

第 1 章	序論：深層学習登場の前と後	
1.1	パターン認識とは	1
1.2	パターン認識の困難さと深層学習による成功	2
1.3	深層学習と従来のパターン認識手法の違い	6

第 2 章	深層学習以前のパターン認識手法	
2.1	深層学習以前のパターン認識の概略	8
2.2	特徴抽出	10
	1. 画像	11
	2. 音声	18
	3. テキスト	20
2.3	機械学習・パターン認識手法	22
	1. 最近傍法	23
	2. 線形識別関数	24
	3. ロジスティック回帰	25
	4. パーセプトロン	26
	5. サポートベクトルマシン	27
	6. アンサンブル学習	32
	7. 確率モデルによる分類	33
2.4	クラスタリング	35
	1. 階層的クラスタリング	35
	2. k -means 法	36
2.5	評価指標	37
	演習問題	41

第3章 深層学習ネットワーク

3.1 深層学習のアイデア	42
3.2 パーセプトロン	46
3.3 多層パーセプトロン	47
3.4 深層学習ネットワークにおける基本レイヤ群	49
1. 全結合層	49
2. 畳込み層	50
3. im2col 変換と行列積による畳込み演算	54
4. 畳込みと全結合の関係	58
5. プーリング層	58
6. ストライド付き畳込み	59
7. 転置畳込みとアンプーリング	61
8. 非線形活性化関数	63
9. 出力関数	64
3.5 基本ネットワーク構造	65
演習問題	67

第4章 ネットワークの学習

4.1 深層学習ネットワークの学習の基本的アイデア	68
4.2 誤差関数	72
4.3 確率的勾配降下法	74
4.4 誤差逆伝播法	76
1. 1入力1出力の全結合層の例	77
2. 一般的な全結合ネットワークでの例	82
4.5 畳込み層の学習	86
4.6 学習の実際	89
4.7 学習した畳込みフィルタの例	91
演習問題	92

第5章 学習のための技術

5.1 学習パラメータの初期値	93
5.2 学習率の設定	95
5.3 データ拡張	97
5.4 ドロップアウト	99
5.5 入力データの正規化	99
5.6 モデルアンサンブル	101
5.7 事前学習とファインチューニング	101
5.8 中間信号の画像特徴量としての利用	103
5.9 距離学習	106
5.10 マルチタスク学習	110
5.11 自己教師学習	112
5.12 ネットワークを小さくする工夫	114
1. 枝刈り	115
2. 量子化	116
3. 知識蒸留	117
演習問題	118

第6章 系列データへの対応

6.1 再帰型ネットワーク	119
1. RNN	119
2. LSTM	121
3. GRU	122
6.2 1次元畳込み	123
6.3 Transformer	124
1. Scaled Dot-Product Attention	126
2. Multi-Head Attention	128
3. Positional Encoding	128
4. 学習と推論	129

演習問題	130
------	-----

第7章 画像認識への適用

7.1 主な画像認識ネットワーク	131
1. LeNet	132
2. AlexNet	132
3. Network-In-Network	134
4. VGG	135
5. GoogleNet	137
6. ResNet	138
7. SE-Net	139
8. アーキテクチャ探索	140
9. 軽量ネットワーク	141
10. Vision Transformer	143
7.2 画像認識ネットワーク内部の可視化	144
7.3 物体検出	147
1. 深層学習による急激な性能向上	147
2. 2段階手法	149
3. 1段階手法	151
7.4 領域分割	153
7.5 人物姿勢推定	156
7.6 動画認識	157
演習問題	160

第8章 画像生成・変換への適用

8.1 エンコーダ・デコーダ型ネットワーク	161
8.2 オートエンコーダ	163
8.3 深層生成モデル	166
1. 変分オートエンコーダ	167

2. 敵対的生成ネットワーク	170
3. フローモデル	175
4. 拡散モデル	177
5. 画像生成結果の評価指標	180
8.4 画像変換	182
1. ペアデータがある場合の画像変換	184
2. 画像ドメイン変換	185
3. 形状とスタイルの特徴分離による画像操作	188
8.5 画像最適化による画像変換	190
1. クラス顕著性マップと敵対画像	190
2. DeepDream	194
3. 画像スタイル変換	194
4. 高速画像スタイル変換	196
演習問題	200

第9章 音声処理への適用

9.1 音声認識ネットワーク	201
1. DNN-HMM	203
2. CTC	204
3. RNN-T	207
4. LAS	209
5. Conformer	210
9.2 音声合成ネットワーク	212
1. WaveNet	213
2. Tacotron	215
3. WaveGlow	217
4. MelGAN	219
演習問題	221

第10章 自然言語処理への適用	
10.1 単語ベクトル	222
10.2 系列変換モデル	224
10.3 事前学習モデル	226
1. BERT	226
2. T5	228
3. GPT-3	229
演習問題	233
第11章 マルチモーダル学習	
11.1 マルチモーダル・クロスモーダル	234
11.2 画像と言語	235
1. 画像キャプション	236
2. 画像質問応答	238
3. テキストに基づく物体検出	239
4. クロスモーダル検索	240
5. テキストからの画像生成・変換	243
6. 画像言語マルチモーダル大規模基盤モデル	246
11.3 画像・映像と音声	250
演習問題	252
演習問題略解	253
参考文献	262
索引	270