Motivation

Pruning

Pruning

# Bayesian Deep Learning with 10 % of the weights

Practical approach to Bayesian deep learning

Rob Romijnders

robromijnders.github.io

PyData Amsterdam, 2018

Motivation

Pruning

### Outline

Introduction Motivation

2 Method

The goal Historical perspective Bayesian inference Parameter posterior Uncertainty Pruning

- 3 Experiments and results Pruning Uncertainties
- 4 Closing

### Motivation

#### iviotivatio

The goal Historical perspective Bayesian inference

Parameter posterior Uncertaint Pruning

Experiment and results

Pruning Uncertaintie

Closina

### Outline

1 Introduction Motivation

2 Method

The goal
Historical perspective
Bayesian inference
Parameter posterior
Uncertainty
Pruning

- 3 Experiments and results
  Pruning
  Uncertainties
- 4 Closing

### Motivation

Metho

The goal Historical perspectiv Bayesian inference

Parameter posterior Uncertaint Pruning

and results
Pruning

Closing

### Problems with neural networks

Neural networks have three problems:

- Neural networks give no uncertainty in predictions

   → easily fooled by adversarial examples
- 2 Neural networks have millions of parameters

#### Introductio

### Motivation

Method

The goal
Historical
perspective
Bayesian
inference
Parameter
posterior

Pruning

Experiment and results

Pruning Uncertaintie

Closin

## Motivation



aorta thoracic / tortuous







calcified\_granuloma / lung / hilum / right

Figure: Uncertainty is important when making diagnoses

> Rob Romijnders

Motivation

perspective

Pruning

Pruning

### Motivation



Figure: Uncertainty is important when making a critical decision

Introduction

Motivation

Metho

The goal Historical perspectiv Bayesian inference

Paramete posterior

Uncertai Pruning

and results
Pruning

Closina

### Motivation

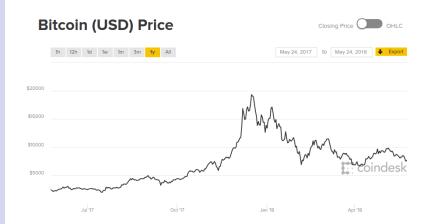


Figure: Uncertainty is important when prediction bitcoin

### Rob Romijnders

Introducti

Motivation

Method
The goal
Historical
perspective

Bayesian inference Parameter posterior Uncertaint

Pruning
Experiment
and results

Pruning Uncertaintie

Closin

## Adversarial attack

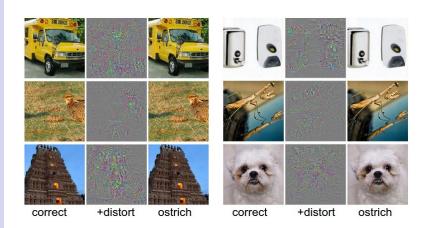


Figure: Uncertainty is necessary to find adversarial examples

Roh Romijnders

Motivation

Pruning

Pruning

# Embedded applications



Figure: Pruning reduces the memory and computation usage (Pruning = dropping parameters)

Rob Romijnders

Introductio

Motivation

. . . . .

Metho The go

Parameter posterior

posterior Uncertai Pruning

Experiment and result

Pruning Uncertaintie

Closin

### Real time inference

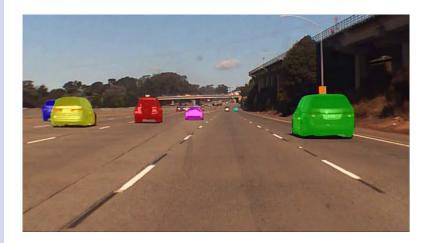


Figure: Pruning reduces the computation requirements

> Rob Romijnders

Introduction

Motivation

### ${\sf Method}$

The goal Historical perspective Bayesian inference Parameter posterior Uncertainty Pruning

and results

Pruning

Closing

Motivation

### The goal

Pruning

### Outline

Introduction

### Method

### The goal

- 3 Experiments and results

#### Roh Romiinders

### The goal

Pruning

Pruning

### Pseudo code

In summary, this talk covers the following pseudo code

```
model = Model()
model.train(data)
model.prune()
```

```
# Actually, the next line is all we care about:
prediction, uncertainty = model.predict(input)
```

> Rob Romijnders

Introduction

Motivation

Metho

The goal

Historic

Bavesia

inferenc

Paramete

posterior

Uncertain

Pruning

experiments

and results

Uncertainti

Pruning

How to make a prediction?

