



**Hochschule
Bonn-Rhein-Sieg**
University of Applied Sciences



Master's Thesis

Project title

Your Name

Submitted to Hochschule Bonn-Rhein-Sieg,
Department of Computer Science
in partial fulfillment of the requirements for the degree
of Master of Science in Autonomous Systems

Supervised by

First supervisor

Second Supervisor

Third Supervisor

Month 20XX

I, the undersigned below, declare that this work has not previously been submitted to this or any other university and that it is, unless otherwise stated, entirely my own work.

Date

Your Name

Abstract

Your abstract

Acknowledgements

Thanks to

Contents

1	Introduction	1
1.1	Motivation	2
1.1.1	2
1.1.2	3
1.2	Challenges and Difficulties	3
1.2.1	3
1.2.2	4
1.2.3	4
1.3	Problem Statement	4
1.3.1	4
1.3.2	6
1.3.3	6
2	State of the Art	7
2.1	Structure of the chapter	7
2.2	A General Framework for Uncertainty Estimation in Deep Learning	7
2.2.1	Overview	7
2.2.2	Integrating MCDO_ADF with a Neural Network and estimating uncertainties	8
2.2.3	Inference procedure	12
2.2.4	Downsides	13
2.3	Deep Evidential Regression	14
2.3.1	Overview	14
2.3.2	Conjugate priors	14
2.3.3	Evidential distribution	14
2.3.4	Evidential learning objectives	16
2.3.5	Estimating uncertainty	18
2.4	Limitations of previous work	19
3	Methodology	21
4	Solution	23
4.1	Proposed algorithm	23
4.2	Implementation details	23
5	Evaluation	25

6	Results	27
6.1	Use case 1	27
6.2	Use case 2	27
6.3	Use case 3	27
7	Conclusions	29
7.1	Contributions	29
7.2	Lessons learned	29
7.3	Future work	29
	Appendix A Design Details	31
	Appendix B Parameters	33
	References	35

List of Figures

2.1	Illustration of the MCDO_ADF technique. Here x_i denotes the input through the i^{th} unit of a given hidden layer, $x_{i(n)}$ denotes the noise variance input to the i th unit of the n th layer. Circles with crosses inscribed denote the dropped out neurons whereas the ones with the Gaussian distribution symbol denote the active units. T values of μ and v are collected from T stochastic forward passes. Image source:	8
2.2	Illustration of forward passes in deterministic and ADF versions of a Neural Network. Here x denotes the input activations, $z^{(n)}$ denotes activation input to the n th layer, y denotes the output, θ_{z^n} represents input activation to the n th layer expressed as a probability distribution and σ corresponds to the noise variance. Image source:	10
2.3	Realizations of the NIG distribution. Image source:	16

List of Tables

1

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit

ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

1.1 Motivation

1.1.1 ...

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce

sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.

1.1.2 ...

1.2 Challenges and Difficulties

1.2.1 ...

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est.

Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

1.2.2 ...

1.2.3 ...

1.3 Problem Statement

1.3.1 ...

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, malesuada id, gravida nec, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, porta sed, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed vitae sem.

Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Maecenas ante. Duis ullamcorper enim. Donec tristique enim eu leo. Nullam molestie elit eu dolor. Nullam bibendum, turpis vitae tristique gravida, quam sapien tempor lectus, quis pretium tellus purus ac quam. Nulla facilisi.

Duis aliquet dui in est. Donec eget est. Nunc lectus odio, varius at, fermentum in, accumsan non, enim. Aliquam erat volutpat. Proin sit amet nulla ut eros consectetur cursus. Phasellus dapibus aliquam justo. Nunc laoreet. Donec consequat placerat magna. Duis pretium tincidunt justo. Sed sollicitudin vestibulum quam. Nam quis ligula. Vivamus at metus. Etiam imperdiet imperdiet pede. Aenean turpis. Fusce augue velit, scelerisque sollicitudin, dictum vitae, tempor et, pede. Donec wisi sapien, feugiat in, fermentum ut, sollicitudin adipiscing, metus.

