

ЗМІСТ

Вступ	7
Позначення	10
Розділ 1. Еволюція глибинного навчання	13
1.1. Основи глибинного навчання	13
1.2. Етапи еволюції штучних нейронних мереж	15
1.3. Глибинне навчання: огляд та основні парадигми	24
1.3.1. Одношаровий перцептрон.....	24
1.3.2. Багатшаровий перцептрон	30
1.3.3. Глибинна нейронна мережа	35
1.3.4. Навчання глибинних нейронних мереж.....	37
1.3.5. Метод навчання “автокодувальник”	38
1.3.6. Навчання глибинних нейронних мереж, базованих на RBM	40
1.3.7. Метод стохастичних градієнтів з ReLU	46
1.3.8. Сучасні парадигми навчання	48
1.4. Історичні тенденції розвитку глибинного навчання	48
1.4.1. Хронологія розвитку глибинних нейронних мережах	48
1.4.2. Підвищення точності, складності та впливу реального світу.....	53
Контрольні питання до розділу 1	56
Розділ 2. Перцептрони та алгоритми навчання перцептрона	57
2.1. Означення та структура перцептрона	57
2.1.1. Лінійні бінарні класифікатори	57
2.1.2. Геометрична інтерпретація процесу навчання перцептрона.....	61
2.1.3. Обмеження процесу навчання перцептронів	62
2.2. Потреба в нелінійності.....	65
2.3. Функція активації перцептронів для прихованого шару	68
2.4. Типи функцій активації для нейронів/перцептрона	69
2.4.1. Лінійна функція активації	69
2.4.2. Бінарна порогова функція активації.....	70
2.4.3. Сигмоїда	71
2.4.4. Функція активації SoftMax.....	72
2.4.5. Функція активації ReLU	74
2.4.6. Функція активації у вигляді гіперболічного тангенса	75
2.5. Правила навчання для багатшарової перцептронної мережі	76
2.5.1. Обчислення градієнта за допомогою методу зворотного поширення помилки	79
2.6. Глибинне навчання і традиційні методи	80
2.7. Методи оптимізації на основі градієнтного спуску	82

2.7.1. Неопуклі функції вартості	84
2.7.2. Сідлові точки багатовимірних функцій вартості	87
Висновки до розділу 2	90
Контрольні питання до розділу 2	90
Розділ 3. Згорткові нейронні мережі	91
3.1. Операція згортки	91
3.1.1. Особливості операції згортки	91
3.1.2. Згортка сигналів у одному вимірі	94
3.2. Аналогові та цифрові сигнали	97
3.3. 2D- і 3D-сигнали	99
3.4. 2D-згортка	100
3.4.1. Двовимірна одинична дельта-функція	100
3.4.2. 2D-згортка сигналу із відгуком LSI-системи на одиничний дельта-імпульс	102
3.4.3. 2D-згортка зображення з імпульсним відгуком різних LSI-систем	104
3.5. Згорткові нейронні мережі	105
3.6. Компоненти згорткових нейронних мереж	106
3.6.1. Шар згортки	107
3.6.2. Шар субдискретизації	108
3.7. Зворотнє поширення помилки через згортковий шар	109
3.8. Пакетна нормалізація	113
Висновки до розділу 3	115
Контрольні питання до розділу 3	116
Розділ 4. Обмежені машини Больцмана та автоматичні шифрувальники	117
4.1. Розподіл Больцмана	117
4.2. Байєсівський підхід: правдоподібність, апіорні та апостеріорні розподіли ймовірностей	119
4.3. Семпсування методами Монте-Карло за схемою марковських ланцюгів	124
4.4. Алгоритм Метрополіса	126
4.5. Обмежені машини Больцмана	127
4.6. Відбір вибірки за Гіббсом	133
4.7. Відбір вибірки за Гіббсом поблоково	134
4.8. Використання семпсування за Гіббсом в обмежених машинах Больцмана	135
4.9. Контрастивна дивергенція	137
4.10. Глибинні мережі довіри	139
4.11. Автокодувальники	141
4.11.1. Навчання ознакам за допомогою автокодувальників	143