Introducti

IVIOTIVATIOI

The gen

Historical perspective

Bayesian inference

Parameter posterior Uncertain

Pruning

and results

Pruning Uncertainties

Closing

## Outline

1 Introduction
Motivation

2 Method

The goa

Historical perspective

Bayesian inference

Parameter posterior Uncertainty

Pruning

Pruning

3 Experiments and results
Pruning

4 Closing

Motivation

Mothod

The goal

Historical perspective Bayesian

Parameter posterior

Uncertain Pruning

and results
Pruning

Closin

# Historical perspective

This content is not new: it has been around for decennia/centuries

## Being Bayesian about neural networks

- Bayes lived in 18th century
- Variational inference for neural networks: Hinton and van Camp (1993)
- Bayesian Neural networks: Neal (1995)

### Uncertainties for a model

Shannon published information theory in 1948

Motivation

Bavesian inference

Pruning

## Outline

Introduction

Method

Bayesian inference

- 3 Experiments and results

### Roh Romijnders

Motivation

Bavesian

inference

Pruning

Pruning

# Bayes rule

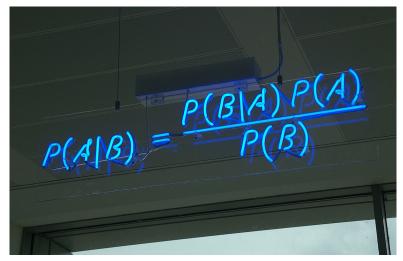


Figure: Every presentation on Bayesian machine learning has this image. So this presentation too