Donec vel nibh ut felis consectetur laoreet. Donec pede. Sed id quam id wisi laoreet suscipit. Nulla lectus dolor, aliquam ac, fringilla eget, mollis ut, orci. In pellentesque justo in ligula. Maecenas turpis. Donec eleifend leo at felis tincidunt consequat. Aenean turpis metus, malesuada sed, condimentum sit amet, auctor a, wisi. Pellentesque sapien elit, bibendum ac, posuere et, congue eu, felis. Vestibulum mattis libero quis metus scelerisque ultrices. Sed purus.

Donec molestie, magna ut luctus ultrices, tellus arcu nonummy velit, sit amet pulvinar elit justo et mauris. In pede. Maecenas euismod elit eu erat. Aliquam augue wisi, facilisis congue, suscipit in, adipiscing et, ante. In justo. Cras lobortis neque ac ipsum. Nunc fermentum massa at ante. Donec orci tortor, egestas sit amet, ultrices eget, venenatis eget, mi. Maecenas vehicula leo semper est. Mauris vel metus. Aliquam erat volutpat. In rhoncus sapien ac tellus. Pellentesque ligula.

Cras dapibus, augue quis scelerisque ultricies, felis dolor placerat sem, id porta velit odio eu elit. Aenean interdum nibh sed wisi. Praesent sollicitudin vulputate dui. Praesent iaculis viverra augue. Quisque in libero. Aenean gravida lorem vitae sem ullamcorper cursus. Nunc adipiscing rutrum ante. Nunc ipsum massa, faucibus sit amet, viverra vel, elementum semper, orci. Cras eros sem, vulputate et, tincidunt id, ultrices eget, magna. Nulla varius ornare odio. Donec accumsan mauris sit amet augue. Sed ligula lacus, laoreet non, aliquam sit amet, iaculis tempor, lorem. Suspendisse eros. Nam porta, leo sed congue tempor, felis est ultrices eros, id mattis velit felis non metus. Curabitur vitae elit non mauris varius pretium. Aenean lacus sem, tincidunt ut, consequat quis, porta vitae, turpis. Nullam laoreet fermentum urna. Proin iaculis lectus.

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

1.3.2 ...

1.3.3 ...

2

State of the Art

2.1 Structure of the chapter

This chapter aims to explain the two state-of-the-art uncertainty estimation methods compared in this research work. The sections 2.2 and 2.3 give an intuitive explanation of the methods.

2.2 A General Framework for Uncertainty Estimation in Deep Learning

2.2.1 Overview

This work proposes a technique to distinctively estimate data and model uncertainties associated with an output of any Neural Network model. The technique is here after referred to as "MCDO_ADF", representing the fact that is a combination of two ideas, Monte-Carlo Drop Out (MCDO) and Assumed-Density-Filtering (ADF). MCDO_ADF treat the two uncertainty components to be related, which sets it apart from most of other uncertainty estimation methods that treat them to be independent. The method employs Bayesian Belief Networks combined with Monte-Carlo sampling for estimating the model uncertainty and relies on the idea of Assumed Density Filtering for estimating data uncertainty associated with an output.

Authors of the MCDO_ADF technique claim it to be a general framework to estimate uncertainties in Neural Networks. They give following reasons to validate their claim:

- Using this uncertainty estimation method does not require any architectural changes in the target Neural Network.
- Applicability of the method to Neural Network models of different tasks.
- Absence of any need of make changes in the optimization process.
- Ability of the technique to be applied to already trained models.

The upcoming sections of this chapter explain the MCDO_ADF technique and also analyze its claimed "generality" by using it in a Resnet8 based Neural Network regression model meant for the application of steering-angle prediction in autonomous cars. (Note: A detailed description of the data set, training and inference procedures of the Neural Network model is available in the next chapter[]).

2.2.2 Integrating MCDO_ADF with a Neural Network and estimating uncertainties

MCDO_ADF as an algorithm

Estimating uncertainty using the MCDO can be formulated as an algorithm consisting of the following steps:

- Transform the Neural Network of interest to its ADF(Assumed Density Filtering) version.
- Collect a predefined number(T) of Monte-Carlo(MC) samples by forwarding inputs and noise variances (x, v) stochastically through the network for T times.
- Computation of output predictions and uncertainties

