

---

<b>Wykaz najważniejszych skrótów</b> .....	<b>8</b>
<b>Przedmowa</b> .....	<b>10</b>
<b>1. Podstawowe pojęcia data mining</b> .....	<b>11</b>
1.1. Wprowadzenie.....	12
1.2. Podstawowe zadania eksploracji danych .....	13
1.3. Główne etapy eksploracji danych .....	15
1.4. Normalizacja i standaryzacja danych .....	17
1.5. Problem obserwacji odstających.....	18
1.6. Problem jakości danych.....	21
1.7. Podstawowa terminologia i oznaczenia.....	22
<b>2. Metody i algorytmy optymalizacji</b> .....	<b>25</b>
2.1. Podstawowe pojęcia optymalizacji gradientowej .....	26
2.2. Toolbox optymalizacyjny Matlab'a .....	30
2.3. Przykłady zastosowania optymalizacji gradientowej.....	33
2.3.1. Optymalizacja minimaksowa .....	33
2.3.2. Programowanie kwadratowe .....	35
2.4. Metoda symulowanego wyżarzania .....	40
2.5. Algorytm genetyczny .....	43
2.6. Algorytmy ewolucyjne.....	48
2.6.1. Strategie ewolucyjne.....	48
2.6.2. Programowanie ewolucyjne .....	50
2.7. Toolbox globalnej optymalizacji w Matlabie .....	51
<b>3. Regresja liniowa</b> .....	<b>55</b>
3.1. Wprowadzenie.....	56
3.2. Ogólny model liniowy regresji .....	56
3.3. Zastosowanie dekompozycji SVD w regresji liniowej.....	59
3.4. Implementacja regresji liniowej w Matlabie .....	64
3.5. Zastosowanie regresji liniowej w klasyfikacji .....	65
<b>4. Regresja logistyczna</b> .....	<b>69</b>
4.1. Wprowadzenie.....	70
4.2. Podstawowe definicje.....	71
4.3. Algorytm określania wartości parametrów w regresji logistycznej.....	74
4.3.1. Wprowadzenie .....	74
4.3.2. Model binarny regresji logistycznej .....	75
4.3.3. Przypadek trzech klas uporządkowanych .....	82
4.3.4. Przypadek trzech klas nieuporządkowanych .....	83

<b>5. Klasyfikatory Bayesa</b> .....	<b>85</b>
5.1. Wprowadzenie.....	86
5.2. Pełny klasyfikator Bayesa.....	86
5.3. Naiwny klasyfikator Bayesa.....	90
5.4. Implementacja naiwnego klasyfikatora Bayesa w Matlabie.....	93
<b>6. Drzewa decyzyjne</b> .....	<b>97</b>
6.1. Wprowadzenie.....	98
6.2. Struktura drzewa decyzyjnego.....	98
6.3. Algorytm tworzenia drzewa decyzyjnego.....	100
6.4. Implementacja modelu drzewa decyzyjnego w Matlabie.....	106
6.5. Losowy las.....	111
6.5.1. Opis metody.....	111
6.5.2. Implementacja lasu losowego w Matlabie.....	113
<b>7. Sztuczne sieci neuronowe MLP i RBF</b> .....	<b>117</b>
7.1. Sieć perceptronu wielowarstwowego MLP.....	118
7.1.1. Struktura sieci.....	118
7.1.2. Algorytmy uczące sieci MLP.....	120
7.1.3. Program komputerowy do uczenia sieci MLP.....	128
7.1.4. Przykłady użycia sieci MLP.....	129
7.2. Sieć radialna RBF.....	134
7.2.1. Struktura sieci RBF.....	134
7.2.2. Algorytmy uczące sieci RBF.....	135
7.2.3. Program komputerowy do uczenia sieci radialnej.....	141
7.2.4. Przykład zastosowania sieci RBF do aproksymacji.....	142
7.3. Zdolności generalizacyjne sieci neuronowych.....	143
7.3.1. Warunki dobrej generalizacji sieci.....	144
7.3.2. Metody zwiększania zdolności generalizacyjnych sieci.....	145
<b>8. Sieć wektorów nośnych SVM</b> .....	<b>149</b>
8.1. Wprowadzenie.....	150
8.2. Sieć liniowa SVM w zadaniu klasyfikacji.....	150
8.3. Sieć nieliniowa SVM w zadaniu klasyfikacji.....	155
8.3.1. Interpretacja mnożników Lagrange'a w rozwiązaniu sieci.....	161
8.3.2. Problem klasyfikacji przy wielu klasach.....	162
8.4. Sieci SVM do zadań regresji.....	163
8.5. Sieć jedнокlasowa SVM.....	165
8.6. Przegląd algorytmów rozwiązania zadania dualnego.....	168
8.7. Program komputerowy uczenia sieci SVM.....	170
8.8. Przykład zastosowania sieci SVM.....	173
8.9. Porównanie sieci SVM z innymi rozwiązaniami neuronowymi.....	175