Motivation

Bavesian

#### inference

Pruning

Pruning

## Bayes rule

$$posterior \propto likelihood imes prior$$
 $p(w|data) \propto p(data|w)p(w)$ 

$$logp(w|data) = logp(data|w) + logp(w) + constant$$

Motivation

perspective

Bayesian

inference

Pruning

Pruning

## Bayes rule

We have been using Bayes' rule all the time

Motivation

Bavesian

inference

Pruning

Pruning

# Weight decay .. L2 regularisation

$$-log\ posterior = -log\ likelihood \ -log\ prior \ +constant$$
 
$$loss = classification\ loss \ +\lambda \sum_i w_i^2 \ +constant$$

Rob Romijnders

Introduction

Motivation

1110111111111

Metho

The

Histor

perspec

Bayesian

inference

Paramet

Uncertain Pruning

Experime

Pruning
Uncertaintie

Closina

# Stochastic gradient descent

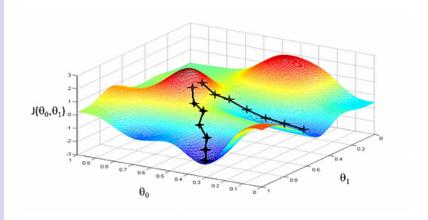


Figure: But we inferred only one parameter vector

Motivation

perspective

Bavesian

inference

Pruning

Pruning

# From this...

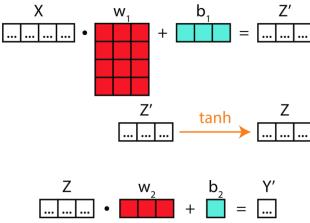


Figure: Used with kind permission of Eric Ma

Motivation

perspective

Bavesian

inference

Pruning

Pruning

...to this

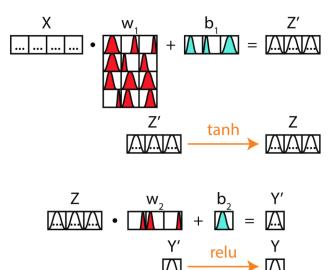


Figure: Used with kind permission of Eric Ma

#### Rob Romijnders

Motivation

perspective

Bayesian

inference

Pruning

Pruning

### Parameters of a Gaussian

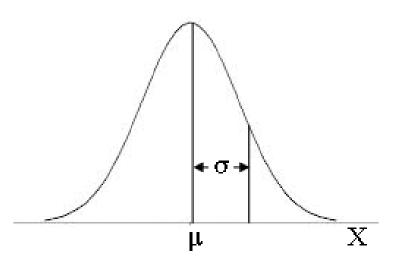


Figure: For a Gaussian, we need parameters  $\mu$  and  $\sigma$ 

Introducti

Motivation

Metho

The goal Historical perspectiv Bayesian

Parameter posterior

Uncertain

Experiment

Pruning Uncertainties

Closing

## Outline

1 Introduction
Motivation

2 Method

The goal Historical perspective Bayesian inference

Parameter posterior

Uncertainty Pruning

3 Experiments and results
Pruning
Uncertainties

4 Closing

Introduction

Motivation

Madelan

The goal Historical perspecti

Parameter posterior

Uncertaint

Experiment and results Pruning

Closing

# Posterior probability

### The parameter posterior will:

- ullet Enable more samples for prediction o uncertainty over prediction
- ullet Tell us which parameters have high zero-probability ightarrow pruning

#### Parameter posterior

Pruning

### Loss functions

### Loss functions

### old loss

$$= \textit{classification loss} + \sum_{\textit{i}} \underbrace{\lambda \textit{w}_{\textit{i}}^2}_{\text{L2 penalty}} + \textit{constant}$$

### new loss

$$= classification \ loss + \sum_{i} \underbrace{\frac{1}{2} \lambda \mu_{i}^{2}}_{\text{L2 penalty}} - \underbrace{\log \sigma_{i} + \frac{1}{2} \lambda \sigma_{i}^{2}}_{\text{penalty on } \sigma} + constant$$

### Interpretation

Parameter posterior

Pruning

Pruning

## Intuition

$$loss = \underbrace{classification \ loss + \sum_{i} \frac{1}{2} \lambda \mu_{i}^{2}}_{\text{loss on location of weights}} - \underbrace{\log \sigma_{i} + \frac{1}{2} \lambda \sigma_{i}^{2}}_{\text{loss on } \sigma} + constant$$

Meth

The goal Historical perspectiv Bayesian inference

Parameter posterior

Uncertaint

Experiment and results Pruning

Oncerta

# Summary

- What do we care about?
   Uncertainties and pruning
- How we do it?
   Find many parameter vectors and average
- How we do that?
   Bayesian inference
- How we do that?
   Approximate the parameter posterior
- What do we do in the end?
   Minimize the loss function on the previous slide

milioducti

Motivation

Method

The goal Historical perspective Bayesian inference

Uncertainty

Pruning

Experiment

Pruning
Uncertainties

Closing

### Outline

1 Introduction
Motivation

2 Method

The goal
Historical perspective
Bayesian inference
Parameter posterior

Uncertainty

Pruning

3 Experiments and results
Pruning
Uncertainties

4 Closing

#### Rob Romijnders

Motivation

Uncertainty

Pruning

Pruning

# Use entropy as uncertainty metric

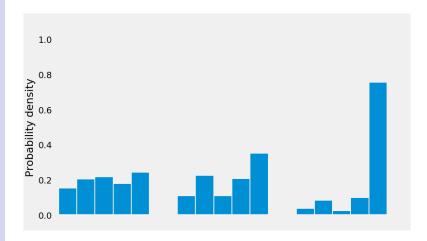


Figure: Which prediction has least uncertainty?

#### Rob Romijnders

Introduction

Motivation

Metho

The goal Historical perspectiv Bayesian

inference Parameter

Uncertainty

Uncertain Pruning

Experiment and results
Pruning

c. .