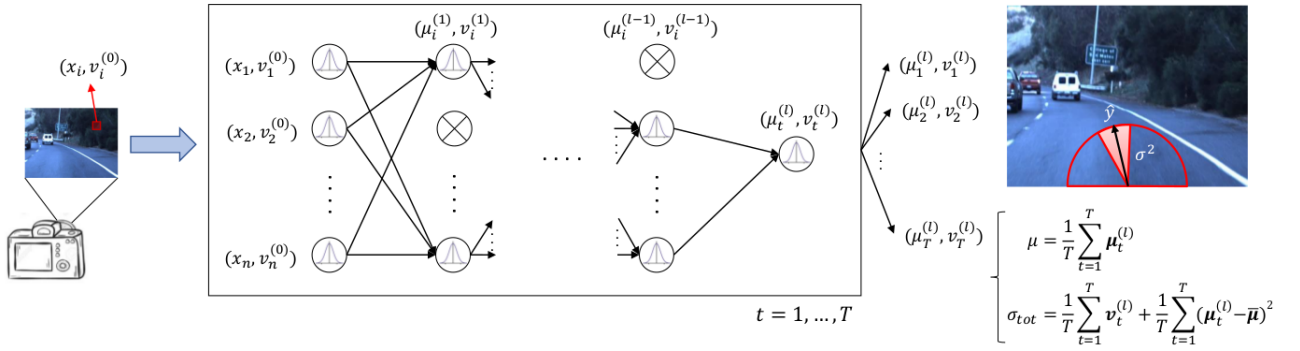


Figure 2.1: Illustration of the MCDO_ADF technique. Here x_i denotes the input through the i^{th} unit of a given hidden layer, $x_{i(n)}$ denotes the noise variance input to the i th unit of the n th layer. Circles with crosses inscribed denote the dropped out neurons whereas the ones with the Gaussian distribution symbol denote the active units. T values of μ and v are collected from T stochastic forward passes. Image source:

Assumed Density Filtering (ADF)

The MCDO_ADF technique considers sensor noise to be the primary source of data uncertainty in neural network predictions and therefore feeds it to the Neural Network model during the inference. In order to propagate the input data distribution (parameterized by the input as its mean and sensor noise as variance) the technique of ADF is used. Briefly in the context of MCDO_ADF, ADF replaces every input activation into a probability distribution and also approximates the same using a tractable Gaussian distribution and makes both the mean and noise variance available in the output layer. Following points describe Assume Density Filtering in a more detailed manner:

- Assumed Density Filtering(ADF) is a technique in Bayesian machine learning to approximate intractable and complex distribution with distributions that are easy to handle. In the case of Bayesian Inference, ADF aims to project the true posterior onto a distribution of choice. The exponential family of distributions are a popular choice.
- In the case of MCDO_ADF there is a need to propagate the input data distribution so that the values of its mean and variance (noise variance) are available in the output layer.
- The input data distribution is considered to be Gaussian in nature. Every intermediate layer outputs the transformed version of the input distribution. However, when it propagates through non-linearities in a Neural Network the resulting distribution need not be essentially another Gaussian. Such a distribution emerging out of non-linear blocks is also conditioned by distribution over activations of the preceeding layers. Therefore, the resulting distribution becomes intractable.
- Such intractable and complex distributions are estimated using ADF by:
 - Assuming conditional independence between distribution outputted from a given layer with its preceeding layers.
 - Approximating the complex distribution with a Gaussian distribution whose pair has the least possible value of Kullback-Leibler divergence(\mathbb{KL}). ADF achieves this by matching the first two moments of the distributions.
 - In practice, this is achieved by optimizing the global variational objective(\mathbb{J}).

In practice, every building block of a Neural Network has its corresponding ADF version and therefore during the inference time the entire model has to be transformed to its ADF equivalent. This gives the ability to the Neural Network model to propagate and output distributions which represent data uncertainty.

