

---

# Table of Contents

Preface.....	xiii
--------------	------

---

## Part I. The Fundamentals of Machine Learning

<b>1. The Machine Learning Landscape.....</b>	<b>3</b>
What Is Machine Learning?	4
Why Use Machine Learning?	4
Types of Machine Learning Systems	7
Supervised/Unsupervised Learning	8
Batch and Online Learning	14
Instance-Based Versus Model-Based Learning	17
Main Challenges of Machine Learning	22
Insufficient Quantity of Training Data	22
Nonrepresentative Training Data	24
Poor-Quality Data	25
Irrelevant Features	25
Overfitting the Training Data	26
Underfitting the Training Data	28
Stepping Back	28
Testing and Validating	29
Exercises	31
<b>2. End-to-End Machine Learning Project.....</b>	<b>33</b>
Working with Real Data	33
Look at the Big Picture	35
Frame the Problem	35
Select a Performance Measure	37

Check the Assumptions	40
Get the Data	40
Create the Workspace	40
Download the Data	43
Take a Quick Look at the Data Structure	45
Create a Test Set	49
Discover and Visualize the Data to Gain Insights	53
Visualizing Geographical Data	53
Looking for Correlations	55
Experimenting with Attribute Combinations	58
Prepare the Data for Machine Learning Algorithms	59
Data Cleaning	60
Handling Text and Categorical Attributes	62
Custom Transformers	64
Feature Scaling	65
Transformation Pipelines	66
Select and Train a Model	68
Training and Evaluating on the Training Set	68
Better Evaluation Using Cross-Validation	69
Fine-Tune Your Model	71
Grid Search	72
Randomized Search	74
Ensemble Methods	74
Analyze the Best Models and Their Errors	74
Evaluate Your System on the Test Set	75
Launch, Monitor, and Maintain Your System	76
Try It Out!	77
Exercises	77
<b>3. Classification.....</b>	<b>79</b>
MNIST	79
Training a Binary Classifier	82
Performance Measures	82
Measuring Accuracy Using Cross-Validation	83
Confusion Matrix	84
Precision and Recall	86
Precision/Recall Tradeoff	87
The ROC Curve	91
Multiclass Classification	93
Error Analysis	96
Multilabel Classification	100
Multioutput Classification	101

Exercises	102
<b>4. Training Models.....</b>	<b>105</b>
Linear Regression	106
The Normal Equation	108
Computational Complexity	110
Gradient Descent	111
Batch Gradient Descent	114
Stochastic Gradient Descent	117
Mini-batch Gradient Descent	119
Polynomial Regression	121
Learning Curves	123
Regularized Linear Models	127
Ridge Regression	127
Lasso Regression	130
Elastic Net	132
Early Stopping	133
Logistic Regression	134
Estimating Probabilities	134
Training and Cost Function	135
Decision Boundaries	136
Softmax Regression	139
Exercises	142
<b>5. Support Vector Machines.....</b>	<b>145</b>
Linear SVM Classification	145
Soft Margin Classification	146
Nonlinear SVM Classification	149
Polynomial Kernel	150
Adding Similarity Features	151
Gaussian RBF Kernel	152
Computational Complexity	153
SVM Regression	154
Under the Hood	156
Decision Function and Predictions	156
Training Objective	157
Quadratic Programming	159
The Dual Problem	160
Kernelized SVM	161
Online SVMs	164
Exercises	165

<b>6. Decision Trees.....</b>	<b>167</b>
Training and Visualizing a Decision Tree	167
Making Predictions	169
Estimating Class Probabilities	171
The CART Training Algorithm	171
Computational Complexity	172
Gini Impurity or Entropy?	172
Regularization Hyperparameters	173
Regression	175
Instability	177
Exercises	178
<b>7. Ensemble Learning and Random Forests.....</b>	<b>181</b>
Voting Classifiers	181
Bagging and Pasting	185
Bagging and Pasting in Scikit-Learn	186
Out-of-Bag Evaluation	187
Random Patches and Random Subspaces	188
Random Forests	189
Extra-Trees	190
Feature Importance	190
Boosting	191
AdaBoost	192
Gradient Boosting	195
Stacking	200
Exercises	202
<b>8. Dimensionality Reduction.....</b>	<b>205</b>
The Curse of Dimensionality	206
Main Approaches for Dimensionality Reduction	207
Projection	207
Manifold Learning	210
PCA	211
Preserving the Variance	211
Principal Components	212
Projecting Down to $d$ Dimensions	213
Using Scikit-Learn	214
Explained Variance Ratio	214
Choosing the Right Number of Dimensions	215
PCA for Compression	216
Incremental PCA	217
Randomized PCA	218

Kernel PCA	218
Selecting a Kernel and Tuning Hyperparameters	219
LLE	221
Other Dimensionality Reduction Techniques	223
Exercises	224

---

## Part II. Neural Networks and Deep Learning

<b>9. Up and Running with TensorFlow.....</b>	<b>229</b>
Installation	232
Creating Your First Graph and Running It in a Session	232
Managing Graphs	234
Lifecycle of a Node Value	235
Linear Regression with TensorFlow	235
Implementing Gradient Descent	237
Manually Computing the Gradients	237
Using autodiff	238
Using an Optimizer	239
Feeding Data to the Training Algorithm	239
Saving and Restoring Models	241
Visualizing the Graph and Training Curves Using TensorBoard	242
Name Scopes	245
Modularity	246
Sharing Variables	248
Exercises	251
<b>10. Introduction to Artificial Neural Networks.....</b>	<b>253</b>
From Biological to Artificial Neurons	254
Biological Neurons	255
Logical Computations with Neurons	256
The Perceptron	257
Multi-Layer Perceptron and Backpropagation	261
Training an MLP with TensorFlow’s High-Level API	264
Training a DNN Using Plain TensorFlow	265
Construction Phase	265
Execution Phase	269
Using the Neural Network	270
Fine-Tuning Neural Network Hyperparameters	270
Number of Hidden Layers	270
Number of Neurons per Hidden Layer	272
Activation Functions	272

