
Spis treści

Słowo wstępne	11
Przedmowa	13
1. Inteligencja zbiorowa — wprowadzenie	21
Czym jest inteligencja zbiorowa?	22
Czym jest uczenie maszynowe?	23
Ograniczenia uczenia maszynowego	24
Rzeczywiste przykłady	24
Inne zastosowania algorytmów uczących	25
2. Tworzenie rekomendacji	27
Filtrowanie grupowe	27
Gromadzenie preferencji	28
Znajdowanie podobnych użytkowników	29
Rekomendowanie pozycji	34
Dopasowywanie produktów	36
Tworzenie systemu rekomendowania odnośników del.icio.us	38
Filtrowanie oparte na pozycjach	42
Zastosowanie zbioru danych MovieLens	45
Filtrowanie oparte na użytkownikach czy pozycjach?	46
Ćwiczenia	47
3. Wykrywanie grup	49
Porównanie uczenia nadzorowanego z nienadzorowanym	49
Wektory wyrazów	50
Grupowanie hierarchiczne	53
Rysowanie dendrogramu	57
Grupowanie kolumn	59
Grupowanie k-średnich	61

Klasyfikacja preferencji	64
Wyświetlanie danych w dwóch wymiarach	68
Inne rzeczy, które mogą być grupowane	71
Ćwiczenia	72
4. Wyszukiwanie i klasyfikowanie	73
Co znajduje się w wyszukiwarce?	73
Prosty przeszukiwacz	75
Budowanie indeksu	77
Odpytywanie	81
Klasyfikacja oparta na treści	83
Użycie odnośników zewnętrznych	87
Uczenie na podstawie kliknięć	91
Ćwiczenia	101
5. Optymalizacja	103
Podróż grupy osób	104
Reprezentowanie rozwiązań	105
Funkcja kosztu	106
Wyszukiwanie losowe	108
Metoda największego wzrostu	109
Symulowane wyżarzanie	111
Algorytmy genetyczne	113
Wyszukiwania rzeczywistych lotów	117
Optymalizowanie pod kątem preferencji	122
Wizualizacja sieci	125
Inne możliwości	130
Ćwiczenia	130
6. Filtrowanie dokumentów	133
Filtrowanie spamu	133
Dokumenty i wyrazy	134
Trenowanie klasyfikatora	135
Obliczanie prawdopodobieństw	137
Naiwny klasyfikator	139
Metoda Fishera	142
Utrwalanie klasyfikatorów po przeprowadzonym treningu	146
Filtrowanie kanałów informacyjnych blogów	148
Poprawianie wykrywania właściwości	150
Użycie interfejsu Akismet	152
Alternatywne metody	153
Ćwiczenia	154

7. Modelowanie przy użyciu drzew decyzyjnych	157
Przewidywanie rejestracji	157
Wprowadzenie do drzew decyzyjnych	159
Uczenie drzewa	160
Wybór najlepszego podziału	162
Budowanie drzewa rekurencyjnego	164
Wyświetlanie drzewa	166
Klasyfikowanie nowych obserwacji	168
Przycinanie drzewa	169
Radzenie sobie z brakującymi danymi	171
Radzenie sobie z wynikami liczbowymi	172
Modelowanie cen domów	173
Modelowanie „atrakcyjności”	176
Kiedy stosować drzewa decyzyjne?	178
Ćwiczenia	179
8. Budowanie modelu cen	181
Budowanie przykładowego zbioru danych	181
Metoda k-najbliższych sąsiadów	183
Sąsiednie elementy z określoną wagą	186
Walidacja krzyżowa	189
Zmienne heterogeniczne	191
Optymalizowanie skali	194
Rozkłady niejednolite	196
Użycie rzeczywistych danych — interfejs API serwisu eBay	200
Kiedy używać metody k-najbliższych sąsiadów?	207
Ćwiczenia	207
9. Zaawansowane klasyfikowanie: metody jądrowe i maszyny wektorów nośnych	209
Zbiór danych swatki	209
Trudności związane z danymi	211
Podstawowa klasyfikacja liniowa	213
Właściwości skategoryzowane	217
Skalowanie danych	218
Metody jądrowe	220
Maszyny wektorów nośnych	223
Zastosowanie biblioteki LIBSVM	225
Dopasowywanie w serwisie Facebook	227
Ćwiczenia	232

10. Znajdowanie niezależnych właściwości	233
Zbiór artykułów	234
Wcześniejsze rozwiązania	237
Nieujemna faktoryzacja macierzy	240
Wyświetlanie wyników	246
Użycie danych rynku giełdowego	249
Ćwiczenia	254
11. Inteligencja rozwojowa	255
Czym jest programowanie genetyczne?	255
Programy w postaci drzew	258
Tworzenie populacji początkowej	261
Testowanie rozwiązania	263
Krzyżowanie	267
Budowanie środowiska	269
Prosta gra	272
Dalsze możliwości	276
Ćwiczenia	278
12. Algorytmy — podsumowanie	281
Klasyfikator bayesowski	281
Klasyfikator drzew decyzyjnych	285
Sieci neuronowe	288
Maszyny wektorów nośnych	292
Metoda k-najbliższych sąsiadów	296
Grupowanie	299
Skalowanie wielowymiarowe	303
Nieujemna faktoryzacja macierzy	305
Optymalizacja	307
A Zewnętrzne biblioteki	311
Universal Feed Parser	311
Python Imaging Library	311
Beautiful Soup	312
pysqlite	313
NumPy	314
matplotlib	315
pydelicious	316

B Formuły matematyczne317

Odległość euklidesowa 317

Współczynnik korelacji Pearsona 317

Średnia ważona 318

Współczynnik Tanimoto 319

Prawdopodobieństwo warunkowe 319

Niejednorodność Giniego 320

Entropia 321

Wariancja 321

Funkcja Gaussa 322

Iloczyny skalarne 322

Skorowidz324