Data uncertainty estimation

The ADF transformed Neural Network produces two outputs from the final layer: mean ($\mu_{t(l)}$) and variance($v_{t(l)}$) of the propagated distribution, as shown in the figure 2.1. The pair of values is outputted for each of the T stochastic forward passes (described in the next paragraph) and the mean of T variance values is considered to be the value of data uncertainty. Likewise, the mean of T predictions is considered to be the model's prediction for the given input.

$$prediction = \mu = \frac{1}{T} \sum_{t=1}^T \mu_{t(l)} \quad (2.1)$$

$$data_uncertainty = \sigma_{data} = \frac{1}{T} \sum_{t=1}^T v_{t(l)} \quad (2.2)$$

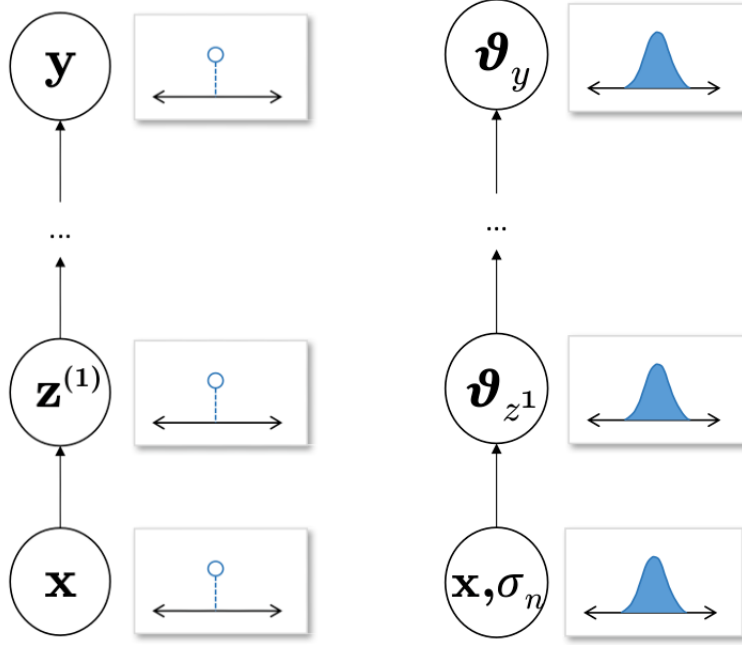


Figure 2.2: Illustration of forward passes in deterministic and ADF versions of a Neural Network. Here x denotes the input activations, $z^{(n)}$ denotes activation input to the n th layer, y denotes the output, θ_{z^n} represents input activation to the n th layer expressed as a probability distribution and σ corresponds to the noise variance. Image source:

Monte-Carlo Dropout (MCDO)

This uncertainty estimation method relies on the idea of Monte-Carlo (MC) sampling to estimate model uncertainty associated any prediction. In practice, MC sampling is achieved by enabling dropout during the test time and obtaining the desired number of samples (T), which are nothing but outputs of the Neural Network model during different forward passes of the input. Enabling dropout introduces stochasticity during those forward passes.

Following points briefly describe the dropout technique in a general context:

- Dropout(\square) is primarily a regularization technique used while training Neural Networks in order to avoid over-fitting.
- During dropout certain nodes of a given Neural Network layer are not considered for training. The nodes are ignored with a probability equal to the dropout rate (often denoted by \mathbf{p}).
- Using dropout during training makes Neural Network layers to adapt in such a way that they cope with mistakes made by the prior layers.

In the context of Bayesian inference, the Dropout technique is used to approximate the posterior distribution over weights of a given Neural Network, when the training data and labels are given

$(P(W|X, Y))$. The approximation is obtained by applying dropout at the test-time. This makes it possible to obtain multiple predictions for any given input from different architectures resulting from application of dropout to the base Neural Network model. The different architectures obtained along with their weights can be considered as Monte-Carlo(MC) samples from the space of all possible architectures. The number of MC samples to be obtained is a hyper-parameter and denoted by T . In another perspective, T equals the number of forward passes through different architectures with different sets of weights $\{W_1^t, \dots, W_L^t\}_{t=1}^T$ (L denotes the number of dropout applied layers in the Neural Network). The first and second moments (mean and variance) of predictions obtained from these stochastic forward passes of given input are utilized to compute model uncertainty(explained in ??). One of the highlights of this technique is that its usage does not require any architectural changes and also can applied to already trained Neural Net models. The hyper-parameters T and p significantly impact the effectiveness of this technique. In the case of p a very high value (close to 1) increases sparsity in nodes and also results in longer convergence-time while a low value eliminates the MC-sampling utility. For our experiment, the value of $p = 0.02$ is used. The hyper-parameter T significantly impacts the inference time of a Neural Network model and therefore has to be chosen optimally based on the run-time requirement of the system where the model would be deployed.

