Table of contents

Nomenclature **xii**

1 Introduction: The Importance of Knowing What We Don't Know **1**
1.1 Deep learning . . . 2
1.2 Model uncertainty . . . 7
1.3 Model uncertainty and AI safety . . . 9
1.3.1 Physician diagnosing a patient . . . 9
1.3.2 Autonomous vehicles . . . 9
1.3.3 Critical systems and high frequency trading . . . 10
1.4 Applications of model uncertainty . . . 11
1.4.1 Active learning . . . 11
1.4.2 Efficient exploration in deep reinforcement learning . . . 12
1.5 Model uncertainty in deep learning . . . 13
1.6 Thesis structure . . . 16

2 The Language of Uncertainty **17**
2.1 Bayesian modelling . . . 17
2.1.1 Variational inference . . . 19
2.2 Bayesian neural networks . . . 20
2.2.1 Brief history . . . 20
2.2.2 Modern approximate inference . . . 23
2.2.3 Challenges . . . 27

3 Bayesian Deep Learning **29**
3.1 Advanced techniques in variational inference . . . 30
3.1.1 Monte Carlo estimators in variational inference . . . 30
3.1.2 Variance analysis of Monte Carlo estimators in variational inference 34
3.2 Practical inference in Bayesian neural networks . . . 37

3.2.1 Stochastic regularisation techniques . . . 39
3.2.2 Stochastic regularisation techniques as approximate inference . . . 41
3.2.3 KL condition . . . 44
3.3 Model uncertainty in Bayesian neural networks . . . 47
3.3.1 Uncertainty in classification . . . 51
3.3.2 Difficulties with the approach . . . 54
3.4 Approximate inference in complex models . . . 56
3.4.1 Bayesian convolutional neural networks . . . 56
3.4.2 Bayesian recurrent neural networks . . . 58

4 Uncertainty Quality **62**
4.1 Effects of model structure on uncertainty . . . 62
4.2 Effects of approximate posterior on uncertainty . . . 63
4.2.1 Regression . . . 66
4.2.2 Classification . . . 71
4.3 Quantitative comparison . . . 73
4.4 Bayesian convolutional neural networks . . . 78
4.4.1 Model over-fitting . . . 80
4.4.2 MC dropout in standard convolutional neural networks . . . 80
4.4.3 MC estimate convergence . . . 82
4.5 Recurrent neural networks . . . 83
4.6 Heteroscedastic uncertainty . . . 85

5 Applications **88**
5.1 Recent literature . . . 88
5.1.1 Language applications . . . 89
5.1.2 Medical diagnostics and bioinformatics . . . 89
5.1.3 Computer vision and autonomous driving . . . 90
5.2 Active learning with image data . . . 91
5.3 Exploration in deep reinforcement learning . . . 95
5.4 Data efficiency in deep reinforcement learning . . . 97
5.4.1 PILCO . . . 98
5.4.2 Deep PILCO . . . 100
5.4.3 Experiment . . . 102

6 Deep Insights **105**
6.1 Practical considerations for getting good uncertainty estimates . . . 105

6.2 What determines what our uncertainty looks like? . . . 106
6.3 Analytical analysis in Bayesian linear regression . . . 107
6.4 ELBO correlation with test log likelihood . . . 111
6.5 Discrete prior models . . . 118
6.6 Dropout as a proxy posterior in spike and slab prior models . . . 121
6.6.1 Historical context . . . 121
6.6.2 Spike and slab prior models . . . 122
6.6.3 Related work . . . 123
6.6.4 Approximate inference with free-form variational distributions . . . 123
6.6.5 Proxy optimal approximating distribution . . . 124
6.6.6 Spike and slab and dropout . . . 126
6.7 Epistemic, Aleatoric, and Predictive uncertainties . . . 127

7 Future Research **133**

References **137**

Appendix A KL condition **149**

Appendix B Figures **153**

Appendix C Spike and slab prior KL **159**