Embedded AI: Principles, Algorithms, and Applications

Vijay Janapa Reddi (Harvard University) and Song Han (MIT)

2023-09-05

Table of contents

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

The Philosophy Behind the Book

Prerequisites

Conventions Used in This Book

How to Contact Us

Contributors

1 About Us

1.1 Who's This Book For

1.2 Course Structure

1.3 Course Requirements

1.4 Course Materials

1.5 What You'll Learn

2 Introduction

2.1 AI for Beginners

2.2 Machine Learning and Deep Learning

2.3 Machine Learning

2.4 Deep Learning

2.5 Applications of Deep Learning

2.6 Quiz

3 Embedded ML

3.1 CloudML

3.2 EdgeML

3.3 TinyML

3.3.1 TinyML for IoT Systems

3.3.2 How does TinyML Work

3.3.3 Resources are Limited, but so is the Competition

3.4 Exercises

4 Deep Learning Primer

4.1 What are Neural Networks

4.2 What is Deep Learning Training

4.3 What is Deep Learning Inference

5 Machine Learning Workflow

5.1 Data Collection

5.2 Pre-Processing

5.3 Training

5.4 Optimization

5.5 Deployment

5.6 Evaluation

5.7 Quiz

6 Data Collection

6.1 Data Sources

6.2 Training Data

6.3 Training Data Splits

6.4 Data Labeling

6.5 Types of Data

7 Pre-processing

7.1 What is Data Pre-processing?

7.2 What's Involved with Data Pre-processing?

7.3 What's The Importance Of Data Pre-Processing?

8 Feature Engineering

coming soon.

9 Model Training

9.1 Selecting a Training Dataset

9.2 Neural Network Architectures

9.2.1 Multilayer Perceptron (MLP)

9.2.2 Convolutional Neural Networks

9.2.3 Recurrent Neural Networks

9.2.4 Transformers

9.3 Back Propagation

9.4 Convergence

9.5 Overfitting and Underfitting

9.6 Hyperparameters

9.6.1 Epochs

9.6.2 Learning Rate

9.7 Transfer Learning

9.7.1 Optimizer

9.8 Summary

9.9 Quiz

10 Optimizations

10.1 Software Optimizations

10.1.1 Compression

10.1.2 Quantization

10.1.3 Weight Pruning

10.1.4 Knowledge Distillation

10.2 Hardware Optimizations

10.2.1 GPUs

10.2.2 TPUs

10.2.3 NPUs

11 Deployment

12 MLOps

References