

**Podstawy matematyki w data science : algebra liniowa, rachunek
prawdopodobieństwa i statystyka / Thomas Nield. – Gliwice, © 2023**

Spis treści

Przedmowa	7
1. Podstawy matematyki oraz rachunku różniczkowego i całkowego	11
Teoria liczb	11
Kolejność działań	13
Zmienne	14
Funkcje	15
Sumowanie	19
Potęgowanie	20
Logarytmy	23
Liczba Eulera i logarytmy naturalne	24
Liczba Eulera	24
Logarytmy naturalne	28
Granice	28
Pochodne	30
Pochodne cząstkowe	33
Reguła łańcuchowa	35
Całki	37
Podsumowanie	41
Ćwiczenia	42
2. Prawdopodobieństwo	43
Zrozumieć prawdopodobieństwo	43
Prawdopodobieństwo a statystyka	45
Matematyka prawdopodobieństw	46
Prawdopodobieństwa łączne	46
Prawdopodobieństwa alternatywne	47
Prawdopodobieństwo warunkowe i twierdzenie Bayesa	48
Łączne i alternatywne prawdopodobieństwa warunkowe	51
Rozkład dwumianowy	52
Rozkład beta	54
Podsumowanie	60
Ćwiczenia	60
3. Statystyka opisowa i wnioskowanie statystyczne	62
Czym są dane?	62
Statystyka opisowa a wnioskowanie statystyczne	64
Populacje, próby i obciążenie	64

Statystyka opisowa	68
Średnia i średnia ważona	68
Mediana	69
Dominanta	71
Wariancja i odchylenie standardowe	71
Rozkład normalny	75
Dystrybuanta odwrotna	81
Standaryzacja Z	82
Wnioskowanie statystyczne	84
Centralne twierdzenie graniczne	84
Przedziały ufności	87
Wartości p	89
Testowanie hipotez	90
Rozkład t: analizowanie małych prób	97
Big data i błąd teksańskiego snajpera	98
Podsumowanie	99
Ćwiczenia	100
4. Algebra liniowa	101
Co to jest wektor?	101
Dodawanie i łączenie wektorów	105
Skalowanie wektorów	106
Powłoka i zależność liniowa	108
Przekształcenia liniowe	110
Wektory bazowe	110
Mnożenie macierzy przez wektor	113
Mnożenie macierzy	117
Wyznaczniki	119
Specjalne rodzaje macierzy	122
Macierz kwadratowa	122
Macierz jednostkowa	122
Macierz odwrotna	122
Macierz diagonalna	123
Macierz trójkątna	123
Macierz rzadka	123
Układy równań i macierze odwrotne	124
Wektory i wartości własne	127
Podsumowanie	129
Ćwiczenia	130
5. Regresja liniowa	131
Podstawowa regresja liniowa	132
Reszty i kwadraty błędu	136
Znajdowanie najlepiej dopasowanej linii	138
Równanie w formie zamkniętej	139
Techniki wykorzystujące macierze odwrotne	140

Metoda gradientu prostego	142
Nadmierne dopasowanie i wariancja	148
Metoda stochastycznego gradientu prostego	149
Współczynnik korelacji	151
Istotność statystyczna	153
Współczynnik determinacji	157
Błąd standardowy estymacji	158
Przedziały przewidywania	159
Podział danych na treningowe i testowe	162
Wielokrotna regresja liniowa	167
Podsumowanie	168
Ćwiczenia	168
6. Regresja logistyczna i klasyfikacja	169
Na czym polega regresja logistyczna?	169
Przeprowadzanie regresji logistycznej	172
Funkcja logistyczna	172
Dopasowywanie krzywej logistycznej	174
Regresja logistyczna z wieloma zmiennymi	179
Logarytm szansy	182
R-kwadrat	185
Wartości p	189
Podziały na dane treningowe i testowe	191
Macierz błędów	192
Twierdzenie Bayesa a klasyfikacja	195
Krzywa ROC/pole pod krzywą	196
Nierównowaga klas	197
Podsumowanie	198
Ćwiczenia	198
7. Sieci neuronowe	200
Kiedy używać sieci neuronowych i uczenia głębokiego?	200
Prosta sieć neuronowa	201
Funkcje aktywacji	204
Propagacja w przód	208
Propagacja wsteczna	213
Obliczanie pochodnych względem wag i biasów	213
Metoda gradientu stochastycznego	217
Używanie scikit-learn	220
Ograniczenia sieci neuronowych i uczenia maszynowego	221
Podsumowanie	224
Ćwiczenie	225
8. Porady zawodowe i droga naprzód	226
Nowa definicja data science	227
Krótką historia data science	229

Szukanie przewagi	231
Biegłość w SQL-u	231
Biegłość w programowaniu	233
Wizualizacja danych	236
Znajomość branży	238
Produktywna nauka	239
Praktyk czy doradca?	240
Na co trzeba uważać w pracy związanej z data science?	242
Definicja roli	242
Skupienie organizacyjne i akceptacja	243
Adekwatne zasoby	244
Rozsądne cele	245
Konkurowanie z istniejącymi systemami	246
Twoja rola nie jest tym, czego się spodziewałeś	248
Czy Twoja praca marzeń nie istnieje?	249
Co dalej?	249
Podsumowanie	250
A. Tematy dodatkowe	251
B. Odpowiedzi do ćwiczeń	269
Skorowidz	281