Model uncertainty estimation

The MCDO_ADF technique estimates model uncertainty using predictions generated from the Neural Network model during multiple(T) forward passes, while the dropout is enabled. A given input is processed by the model T times, with a new combination of neurons considered for almost every forward pass. This produces an effect of gathering predictions from an ensemble consisting of T Neural Net models with different architectures. The variance of T gathered predictions is the estimated model uncertainty. In the following equations $\mu_{t(i)}$ signifies the mean output from the ADF transformed version of the Model during t^{th} forward pass.

$$model_uncertainty = \sigma_{model} = \frac{1}{T} \sum_{t=1}^T (\mu_{t(i)} - \bar{\mu})^2 \quad (2.3)$$

$$where, \bar{\mu} = \frac{1}{T} \sum_{t=1}^T \mu_{t(i)} \quad (2.4)$$

Combining ADF and MCDO

The MCDO_ADF method considers a relationship to exist between the two components of uncertainty (data and model). The relationship is realized in this technique by combining both the ideas of ADF and MCDO. During inference,

- The given Neural Network model is transformed to its ADF equivalent so that the output layer produces both predictions(mean) and noise variance as the model's final outputs.

- For estimating model uncertainty, dropout is enabled in the ADF transformed version of the original Neural Network following which T MC samples are collected during T stochastic forward passes. It is this application of dropout on the ADF transformed version that produces the "effect of ensembling T ADF Neural Networks" and also considers any relationship between the two uncertainty components.
- Combining ADF and MCDO leads to another intuitive realization about the uncertainty components in this setup. Even when a particular input fed to the Neural Net model was observed frequently during training, if corrupted due to sensor noise then it will not only have high values of both data and model uncertainties.

Total uncertainty

The predictive uncertainty is estimated by summing up both its components (data and model) and is given by the following equation.

$$predictive_uncertainty = \sigma_{total} = \frac{1}{T} \sum_{t=1}^T ((\mu_{t(l)} - \bar{\mu})^2 + v_{t(l)}) \quad (2.5)$$

In summary, both ADF and MCDO techniques approximate probability distributions of data and model with Normal distributions respectively. ADF propagates the input data distribution and approximates them as Gaussians in every Neural Network layer while MCDO approximates the distribution around weights by sampling and forms a Gaussian distribution out of the samples. The variances of these Gaussian distributions are considered to be the uncertainty components and are summed up to yield the predictive uncertainty.

2.2.3 Inference procedure

The MCDO_ADF method can be applied to already trained deterministic version of the Neural Network models as mentioned in 2.2.1. However, it is also possible to train the Neural Network model of interest with dropout enabled and use the same for inference. In order to estimate the predictive uncertainty during inference:

- Every layer of the Neural Network has to be replaced with its ADF equivalent so that they are equipped with the ability to propagate data distributions. The implementation of ADF equivalents for most of the Neural Network building blocks is available in [].
- The value of noise variance (a constant value) obtained from the sensor's data sheet is fed along with the input data to the ADF transformed input layer of the network. During propagation through intermediate layers, it is ensured that at least a minimum value of variance is propagated. In the case of experiment described in the next chapter, a minimum value of 0.001 is used.
- Every input along with the noise variance undergoes T stochastic forward passes through the network to generate T predictions.

The experiment described in the next chapter discusses more on practical aspects of this technique.

2.2.4 Downsides

- The need for multiple (T) forward passes to obtain MC samples is computationally expensive and cannot be afforded in the case of real-time systems. While there is an option to reduce the value of T , it increases the difference between approximated and underlying distribution over weights of the model, thereby affecting the method's performance.
- The method considers sensor noise to be the only source of data uncertainty. Also, it treats the noise to be additive Gaussian in nature. However, sensor noise is just one of the factors contributing to data uncertainty. For instance, in the case of image data, usage of a lossy compression technique can also contribute to its noise. Also, it is possible for a given sensor to produce data whose noise levels differ. As it is impossible to consider and model every possible noise source, it is important for an uncertainty estimation method to learn to differentiate noise and useful information from given data.
- The authors of MCDO_ADF quote its ability to be applied to already trained models as one of the key reasons for its generality. However, retraining a Neural Network model is something feasible in most of the cases.
- As hyper parameters such as drop-out rate(p), number of MC samples(T) and noise variance have a major role to play in this technique, it adds to the responsibilities of the practitioner to optimally choose them based on the problem at hand.

