

# Spis treści

<b>Przedmowa</b>	<b>7</b>
<b>1 Analiza składowych głównych</b>	<b>9</b>
1.1. Przykładowa analiza . . . . .	10
1.2. Regresja składowych głównych (PCR) i regresja częściowych najmniejszych kwadratów (PLSR) . . . . .	17
1.3. Zadania . . . . .	22
<b>2 LDA i QDA</b>	<b>25</b>
2.1. Dwie klasy, dwie zmienne . . . . .	29
2.2. Trzy klasy, dwie zmienne . . . . .	36
2.3. Trzy klasy, trzy zmienne . . . . .	42
2.4. Wiele klas, wiele zmiennych . . . . .	44
2.5. Zadania . . . . .	46
<b>3 Dyskryminacja logistyczna</b>	<b>51</b>
3.1. Zadania . . . . .	60
<b>4 Ocena jakości klasyfikatorów. Krzywa ROC i parametr AUC</b>	<b>63</b>
4.1. Ocena jakości reguły klasyfikacyjnej . . . . .	63
4.2. Metody estymacji prawdopodobieństwa błędnej klasyfikacji . . . . .	64
4.2.1. Krosvalidacja $n$ -krotna (leave-one-out, one-leaving-out) . . . . .	64
4.2.2. Krosvalidacja 10-krotna . . . . .	66
4.2.3. Metoda oceny błędu na podstawie wielokrotnego podziału próby . . . . .	68
4.3. Krzywe ROC . . . . .	69

4.3.1. Binormalna krzywa ROC . . . . .	72
<b>5 Estymacja gęstości</b>	<b>77</b>
5.1. Wprowadzenie . . . . .	77
5.2. Estymacja gęstości jednowymiarowej . . . . .	79
5.3. Klasyfikacja na podstawie gęstości jednowymiarowej . . . . .	83
5.4. Estymacja gęstości dwuwymiarowej . . . . .	84
5.5. Klasyfikacja na podstawie gęstości dwuwymiarowej . . . . .	85
5.6. Metoda typu „najbliższy sąsiad” ( $kNN$ ) . . . . .	88
5.7. Zadania . . . . .	90
<b>6 Drzewa klasyfikacyjne i regresyjne</b>	<b>93</b>
6.1. Drzewa regresyjne . . . . .	93
6.2. Drzewa klasyfikacyjne . . . . .	94
6.3. Drzewa regresyjne: dane fitness . . . . .	95
6.4. Drzewa klasyfikacyjne: dane Cars93 . . . . .	105
6.5. Redukcja zmiennych w przypadku drzew klasyfikacyjnych . . . . .	110
6.6. Zadania . . . . .	114
<b>7 Metody nieparametrycznej estymacji funkcji regresji</b>	<b>117</b>
7.1. Nieliniowe parametryczne metody regresyjne . . . . .	117
7.2. Nieparametryczne metody regresyjne . . . . .	118
7.2.1. Metody nieparametryczne - niski wymiar wektora atrybutów	118
7.2.2. Metody nieparametryczne - wysoki wymiar wektora atrybutów	124
7.3. Zadania . . . . .	132
<b>8 Metody łączenia klasyfikatorów: bagging, boosting i lasy losowe</b>	<b>137</b>
8.1. Klasyfikator AdaBoost dla drzewa . . . . .	140
8.2. Klasyfikator AdaBoost dla LDA . . . . .	142
8.3. Zadania . . . . .	147
<b>9 Analiza skupień i skalowanie wielowymiarowe</b>	<b>149</b>
9.1. Zadania . . . . .	156
<b>10 Przykład analizy: dane Image Segmentation</b>	<b>159</b>

10.1. Wczytanie danych i wybranie atrybutów . . . . .	159
10.2. PCA . . . . .	162
10.3. LDA . . . . .	166
10.4. QDA . . . . .	167
10.5. Drzewa klasyfikacyjne . . . . .	169
10.5.1. LDA z pominięciem kilku klas . . . . .	175
10.6. kNN . . . . .	175
10.7. Klasyfikacja na podstawie składowych głównych . . . . .	179
<b>11 Dodatek</b>	<b>183</b>
11.1. Używane zbiory danych . . . . .	183