

Contents

Summary	v
Contents	xxi
1 Introduction	1
1.1 Motivation	2
1.2 Overview of Contributions	3
1.3 Thesis Structure	3
2 Background	7
2.1 Bayes Decision Rule	7
2.1.1 Classification	7
2.1.2 Regression and Unconditional Modeling	10
2.2 Calibration and Refinement	11
2.2.1 Proper Scoring Rules	12
2.2.2 Decomposing Proper Scoring Rules	14
2.2.3 Illustrating PSR decomposition with an example	16
2.2.4 Calibration, Refinement and Bayes Decision Rule	19
2.2.5 Measuring Calibration	21
2.3 How does learning techniques meet model calibration and refinement	22
2.3.1 Maximum Likelihood Estimation	22
2.3.2 Empirical Risk Minimization	23
2.3.3 Overfitting	24
2.3.4 Regularization	25
2.3.5 Bayesian Learning	26

2.3.6	Model misspecification	29
3	Implicit Calibration of Deep Neural Networks using Mixup Training	31
3.1	From Empirical to Vicinal Risk minimization	31
3.1.1	Mixup Training	32
3.2	Does Mixup Training really achieves Model Calibration?	33
3.2.1	Data Augmentation and Model Calibration	34
3.2.2	Mixup and Model Calibration	35
3.3	The Auto Regularized Confidence Loss Function	37
3.3.1	Proposed Solution	39
3.3.2	Motivation behind the two loss variants	40
3.4	Experimental Evaluation	41
3.4.1	Experimental Details	41
3.4.2	Reported Results	43
3.4.3	Analysis of Results	44
3.4.4	A final insight on the experiments	45
3.5	Conclusions	46
4	Recalibration of Deep Probabilistic Models using Bayesian Neural Networks	49
4.1	Introduction to Post Calibration	50
4.1.1	Deep Neural Networks are Uncalibrated	51
4.1.2	The benefits of post-calibration techniques	52
4.2	Bayesian Neural Networks as Post-Calibration technique	52
4.2.1	Bayesian Modeling and Calibration	55
4.2.2	Proposed Solution	56
4.2.3	Chapter Summary	65
4.3	Experiments	66
4.3.1	Experiments set up	66
4.3.2	Bayesian vs Non-Bayesian Linear Regression	68
4.3.3	Selecting optimal K on validation	69
4.3.4	Calibration performance of BNN	70
4.3.5	Comparison Against state-of-the-art calibration techniques	73
4.3.6	Qualitative Analysis	76
4.4	Discussion	77
4.5	Conclusions and Future Work	77
5	Transformed Gaussian Process as a new prior over functions	79
5.1	Standard Gaussian Process	80
5.1.1	Introduction	80

5.1.2	Bayesian predictions using GP	82
5.1.3	Bayesian Model Selection	83
5.1.4	Benefits of Bayesian Learning Using Gaussian Processes .	84
5.1.5	Drawbacks of Bayesian Learning Using Gaussian Processes	84
5.2	Sparse Gaussian Process	85
5.3	Transformed Gaussian Processes	88
5.3.1	Model Description	89
5.3.2	Input-dependent Flows	90
5.3.3	Bayesian Priors on Flows	92
5.3.4	Induced Distributions	93
5.4	Inference in the Transformed Gaussian Process	94
5.4.1	Sparse Prior	94
5.4.2	Choice of the Variational Distribution	96
5.4.3	Evidence Lower Bound	96
5.4.4	Input Dependent Flows	99
5.4.5	Computational benefits of the approximate posterior . . .	101
5.5	Warped Gaussian Processes	101
5.6	Predictions	103
5.7	Experimental Evaluation	104
5.7.1	Bayesian Input Dependent TGP	104
5.7.2	Calibration Properties of the TGP	105
5.7.3	Computational Performance of the TGP	112
5.7.4	Uncertainty handled by the GP and TGP	113
5.7.5	Applications	114
5.8	Conclusions and Future work	116
6	Conclusions and Future Work	117
A	Additional Calibration Results in Chapter 3	119
A.1	Additional Loss Analysis	119
A.1.1	Accuracy Improvement	119
A.1.2	Further discussion about hyperparameter search	120
A.2	Additional Results	121
B	Experimental Details of the Transformed Gaussian Process	123
B.1	Flow Parameters Initialization	123
B.1.1	Initializing Flows from Data	124
B.1.2	Initializing Flows approximately to Identity	125
B.1.3	Initialization of Input-dependent flows	125
B.2	Experiment Details	125
B.2.1	Regression and Classification	125

B.2.2 Real World Experiments	127
Bibliography	129