****Downsides of MCDO and ADF to be read and listed****

2.3 Deep Evidential Regression

2.3.1 Overview

Deep Evidential Regression proposes a method (hereafter referred to as "DER") to estimate predictive uncertainty primarily in Neural Networks for regression, by simultaneously learning a hierarchy of distributions. The learned hierarchy consists of two levels of distributions: 1. A lower level Gaussian distribution over data, with parameters (mean μ and variance σ^2) 2. A higher level(also called Evidential) Normal-Inverse Gamma distribution over the parameters of the lower level distribution. In the perspective of the Bayesian Inference, the higher-order distribution can be taken as a prior over the the lower-order distribution which is obtained by evaluating likelihood of known data points for a particular choice of μ and σ^2 . The evidential distribution evaluated at any particular instance(a combination of μ and σ^2), provides the subjective belief mass of the corresponding lower-order distribution there. This subjective belief mass is also called as "evidence". Lack of evidence means existence of uncertainty and therefore the value of evidence is used to quantify predictive uncertainty.

In order to put the above mentioned ideas into practice, DER provides a loss function whose objectives are to:

- Fit the training data to the evidential model.
- Learn the evidential prior which would provide uncertainty estimates during inference.
- In simple words, to learn the parameters of the higher-order evidential distribution.

The upcoming sections of this chapter explain the method in a detailed manner.

2.3.2 Conjugate priors

2.3.3 Evidential distribution

From the perspective of Bayesian inference

Let us consider a regression problem with an available dataset D with N pairs of data labels and targets represented by $(x_1, y_1), \dots, (x_N, y_N)$. The DER method assumes that the targets are drawn independent and identically from a Gaussian distribution with unknown mean and variance represented by μ and σ^2 respectively.

$$(y_1, \dots, y_N) \sim \mathcal{N}(\mu, \sigma^2) \quad (2.6)$$

The parameters μ and σ^2 are considered to be random variables that follow Gaussian and Inverse-Gamma distributions respectively.

$$\mu \sim \mathcal{N}(\gamma, \sigma^2 \lambda^{-1}) \quad (2.7)$$

$$\sigma^2 \sim \Gamma^{-1}(\alpha, \beta) \quad (2.8)$$

where $\alpha, \beta, \gamma, \lambda$ denote parameters of the higher-order Normal Inverse Gamma (NIG) distribution. Let $\theta = (\mu, \sigma^2)$ denote the parameters of one instance of Gaussian distribution generating targets y_i and $m = (\alpha, \beta, \gamma, \lambda)$ denote the set of NIG distribution parameters.

We are interested to model the distribution around θ . Applying Bayes Rule, we get

$$P(\theta|m) = \frac{P(m|\theta)P(\theta)}{P(m)}, \quad (2.9)$$

posterior. dist. over θ for the given value of $m = \frac{\text{likelihood of } m \text{ evaluated at the given value of } \theta \times \text{prior over } \theta}{\text{likelihood of } m \text{ evaluated at all possible values of } \theta}$

Here, the prior over θ is a NIG distribution and the likelihood function is Gaussian in nature. Therefore, the posterior takes the form of an NIG distribution expressed as follows:

$$P(\mu, \sigma^2 | \gamma, \lambda, \alpha, \beta) = \frac{\sqrt{\lambda}}{\sigma\sqrt{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{2\beta + \lambda(\gamma - \mu)^2}{2\sigma^2}\right) \quad (2.10)$$

Significance of NIG parameters

γ and α are shape and location parameters of NIG distribution respectively. β refers to the inverse-scale (rate) parameter. This means that spread of the distribution is inversely related to β . There is a relationship that exists between parameters of the NIG distribution. γ can be interpreted as the sample mean of λ virtual observations, determining NIG's location. On the other hand, spread of the NIG distribution can be considered to have calculated from 2α virtual observations whose sample mean equals γ and their squared deviations summing to 2β . DER considers the count of virtual observations as evidence(ϕ) in support of the data sample at hand.