# Use entropy as uncertainty metric

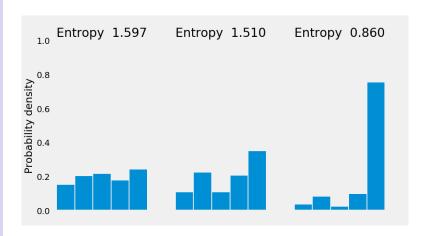


Figure: Which prediction has least uncertainty?

Motivation

111011144110

The goa

The goal Historical perspective Bayesian inference

Parameter posterior

Pruning

and results
Pruning

Officertal

## Outline

IntroductionMotivation

2 Method

The goal
Historical perspectiv
Bayesian inference
Parameter posterior
Uncertainty
Pruning

- 3 Experiments and results
  Pruning
  Uncertainties
- 4 Closing

Introduction

Motivation

Meth

The goal

Historical perspective Bayesian

Parameter

posterior

Uncertaint

#### Pruning

# and result

Pruning Uncertaintie

Clasian

## From this...

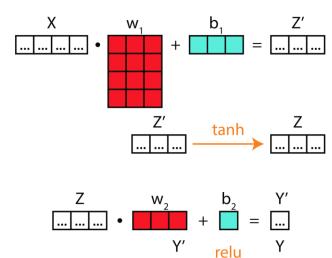


Figure: Used with kind permission of Eric Ma

Introductio

Motivation

8.4

ivietno

The goal Historica

perspectiv Bavesian

inference

raramei

I la acada i a

Pruning

# and results

Uncertaintie

Closing

## From this...

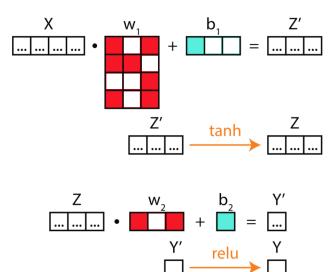


Figure: Used with kind permission of Eric Ma

#### Rob Romijnders

Introduction

Motivation

Metho

The goal Historical perspection

inference Parameter posterior

Uncert

Pruning

and results
Pruning

Clasia

### Pruning according to posterior

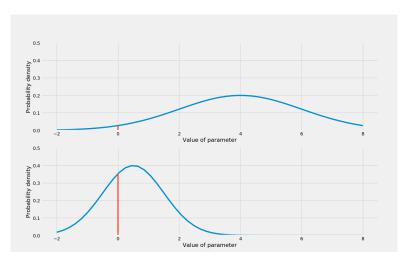


Figure: Which parameter would you rather prune?

#### Roh Romiinders

Pruning

Experiments and results

Pruning

#### Data sets

#### Fun

No deep learning project is complete without **MNIST** 

#### Serious

Two most common applications of deep learning:

- Image recognition: CIFAR10 data set
- Time series classification: UCR ECG's
  - Train set only 500 time series → Bayesian's don't overfit

