Bayesian Neural Network

By Marc Pedersen, Daniel Dalsgaard Jonsson and Gustav Stolberg-Larsen

University of Copenhagen - The Niels Bohr Institute - Advanced Methods in Applied Statistics 2022

Outline

Introduction

Review

Discussion

Conclusion

References

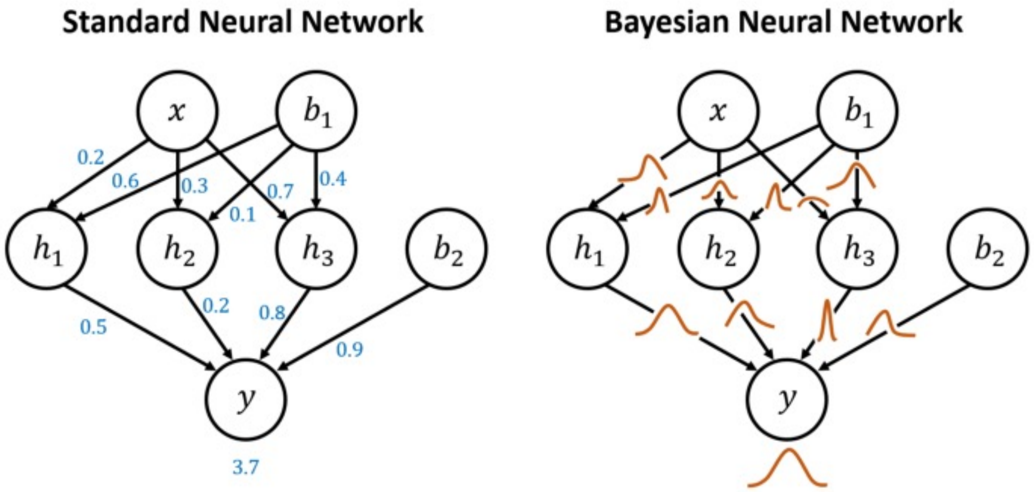


Figure: https://towardsdatascience.com/why-you-should-use-bayesian-neural-network-aaf76732c150

Introduction

Introduction

Review

Discussion

Conclusion

References

Traditional Artificial neural networks (ANN) tend to overfit. $\boldsymbol{l}_i = s_i(\boldsymbol{W}_i \boldsymbol{l}_{i-1} + \boldsymbol{b}_i)$

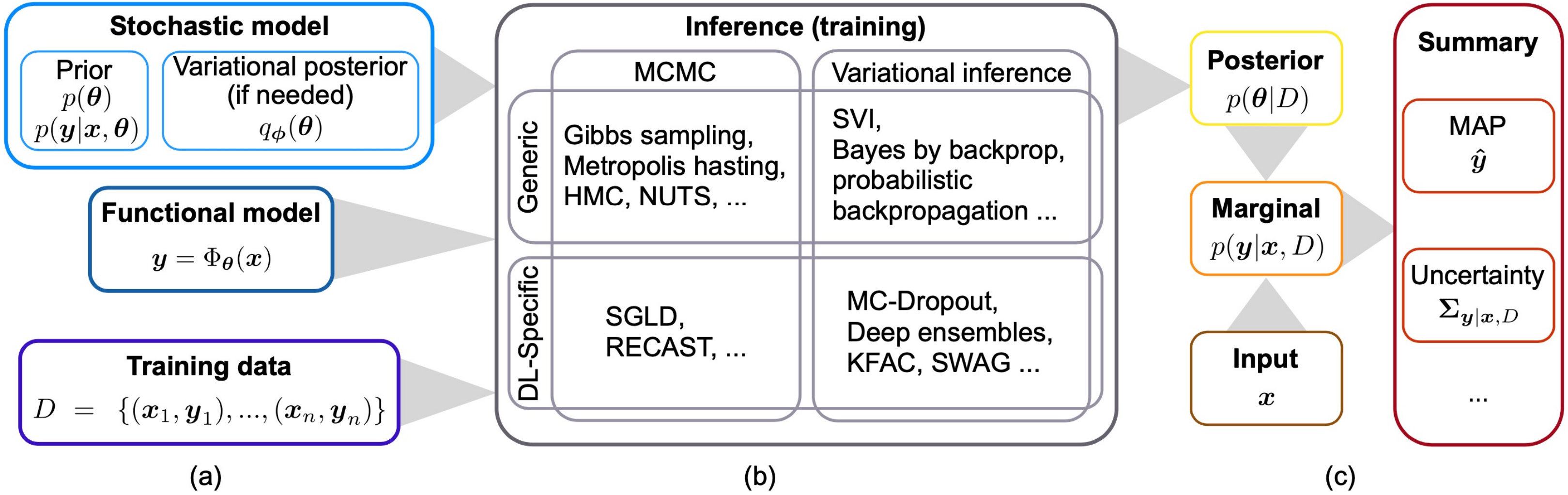
Bayesian neural networks (BNN) are NNs in a Bayesian framework