$$\phi = \lambda + 2\alpha \quad (2.11)$$

The distribution hierarchy

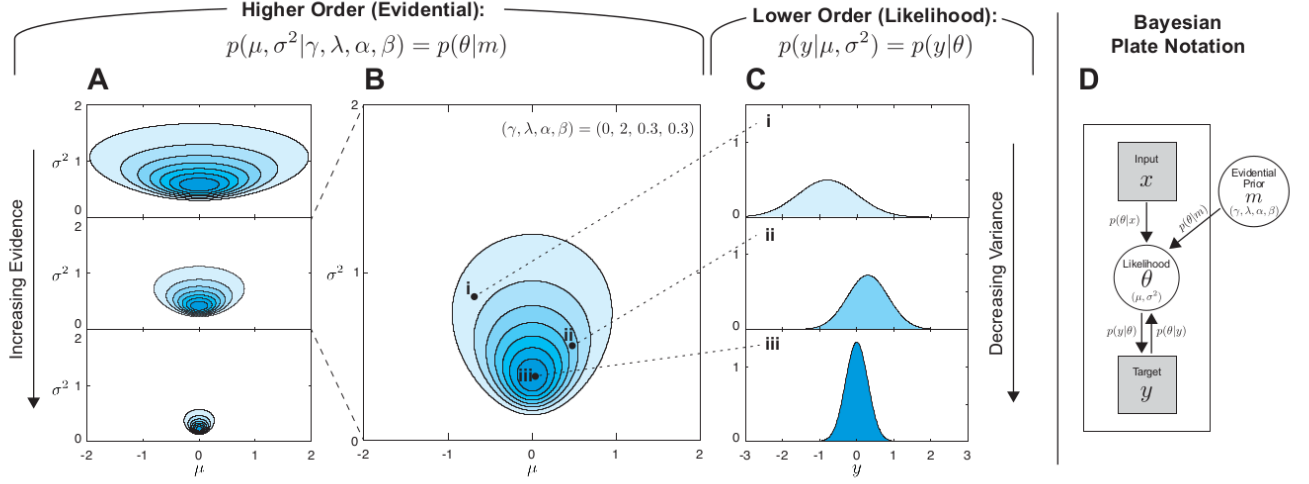


Figure 2.3: Realizations of the NIG distribution. Image source:

Figure 2.3 illustrates the impact of increase in evidence on the shape and spread of the NIG distribution (column **A**) and various realizations of the lower-order likelihood distribution (column **C**) from a given instance of NIG distribution (column **B**). Following are some of the key insights that can be obtained from the illustration:

- With increase in evidence (as expressed in 2.11) the belief mass increasingly concentrates around a specific value pair of μ and σ^2 in the column **A** meaning a reduction in uncertainty.
- Column **B** illustrates the evidential distribution centered around a particular value pair of μ and σ^2 . This intuitively means that every point on the distribution corresponds to parameters of a possible likelihood distribution.
- Sampling the higher order distribution at various locations yield likelihood distributions of varying levels of evidence associated with them. Column **C** in the illustration shows such realizations. Darker the blue shade used to represent a likelihood-distribution, higher the evidence level associated with it.

2.3.4 Evidential learning objectives

As described in the overview (2.3.1), the objectives of DER are two fold: 1. Maximize the model fit and 2. Minimize the evidence measure in an event of error. The method realizes these objectives in form of a loss function/s(two forms) which is integrated to the Neural Network of choice and optimized during training.

Maximizing the model fit

This objective of DER focuses on learning the underlying patterns in the data and also increasing the belief mass/evidence in favor of right predictions.

Rewriting the equation 2.9

$$P(\theta|m) = \frac{P(m|\theta)P(\theta)}{P(m)}$$

Let us assume that we observe our target y_i which when added to the above equation yields,

$$P(\theta|y_i, m) = \frac{P(y_i|\theta, m)P(\theta|m)}{P(y_i|m)} \quad (2.12)$$

Before interpreting the above equation it is important to recollect the fact that in Bayesian Inference every Random Variable(RV) involved is considered to be jointly distributed. In the case of above equation, Random Variables Y, Θ and M corresponding to y_i, θ and m are jointly distributed.