---

<b>9. Zespoły klasyfikatorów i predyktorów</b> .....	<b>179</b>
9.1. Wprowadzenie.....	180
9.2. Zespół klasyfikatorów.....	181
9.2.1. Integracja większościowa zespołu.....	181
9.2.2. Zastosowanie dekompozycji PCA w integracji.....	185
9.2.3. Zastosowanie naiwnej reguły Bayesa w integracji.....	186
9.2.4. Metoda integracji Kullbacka–Leiblera.....	188
9.3. Zespół predyktorów.....	191
9.3.1. Integracja poprzez uśrednianie wyników.....	191
9.3.2. Integracja wykorzystująca PCA.....	193
9.3.3. Integracja przy zastosowaniu ICA.....	193
9.3.4. Integracja zespołu przy zastosowaniu sieci neuronowej.....	195
9.4. Przykład użycia zespołu w prognozowaniu obciążenia w PSE.....	196
<b>10. Metody oceny jakości klasyfikatorów i predyktorów</b> .....	<b>199</b>
10.1. Miary jakości predykcji.....	200
10.2. Badanie jakości rozwiązań w zadaniach klasyfikacji.....	203
10.2.1. Miary jakości klasyfikatora.....	204
10.2.2. Charakterystyka ROC.....	206
10.3. Metody poprawy jakości klasyfikatora.....	209
10.3.1. Metoda różnicowania kosztu błędnej klasyfikacji.....	210
10.3.2. Metody równoważenia klas.....	211
10.3.3. Problemy klasyfikacji wieloklasowej.....	212
10.4. Obiektywna ocena zdolności generalizacyjnych systemu predykcyjnego i klasyfikacyjnego.....	213
<b>11. Transformacje i metody redukcji wymiaru danych</b> .....	<b>215</b>
11.1. Kryteria doboru rzędu zredukowanego modelu.....	216
11.2. Transformacje bazujące na PCA.....	217
11.2.1. Istota przekształcenia PCA.....	218
11.2.2. Implementacja przekształcenia PCA w Matlabie.....	222
11.2.3. Wykorzystanie PCA w kompresji.....	223
11.2.4. PCA w zastosowaniu do ilustracji rozkładu danych wielowymiarowych.....	225
11.3. Nieliniowe przekształcenie PCA.....	227
11.4. Przekształcenie LDA Fishera.....	230
11.5. Ślepa separacja sygnałów.....	232
11.5.1. Sformułowanie problemu ślepej separacji.....	232
11.5.2. Algorytmy bazujące na statystykach drugiego rzędu.....	235
11.5.3. Metody bazujące na statystykach wyższego rzędu.....	236
11.5.4. Toolbox ICALAB.....	238
11.6. Rzutowanie Sammona.....	243

