.A	Experimental Set-up and Implementation Details	69

3.A.1	Hardware	69
3.A.2	Software	69
3.A.3	Hyperparameters	69
3.B	Additional Results	70
3.B.1	Regression on Airline Delays Dataset	70
3.B.2	Extra UCI Repository Datasets	70
4	Cycle-Consistent Generative Adversarial Networks as a Bayesian Approximation	73
4.1	Introduction	73
4.2	Implicit Latent Variable Models	74
4.2.1	Prescribed Likelihood	75
4.2.2	Implicit Prior	76
4.3	Variational Inference	76
4.3.1	Prescribed Variational Posterior	76
4.3.2	Reverse KL Variational Objective	77
4.3.3	Approximate Divergence Minimisation	78
4.4	Symmetric Joint-Matching Variational Inference	79
4.4.1	Variational Joint	80
4.4.2	Forward KL Variational Objective	80
4.5	CycleGAN as a Special Case	82
4.5.1	Basic CycleGAN Framework	82
4.5.2	Cycle-consistency as Conditional Entropy Maximisation	83
4.5.3	Distribution Matching as Approximate Divergence Minimisation	84
4.6	Related Work	87
4.7	Experiments	88
4.8	Summary	90
	Addendum	93
4.A	Relation to KL Importance Estimation Procedure (KLIEP)	93
4.B	Summary of Definitions	94
5	Bayesian Optimisation by Classification with Deep Learning and Beyond	97
5.1	Introduction	97
5.2	Optimisation Policies and Density-Ratio Estimation	99
5.2.1	Relative Density-Ratio	99
5.2.2	Improvement-based Acquisition Functions	99
5.2.3	Tree-structured Parzen Estimator	102
5.2.4	Potential Pitfalls	102
5.3	Bayesian Optimisation by Probabilistic Classification	103
5.3.1	Choice of Proportion γ	106
5.3.2	Choice of Probabilistic Classifier	107
5.3.3	Likelihood-Free BO by Weighted Classification	109
5.4	Related Work	110
5.5	Experiments	111
5.5.1	Neural Network Tuning (HPOBench)	111

5.5.2	Neural Architecture Search (NASBench201)	114
5.5.3	Robot Arm Pushing	114
5.5.4	Racing Line Optimisation.	117
5.5.5	Ablation Studies	117
5.6	Discussion	118
5.7	Summary	121
	Addendum	123
5.A	Relative Density-Ratio: Unabridged Notation	123
5.B	Class-posterior Probability	124
5.C	Log Loss	124
5.C.1	Optimum	125
5.C.2	Empirical Risk Minimisation	125
5.D	Implementation of Baselines	126
5.E	Experimental Set-up and Implementation Details	126
5.E.1	BORE-RF	127
5.E.2	BORE-XGB	127
5.E.3	BORE-MLP	128
5.F	Details of Benchmarks	128
5.F.1	HPOBench	128
5.F.2	NASBench201	128
5.F.3	Robot pushing control	129
5.F.4	Racing Line Optimisation	130
5.G	Parameters, hyperparameters, and meta-hyperparameters	130
5.G.1	Parameters	131
5.G.2	Hyperparameters	132
5.G.3	Meta-hyperparameters	133
6	Conclusion	135
6.1	Summary of Contributions	135
6.2	Future Directions	136
6.3	Final Reflection	137
A	Numerical Methods for Improved Decoupled Sampling of Gaussian Processes	139
A.1	Introduction	139
A.2	Decoupled Sampling of Gaussian Processes	140
A.3	Numerical Integration for GP Prior Approximations	143
A.3.1	Monte Carlo Estimation	144
A.3.2	Quasi-Monte Carlo	145
A.3.3	Quadrature	146
A.3.4	Other Approaches	154
A.4	Experiments	156
A.4.1	Prior Approximation	156
A.4.2	Posterior Sample Approximation	160
A.5	Summary	162
	Addendum	165
A.A	Product-to-Sum Identity	165
A.B	Zero in Expectation	165

Bibliography	167
--------------	-----