In the above equation, we determine the probability distribution around θ when it is conditioned under specific values of y_i and m . The two terms in the numerator denote likelihood and prior as described in 2.3.3. The denominator term is termed as "marginal likelihood" or "evidence" in Bayesian inference, as it yields the total probability mass of $Y = y_i$ for all possible realizations of θ in the model parameterized by m . The evidence term is also important to normalize the likelihood so that it represents a probability measure. Column D of the figure 2.3 illustrates Bayesian inference in DER. The marginal likelihood term can be represented mathematically as follows:

$$P(y_i|m) = \int_{\theta} P(y_i|\theta, m)P(\theta|m)d\theta \quad (2.13)$$

$$P(y_i|m) = \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} P(y_i|\mu, \sigma^2)P(\mu, \sigma^2|m)d\mu d\sigma^2 \quad (2.14)$$

The proposed loss function aims to determine the set of parameters m which maximizes the term $P(y_i|m)$ (evidence) for the given target y_i . Similar to Maximum Likelihood Estimation, the objective of maximization of marginal likelihood is re-framed as minimization of negative log of marginal-likelihood(NLL) for computational convenience. The loss function is expressed as:

Alternative to the usual way of minimizing Negative log likelihood the author proposes yet another form for the loss function which minimizes sum-of-squared errors between the prior and data sampled from the likelihood function. Following is the expression for the Sum Of Squared errors(SOS) version of the loss function:

The author claims the SOS version of loss function to be relatively stable while training and also to perform better than the other during evaluation. The portion of loss function \mathcal{L} described in this section only achieves the "model fitting" objective of DER.

Minimizing evidence on errors

The second objective of DER aims to minimize the evidence measure or to inflate uncertainty in the absence of training data. DER expresses this objective by adding a regularizer term to the loss function which penalizes the loss function in an event of its prediction deviating from the ground truth label. The penalty is scaled by evidence which expressed as sum of virtual observations as described in 2.11. Following is the expression for the regularizer term,

$$\mathcal{L}_i^R(w) = \|y_i - \gamma\| \cdot (2\alpha + \lambda) \quad (2.15)$$

Here p refers to the order of norm used to represent the difference between the ground truth label y_i and predicted mean γ . Author uses the value of $p=1$ claiming it to be stable during the training process.

Putting both its objectives together the evidential loss function can be expressed as:

$$\mathcal{L}_i(w) = \mathcal{L}_i^{SOS}(w) + \mathcal{L}_i^R(w) \quad (2.16)$$

2.3.5 Estimating uncertainty

Epistemic uncertainty which quantifies the model's inherent lack of knowledge associated with an output can be expressed as the variance around its predictions.

$$Var(\mu) = \frac{\beta}{(\alpha - 1)\lambda} \quad (2.17)$$

The λ term in the denominator refers to the number of virtual observations. Aleatoric uncertainty can be computed with the following expression:

$$\mathbb{E}[\sigma^2] = \frac{\beta}{\alpha - 1} \quad (2.18)$$

From equations 2.18 and 2.17 both components of uncertainty can be related as follows:

$$epistemic_uncertainty = \frac{aleatoric_uncertainty}{\lambda} \quad (2.19)$$

This means that the epistemic uncertainty component is the mean of aleatoric uncertainty over λ virtual observations.

Predictive uncertainty can be evaluated as the sum of epistemic and aleatoric uncertainty components.

$$predictive_uncertainty = aleatoric_uncertainty + epistemic_uncertainty \quad (2.20)$$

From eqns 2.18 and 2.17 predictive uncertainty can be computed as follows:

$$predictive_uncertainty = \frac{\beta}{\alpha - 1} + \frac{\beta}{(\alpha - 1)\lambda} \quad (2.21)$$

After simplification,

$$predictive_uncertainty = \frac{\beta(1 + \lambda)}{(a - 1)\lambda} \quad (2.22)$$

3

Methodology

How you are planning to test/compare/evaluate your research. Criteria used.

4

Solution

Your main contributions go here

4.1 Proposed algorithm

4.2 Implementation details

5

Evaluation

Implementation and measurements.

6

Results

6.1 Use case 1

Describe results and analyse them

6.2 Use case 2

6.3 Use case 3

7

Conclusions

7.1 Contributions

7.2 Lessons learned

7.3 Future work



Design Details

Your first appendix

B

Parameters

Your second chapter appendix

References

- [1] Author Name. Book title. *Lecture Notes in Autonomous System*, 1001:900–921, 2003. ISSN 0302-2345.