4.11.2. Відстань Кульбака–Лейблера	146
4.11.3. Розріджені автокодувальники	146
4.11.4. Шумозаглушувальні автокодувальники	148
4.12. PCA-і ZCA-відбілювання	149
Висновки до розділу 4	152
Контрольні питання до розділу 4	152
Розділ 5. Рекурентні нейронні мережі	153
5.1. Векторна модель	153
5.2. Векторне подання слів	159
5.3. Технологія Word2Vec	160
5.3.1. Метод CBOW	160
5.3.2. Модель векторного подання слів на основі скіп-грам	164
5.4. Вступ до рекурентних нейронних мереж	167
5.4.1. Моделювання мови	170
5.4.2. Прогнозування наступного слова речення традиційними методами і за допомогою RNN	171
Контрольні питання до розділу 5	173
Розділ 6. Варіанти архітектури нейронних мереж	174
6.1. Сегментація зображення	174
6.1.1. Бінарний пороговий метод, оснований на гістограмах інтенсивності пікселів	175
6.1.2. Метод Оцу	176
6.1.3. Сегментація зображення методом вододілів	177
6.1.4. Сегментація зображень методом кластеризації за допомогою K-середніх	178
6.1.5. Семантична сегментація	180
6.1.6. Метод ковзного вікна	180
6.2. Виявлення об'єктів	181
6.2.1. Мережа R-CNN	182
6.2.2. Мережі Fast і Faster R-CNN	184
6.3. Генеративно-змагальні мережі	185
Контрольні питання до розділу 6	187
Список літератури	188
Додаток А. Лінійна алгебра	192
A.1. Скаляри, вектори, матриці та тензори	192
A.2. Множення матриць і векторів	194
A.3. Матриця ідентичності та зворотні матриці	196
A.4. Лінійна залежність і проміжок	197
A.5. Розміри векторів у машинному навчанні – норми	199

А.6. Спеціальні види матриць та векторів	200
А.7. Власний розклад матриці (Eigendecomposition).....	202
А.8. Розклад сингулярного значення	205
А.9. Псевдоінверсія Мура–Пенроуза.....	206
А.10. Оператор сліду матриці. Сума всіх діагональних елементів матриці	206
А.11. Детермінант	207
А.12. Приклад: аналіз основних компонентів.....	207
Додаток Б. Теорія ймовірностей.....	212
Б.1. Чому ймовірність?	212
Б.2. Випадкові змінні	215
Б.3. Ймовірні розподіли	215
Б.3.1. Дискретні змінні та функції ймовірності.....	215
Б.3.2. Функції неперервних змінних та густина ймовірності	216
Б.4. Відособлений розподіл	217
Б.5. Умовна ймовірність	218
Б.6. Ланцюгове правило умовних ймовірностей	218
Б.7. Незалежність та умовна незалежність.....	218
Б.8. Математичне сподівання, дисперсія та коваріація.....	219
Б.9. Загальні розподіли ймовірностей	220
Б.9.1. Розподіл Бернуллі	221
Б.9.2. Узагальнений розподіл Бернуллі (англ. “Multinoulli Distribution” – розподіл мультинуллі)	221
Б.9.3. Нормальний розподіл (розподіл Гаусса).....	222
Б.9.4. Експоненційний розподіл і розподіл Лапласа	223
Б.9.5. Розподіл Дірака й емпіричний розподіл.....	224
Б.9.6. Суміші розподілів	225
Б.10. Властивості функцій.....	226
Б.11. Правило Байеса	228
Б.12. Технічні деталі неперервних величин	229
Б.13. Теорія інформації.....	231
Б.14. Структуровані ймовірнісні моделі.....	234
Додаток В. Числові методи.....	238
В.1. Переповнення і антипереповнення	238
В.2. Небажана обумовленість.....	240
В.3. Оптимізація градієнтним методом.....	240
В.3.1. Не тільки градієнт: матриці Якобі й Гессе	244
В.4. Оптимізація з обмеженнями	251
В.5. Приклад: лінійний метод найменших квадратів	253
Приклади тестових завдань.....	255