BNNs can better estimate confidence levels

$$p(\theta|D) = \frac{p\left(D_y \middle| D_x, \theta\right) p(\theta)}{\int p\left(D_y \middle| D_x, \theta'\right) p(\theta)\, d\theta'} \propto p\left(D_y \middle| D_x, \theta\right) p(\theta)$$

BNN

Introduction
Review
Discussion
Conclusion
References



MCMC and Variational Inference

Introduction
Review
Discussion
Conclusion
References

	Benefits	Limitations	Use cases	
MCMC (V.A)	Directly samples the posterior	Requires to store a very large number of samples	Small and average models	
Classic methods (HMC, NUTS)(§V-A)	State of the art samplers limit autocorrelation between samples	Do not scale well to large models	Small and critical models	**Can be combined**
SGLD and derivates (§V-E2a)	Provide a well behaved Markov Chain with minibatches	Focus on a single mode of the posterior	Models with larger datasets	
Warm restarts (§V-E2a)	Help a MCMC method explore different modes of the posterior	Requires a new burn-in sequence for each restart	Combined with a MCMC sampler	
Variational inference (V.B)	The variational distribution is easy to sample	Is an approximation	Large scale models	
Bayes by backprop (§V-C)	Fit any parametric distribution as posterior	Noisy gradient descent	Large scale models	**Can be combined**
Monte Carlo-Dropout (§V-E1)	Can transform a model using dropout into a BNN	Lack expressive power	Dropout based models	
Laplace approximation (§V-E2b)	By analyzing standard SGD get a BNN from a MAP	Focus on a single mode of the posterior	Unimodals large scale models	
Deep ensembles (§V-E2b)	Help focusing on different modes of the posterior	Cannot detect local uncertainty if used alone	Multimodals models and combined with other VI methods	

Bayes-by-backprop

Introduction

Review

Discussion

Conclusion

References

Draw epsilon from q

Deterministic transformation

Loss function with variational inference

Normal backpropagation

Algorithm 5 Bayes-by-backprop algorithm.

$\phi = \phi_0$;
for $i = 0$ **to** N **do**
  Draw $\varepsilon \sim q(\varepsilon)$;
  $\boldsymbol{\theta} = t(\varepsilon, \boldsymbol{\phi})$;
  $f(\boldsymbol{\theta}, \boldsymbol{\phi}) = \log(q_\phi(\boldsymbol{\theta})) - \log(p(D_{\boldsymbol{y}}|D_{\boldsymbol{x}}, \boldsymbol{\theta})p(\boldsymbol{\theta}))$;
  $\Delta_\phi f = \text{backprop}_\phi(f)$;
  $\phi = \phi - \alpha\Delta_\phi f$;
end for

Performance Metrics

Introduction

Review

Discussion

Conclusion

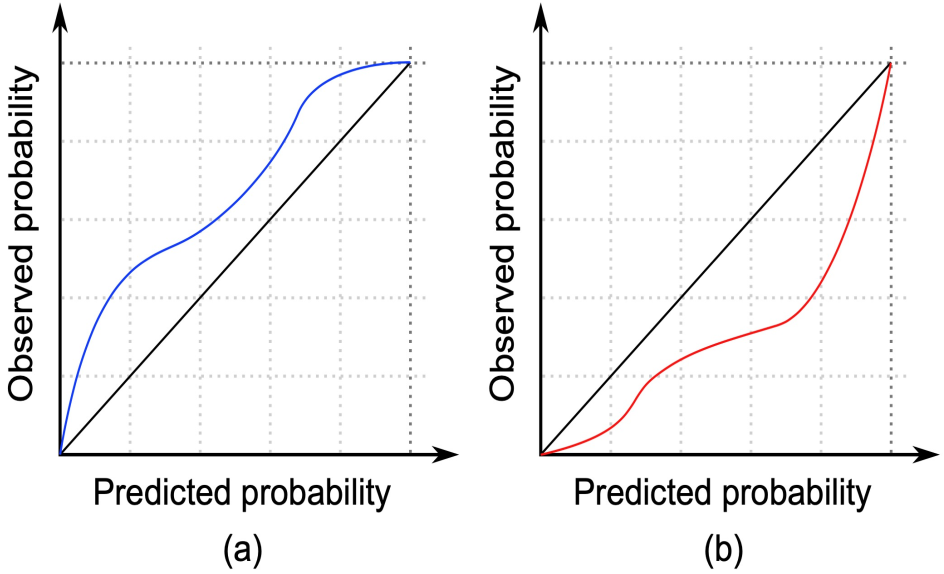
References

Distribution, not point
Estimate

Calibration graph:

- predicted probability p
- observed probability q

If q > p – underconfident
If q < p – overconfident
If q ≈ p – well calibrated



Discussion

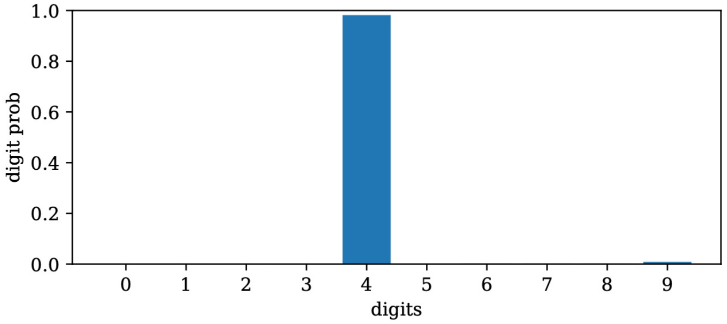
Introduction

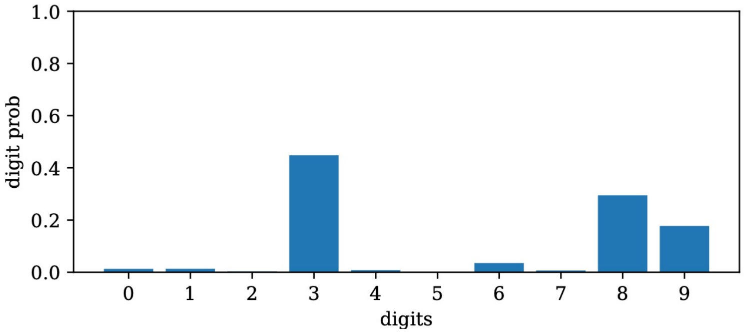
Review

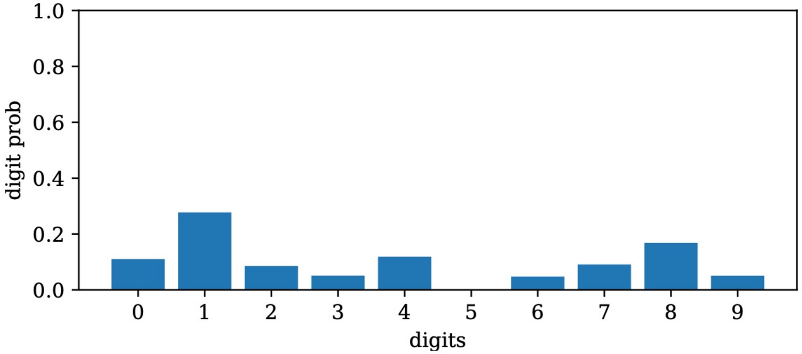
Discussion

Conclusion

References







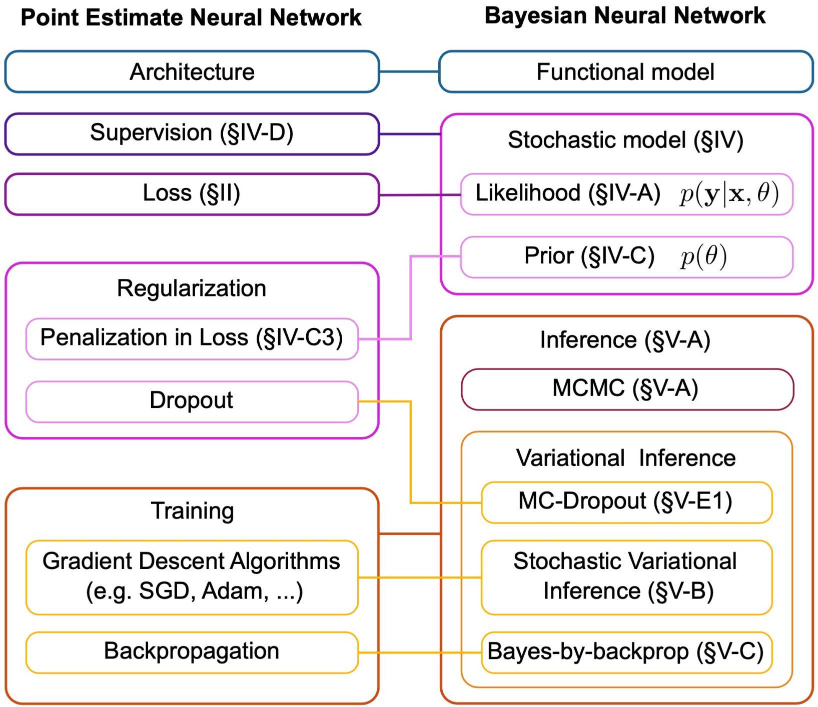
Conclusion

Introduction
Review
Discussion
Conclusion
References

Appropriate confidence intervals

Always use the right tool for the task

- BNN
- ANN



References

Introduction

Review

Discussion

Conclusion

References

Laurent Valentin Jospin, Wray L. Buntine, Farid Boussaïd, Hamid Laga, Mohammed Bennamoun: "*Hands-on Bayesian Neural Networks – A Tutorial for Deep Learning Users*".
Submitted on 14 Jul 2020 (v1). Last revised 3 Jan 2022 (v3).
https://arxiv.org/abs/2007.06823

Figure slide 2:
https://towardsdatascience.com/why-you-should-use-bayesian-neural-network-aaf76732c150

Acknowledgements

Introduction

Review

Discussion

Conclusion

References

Extras

- D. Jason Koskinen for feedback on the relevance of the topic.