11.7.	Transformacja SNE .....	245
11.8.	Toolbox do redukcji danych w Matlabie.....	248
<b>12.</b>	<b>Wybrane metody generacji i selekcji cech diagnostycznych.....</b>	<b>251</b>
12.1.	Wprowadzenie.....	252
12.2.	Metody generacji cech diagnostycznych .....	253
12.3.	Metody selekcji cech diagnostycznych .....	256
12.3.1.	Metoda dyskryminacji Fishera.....	257
12.3.2.	Metoda korelacji danych z klasą .....	258
12.3.3.	Zastosowanie jednowejściowej sieci SVM .....	258
12.3.4.	Wykorzystanie wielowejściowej liniowej sieci SVM .....	259
12.3.5.	Zastosowanie nieliniowej funkcji jądra .....	260
12.3.6.	Selekcja cech bazująca na liniowej regresji krokowej .....	262
12.3.7.	Selekcja przy zastosowaniu algorytmu genetycznego .....	264
12.3.8.	Zastosowanie testu statystycznego Kołmogorowa–Smirnowa.....	266
12.3.9.	Użycie testu Wilcoxon–Manna–Whitneya.....	267
12.3.10.	Selekcja przy zastosowaniu transformacji PCA.....	268
12.3.11.	Selekcja przy wykorzystaniu transformacji ICA.....	269
12.3.12.	Selekcja przy zastosowaniu transformacji LDA .....	270
<b>13.</b>	<b>Metody grupowania danych.....</b>	<b>275</b>
13.1.	Wprowadzenie.....	276
13.2.	Miary odległości między wektorami .....	277
13.3.	Miary odległości między klastrami .....	278
13.4.	Algorytm <i>K-means</i> grupowania .....	279
13.5.	Algorytm hierarchiczny grupowania .....	285
13.6.	Algorytmy rozmyte grupowania .....	288
13.6.1.	Algorytm grupowania górskiego.....	289
13.6.2.	Algorytm <i>c-means</i> .....	291
13.6.3.	Algorytm Gustafsona–Kessela .....	294
13.7.	Miary jakości grupowania danych.....	300
13.7.1.	Miary jakości grupowania danych nieprzypisanych do klas.....	300
13.7.2.	Miary jakości grupowania danych przypisanych do klas .....	305
<b>14.</b>	<b>Analiza zależności asocjacyjnych między danymi .....</b>	<b>311</b>
14.1.	Podstawowe definicje problemu .....	313
14.2.	Generacja zbiorów najczęściej pojawiających się w transakcjach.....	315
14.3.	Generacja reguł asocjacyjnych .....	318
14.4.	Metody oceny reguł asocjacyjnych.....	320
14.4.1.	Miary symetryczne.....	320
14.4.2.	Miary niesymetryczne .....	323

---

<b>15. Metody wizualizacji danych</b> .....	<b>327</b>
15.1. Wprowadzenie.....	328
15.2. Narzędzia wizualizacji graficznej w Matlabie .....	330
15.2.1. Podstawowe funkcje graficzne .....	330
15.2.2. Zaawansowane funkcje graficzne .....	332
15.2.3. Grafika trójwymiarowa .....	336
15.2.4. Wykresy wektorowe .....	341
15.2.5. Rodzina funkcji <i>ezplot</i> .....	342
15.2.6. Tworzenie animacji .....	343
15.2.7. Opisy rysunków.....	344
15.3. Wizualizacja danych wielowymiarowych .....	345
<b>16. Przykłady zastosowań metod eksploracji danych</b> .....	<b>349</b>
16.1. Analiza macierzy ekspresji genów w chorobie nowotworowej .....	350
16.1.1. Wprowadzenie .....	350
16.1.2. Metoda eksploracji danych.....	351
16.1.3. Analiza mikromacierzy ekspresji genów w białaczce.....	352
16.2. Przykład eksploracji danych przy rozpoznaniu komórek krwiotwórczych .....	358
16.2.1. Charakterystyka ogólna komórek krwiotwórczych .....	358
16.2.2. Baza danych komórek w eksperymencie numerycznym.....	360
16.2.3. Selekcja zbioru najważniejszych cech diagnostycznych .....	361
16.2.4. Klasyfikacja końcowa komórek przy użyciu zespołu klasyfikatorów .....	366
<b>Dodatek A</b> .....	<b>367</b>
A.1. Ortogonalność wektorów i macierzy .....	367
A.2. Normy i wskaźniki macierzowe .....	367
A.3. Transformacje i dekompozycje macierzowe .....	370
A.3.1. Dekompozycja <i>LU</i> .....	370
A.3.2. Dekompozycja macierzy według wartości własnych.....	371
A.3.3. Dekompozycja macierzy według wartości osobliwych (SVD).....	372
A.3.4. Dekompozycja <i>QR</i> .....	373
A.3.5. Dekompozycja Schura.....	374
A.3.6. Transformacja Hessenberga .....	374
A.3.7. Transformacja Householdera .....	375
A.3.8. Transformacja Givensa.....	375
A.4. Zaawansowane operacje macierzowe .....	376
A.4.1. Reguły inwersji macierzy złożonych.....	376
A.4.2. Różniczkowanie funkcji względem wektora .....	377
A.4.3. Różniczkowanie macierzy .....	378
<b>Skorowidz</b> .....	<b>379</b>
<b>Literatura</b> .....	<b>385</b>