- 11. Training Deep Neural Nets..... 275**
  - Vanishing/Exploding Gradients Problems 275
    - Xavier and He Initialization 277
    - Nonsaturating Activation Functions 279
    - Batch Normalization 282
    - Gradient Clipping 286
  - Reusing Pretrained Layers 286
    - Reusing a TensorFlow Model 287
    - Reusing Models from Other Frameworks 288
    - Freezing the Lower Layers 289
    - Caching the Frozen Layers 290
    - Tweaking, Dropping, or Replacing the Upper Layers 290
    - Model Zoos 291
    - Unsupervised Pretraining 291
    - Pretraining on an Auxiliary Task 292
  - Faster Optimizers 293
    - Momentum optimization 294
    - Nesterov Accelerated Gradient 295
    - AdaGrad 296
    - RMSProp 298
    - Adam Optimization 298
    - Learning Rate Scheduling 300
  - Avoiding Overfitting Through Regularization 302
    - Early Stopping 303
      - $\ell_1$  and  $\ell_2$  Regularization 303
    - Dropout 304
    - Max-Norm Regularization 307
    - Data Augmentation 309
  - Practical Guidelines 310
  - Exercises 311
  
- 12. Distributing TensorFlow Across Devices and Servers..... 313**
  - Multiple Devices on a Single Machine 314
    - Installation 314
    - Managing the GPU RAM 317
    - Placing Operations on Devices 318
    - Parallel Execution 321
    - Control Dependencies 323
  - Multiple Devices Across Multiple Servers 323
    - Opening a Session 325

The Master and Worker Services	325
Pinning Operations Across Tasks	326
Sharding Variables Across Multiple Parameter Servers	327
Sharing State Across Sessions Using Resource Containers	328
Asynchronous Communication Using TensorFlow Queues	329
Loading Data Directly from the Graph	335
Parallelizing Neural Networks on a TensorFlow Cluster	342
One Neural Network per Device	342
In-Graph Versus Between-Graph Replication	343
Model Parallelism	345
Data Parallelism	347
Exercises	352
<b>13. Convolutional Neural Networks.....</b>	<b>353</b>
The Architecture of the Visual Cortex	354
Convolutional Layer	355
Filters	357
Stacking Multiple Feature Maps	358
TensorFlow Implementation	360
Memory Requirements	362
Pooling Layer	363
CNN Architectures	365
LeNet-5	366
AlexNet	367
GoogLeNet	369
ResNet	372
Exercises	376
<b>14. Recurrent Neural Networks.....</b>	<b>379</b>
Recurrent Neurons	380
Memory Cells	382
Input and Output Sequences	382
Basic RNNs in TensorFlow	384
Static Unrolling Through Time	385
Dynamic Unrolling Through Time	387
Handling Variable Length Input Sequences	387
Handling Variable-Length Output Sequences	388
Training RNNs	389
Training a Sequence Classifier	389
Training to Predict Time Series	392
Creative RNN	396
Deep RNNs	396

Distributing a Deep RNN Across Multiple GPUs	397
Applying Dropout	399
The Difficulty of Training over Many Time Steps	400
LSTM Cell	401
Peephole Connections	403
GRU Cell	404
Natural Language Processing	405
Word Embeddings	405
An Encoder–Decoder Network for Machine Translation	407
Exercises	410
<b>15. Autoencoders.....</b>	<b>413</b>
Efficient Data Representations	414
Performing PCA with an Undercomplete Linear Autoencoder	415
Stacked Autoencoders	417
TensorFlow Implementation	418
Tying Weights	419
Training One Autoencoder at a Time	420
Visualizing the Reconstructions	422
Visualizing Features	423
Unsupervised Pretraining Using Stacked Autoencoders	424
Denoising Autoencoders	426
TensorFlow Implementation	427
Sparse Autoencoders	428
TensorFlow Implementation	429
Variational Autoencoders	430
Generating Digits	433
Other Autoencoders	434
Exercises	435
<b>16. Reinforcement Learning.....</b>	<b>439</b>
Learning to Optimize Rewards	440
Policy Search	442
Introduction to OpenAI Gym	443
Neural Network Policies	446
Evaluating Actions: The Credit Assignment Problem	449
Policy Gradients	450
Markov Decision Processes	455
Temporal Difference Learning and Q-Learning	459
Exploration Policies	461
Approximate Q-Learning	462
Learning to Play Ms. Pac-Man Using Deep Q-Learning	462



Exercises	471
Thank You!	472
<b>A. Exercise Solutions</b> .....	<b>473</b>
<b>B. Machine Learning Project Checklist</b> .....	<b>499</b>
<b>C. SVM Dual Problem</b> .....	<b>505</b>
<b>D. Autodiff</b> .....	<b>509</b>
<b>E. Other Popular ANN Architectures</b> .....	<b>517</b>
<b>Index</b> .....	<b>527</b>

