

# Table of Contents

<b>About the Author .....</b>	<b>xi</b>
<b>About the Technical Reviewer .....</b>	<b>xiii</b>
<b>Acknowledgments .....</b>	<b>xv</b>
<b>Introduction .....</b>	<b>xvii</b>
<b>Chapter 1: Computational Graphs and TensorFlow.....</b>	<b>1</b>
How to Set Up Your Python Environment .....	1
Creating an Environment.....	3
Installing TensorFlow.....	9
Jupyter Notebooks .....	11
Basic Introduction to TensorFlow.....	14
Computational Graphs.....	14
Tensors.....	17
Creating and Running a Computational Graph .....	19
Computational Graph with tf.constant.....	19
Computational Graph with tf.Variable.....	20
Computational Graph with tf.placeholder.....	22
Differences Between run and eval .....	25
Dependencies Between Nodes.....	26
Tips on How to Create and Close a Session .....	27
<b>Chapter 2: Single Neuron.....</b>	<b>31</b>
The Structure of a Neuron.....	31
Matrix Notation.....	35
Python Implementation Tip: Loops and NumPy .....	36
Activation Functions .....	38

## TABLE OF CONTENTS

Cost Function and Gradient Descent: The Quirks of the Learning Rate .....	47
Learning Rate in a Practical Example .....	50
Example of Linear Regression in tensorflow .....	57
Example of Logistic Regression .....	70
Cost Function.....	70
Activation Function.....	71
The Dataset .....	71
tensorflow Implementation.....	75
References .....	80
<b>Chapter 3: Feedforward Neural Networks .....</b>	<b>83</b>
Network Architecture .....	84
Output of Neurons .....	87
Summary of Matrix Dimensions .....	88
Example: Equations for a Network with Three Layers .....	88
Hyperparameters in Fully Connected Networks .....	90
softmax Function for Multiclass Classification .....	90
A Brief Digression: Overfitting.....	91
A Practical Example of Overfitting.....	92
Basic Error Analysis.....	99
The Zalando Dataset .....	100
Building a Model with tensorflow .....	105
Network Architecture.....	106
Modifying Labels for the softmax Function—One-Hot Encoding .....	108
The tensorflow Model.....	110
Gradient Descent Variations.....	114
Batch Gradient Descent.....	114
Stochastic Gradient Descent .....	116
Mini-Batch Gradient Descent.....	117
Comparison of the Variations.....	119
Examples of Wrong Predictions .....	123
Weight Initialization.....	125

Adding Many Layers Efficiently .....	127
Advantages of Additional Hidden Layers.....	130
Comparing Different Networks .....	131
Tips for Choosing the Right Network .....	135
<b>Chapter 4: Training Neural Networks .....</b>	<b>137</b>
Dynamic Learning Rate Decay .....	137
Iterations or Epochs?.....	139
Staircase Decay.....	140
Step Decay .....	142
Inverse Time Decay .....	145
Exponential Decay .....	148
Natural Exponential Decay .....	150
tensorflow Implementation.....	158
Applying the Methods to the Zalando Dataset.....	162
Common Optimizers.....	163
Exponentially Weighted Averages.....	163
Momentum .....	167
RMSProp.....	172
Adam .....	175
Which Optimizer Should I Use? .....	177
Example of Self-Developed Optimizer .....	179
<b>Chapter 5: Regularization .....</b>	<b>185</b>
Complex Networks and Overfitting .....	185
What Is Regularization? .....	190
About Network Complexity .....	191
$\ell_p$ Norm .....	192
$\ell_2$ Regularization .....	192
Theory of $\ell_2$ Regularization .....	192
tensorflow Implementation.....	194
$\ell_1$ Regularization .....	205

## TABLE OF CONTENTS

Theory of $\ell_1$ Regularization and tensorflow Implementation .....	206
Are Weights Really Going to Zero? .....	208
Dropout .....	211
Early Stopping.....	215
Additional Methods .....	216
<b>Chapter 6: Metric Analysis .....</b>	<b>217</b>
Human-Level Performance and Bayes Error.....	218
A Short Story About Human-Level Performance.....	221
Human-Level Performance on MNIST.....	223
Bias .....	223
Metric Analysis Diagram .....	225
Training Set Overfitting .....	225
Test Set .....	228
How to Split Your Dataset.....	230
Unbalanced Class Distribution: What Can Happen.....	234
Precision, Recall, and F1 Metrics.....	239
Datasets with Different Distributions.....	245
K-Fold Cross-Validation .....	253
Manual Metric Analysis: An Example .....	263
<b>Chapter 7: Hyperparameter Tuning.....</b>	<b>271</b>
Black-Box Optimization.....	271
Notes on Black-Box Functions .....	273
The Problem of Hyperparameter Tuning .....	274
Sample Black-Box Problem .....	275
Grid Search .....	277
Random Search.....	282
Coarse-to-Fine Optimization .....	285
Bayesian Optimization .....	289
Nadaraya-Watson Regression .....	290
Gaussian Process .....	291

Stationary Process .....	292
Prediction with Gaussian Processes .....	292
Acquisition Function .....	298
Upper Confidence Bound (UCB) .....	299
Example .....	300
Sampling on a Logarithmic Scale .....	310
Hyperparameter Tuning with the Zalando Dataset.....	312
A Quick Note on the Radial Basis Function .....	321
<b>Chapter 8: Convolutional and Recurrent Neural Networks.....</b>	<b>323</b>
Kernels and Filters .....	323
Convolution .....	325
Examples of Convolution.....	334
Pooling .....	342
Padding .....	345
Building Blocks of a CNN .....	346
Convolutional Layers .....	347
Pooling Layers .....	349
Stacking Layers Together .....	349
Example of a CNN .....	350
Introduction to RNNs .....	355
Notation .....	357
Basic Idea of RNNs .....	358
Why the Name <i>Recurrent</i> ? .....	359
Learning to Count .....	359
<b>Chapter 9: A Research Project.....</b>	<b>365</b>
The Problem Description.....	365
The Mathematical Model .....	369
Regression Problem .....	369
Dataset Preparation .....	375
Model Training.....	384

TABLE OF CONTENTS

**Chapter 10: Logistic Regression from Scratch ..... 391**

- Mathematics Behind Logistic Regression ..... 392
- Python Implementation ..... 395
- Test of the Model ..... 398
  - Dataset Preparation ..... 398
  - Running the Test ..... 400
- Conclusion ..... 401

**Index ..... 403**