Introduct

Motivation

N. A. . . . .

Metho

The goa

perspectiv Bayesian inference

Paramet posterior

Uncertain Pruning

Experiments and results

Pruning Uncertaintie

Closin

### MNIST examples

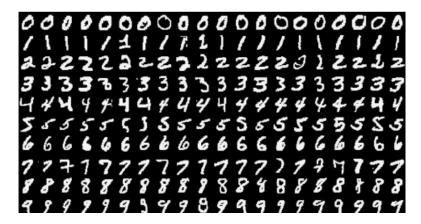


Figure: Examples of MNIST. Train set: 50k samples. Test set: 10k samples

#### Rob Romijnders

Introduction

Motivation

Moth

ivietn

The goal

perspectiv Bayesian inference

Paramet

Uncertai

Pruning

### Experiments and results

Pruning Uncertainti

Closing

### CIFAR examples

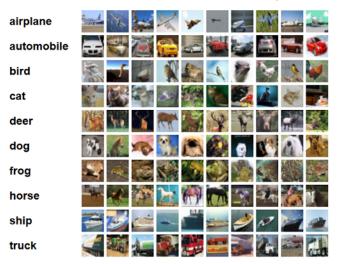


Figure: Examples of CIFAR. Train set: 50k samples. Test set: 10k samples

Introduction

Motivation

Metho

The goa

perspective

Bayesian inference

Paramete

posterior

Uncertai

Pruning

### Experiments and results

Pruning Uncertainti

Closing

### ECG examples

??? MAKE ECG EXAMPLES HERE

Motivation

IVIOLIVALIO

Metho

The goal Historical perspective Bayesian

Parameter posterior

Pruning

and results

Pruning Uncertaintie

Closino

### Outline

IntroductionMotivation

2 Method

The goal
Historical perspective
Bayesian inference
Parameter posterior
Uncertainty
Pruning

3 Experiments and results
Pruning

4 Closing

Introduction Motivation

. . . . . .

The goal Historical perspective Bayesian

Paramet posterior

Uncertain Pruning

Experiment

and results
Pruning

Uncertaintie

Closing

# Remember the model ...to this

$$Z$$
 $W_2$ 
 $Y'$ 
 $Y'$ 
 $Y'$ 
 $Y'$ 
 $Y'$ 

Rob Romijnders

Introduction

Motivation

Metho

The m

Historica perspecti

Bayesian inference

Parame

posterior

Uncertain Pruning

Experime and result

Pruning

Closin

### Pruning MNIST

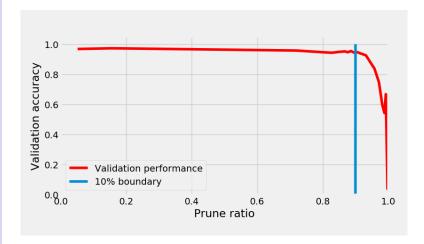


Figure: Pruning curve for MNIST

Rob Romijnders

Introduction Motivation

Motivation

The go

Historical perspective Bayesian inference Parameter posterior

Pruning

and results

Uncertainti

Closin

### Pruning CIFAR

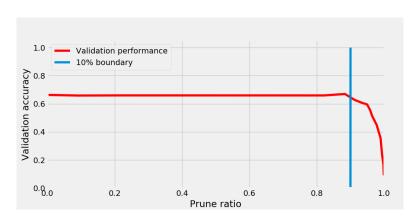


Figure: Pruning curve for CIFAR

#### Rob Romijnders

Introduction

Motivation

Metho

The goal Historical perspectiv Bayesian

Paramete posterior Uncertain

Pruning

and result

Pruning

Clasia

### Pruning ECG

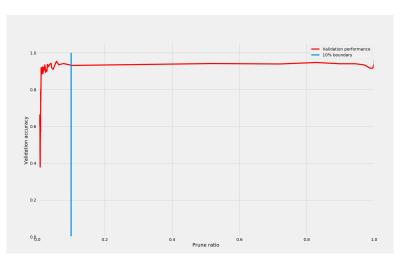


Figure: Pruning curve for ECG

Motivation

....

The goal

perspective Bayesian inference Parameter

Uncertaint

Experimen

Pruning Uncertainties

Closing

### Outline

IntroductionMotivation

2 Method

The goal
Historical perspective
Bayesian inference
Parameter posterior
Uncertainty
Pruning

3 Experiments and results
Pruning
Uncertainties

4 Closing

Introductio

Motivation

Markey

Metho

Historical perspectiv Bayesian

Bayesian inference Parameter

posterior Uncertaint

Uncertain Pruning

and results

Pruning Uncertainties

Closing

### Experiment uncertainty

How to mutilate images to raise uncertainty?

- Add noise
- Warping

#### Rob Romijnders

Introduction

Motivation

Metho

The goal Historical perspective Bayesian inference

Parameter posterior Uncertaint

Pruning

and results

Pruning Uncertainties

Closing

Here will be slides with experiments to show uncertainty on CIFAR10 when we add noise or rotate or warp  $\,$ 

Motivation

IVIOLIVALIO

Metho

The goal Historical perspection Bayesian inference

Parameter posterior Uncertaint Pruning

Experiment and results

Pruning Uncertaintie

Closing

### Take aways

- Get uncertainty for critical predictions
- Robust against adversarial attacks
- Prune networks for small memory and small compute

Introduction

Motivation

Method

The goal Historical perspective Bayesian

inference Parameter posterior

Uncertain Pruning

and results
Pruning
Uncertaintie

Closing

# Questions?

robromijnders.github.io

#### Material

github.com/RobRomijnders/weight\_uncertainty

- All code
- Further reading
- More explanation

> Rob Romijnders

Introduction Motivation

Metho

Historical perspective Bayesian

Paramete

posterior Uncertain

Uncertain Pruning

Experiment and results

Pruning Uncertaintie

Closing

## Additional slides

#### Introduction

Motivation

Metho

The goal Historical perspective Bayesian inference Parameter posterior

Pruning
Experimentand results

Pruning Uncertaintie

#### Closing

### Learning the sigma's



Figure: The VI objective increases the sigma's by itself!!

Motivation

perspective

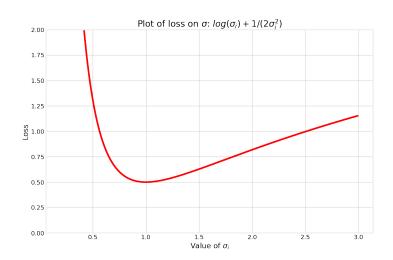
Pruning

Pruning

Closing

### Loss on $\sigma$

#### What does the loss for $\sigma$ look like?



#### Roh Romiinders

Pruning

Closing

### Make predictions

### Sampling

Make multiple predictions with sampled parameters. One can think of this sampling as an ensemble method

```
make_prediction(input):
    for param_vec in param_vecs:
        yield model.get_output(input, param_vec)
prediction = np.mean(make_prediction(input))
```

Introduction

AA .: .:

Motho

The

Historica

Bayesiar

Parame

posterior

Uncertain Pruning

\_\_\_\_\_

and results

Pruning Uncertaintie

#### Closing

```
Pseudo code
```

Pseudo code for training our neural network

```
# OLD CODE
while not converged:
  # Get the loss
  x, y = sample_batch()
  loss = loss\_function(x, y, w)
  #Update the parameters
  w_grad = gradient(loss, w)
  w = update(w, w_grad)
# NEW CODE
while not converged:
  # Get the loss
  x, y = sample_batch()
  w = approximation.sample()
  loss = loss\_function(x, y, w)
  # Update the approximation
  w_grad = gradient(loss, w)
  approximation = update(approximation, w_grad)
```

```
Bayesian Deep
Learning with
10 % of the
weights
```

#### Introduction

Motivation

Metho

IVIELIIO

The goal Historical perspective

Bayesian inference Parameter

Parameter

Uncertain Pruning

Experiments and results

Pruning Uncertainties

#### Closing

```
while not converged:
    # Get the loss
    x, y = sample_batch()
    w = approximation.sample()
    loss = loss_function(x, y, w)

# Update the approximation
    w_grad = gradient(loss, w)
    approximation = update(approximation, w_grad)
```

Introduction

Motivation

Madelan

The goal Historica perspect

perspective Bayesian inference Parameter

posterior Uncertain Pruning

Experime

Pruning

Closing

### Research

Pruning: speed

Bayesian compression for deep learning, Louizos @ NIPS2017

Uncertainty: adversarial attack

Adversarial phenomenon in Bayesian deep learning, Rawat, 2017

#### Roh Romiinders

Pruning

Closing

### Gaussian approximation

Approximate with a normal distribution

- Captures local structure of the posterior, which indicates the uncertainty
- Simple for parameter pruning

# Anything is better than point estimation !!!