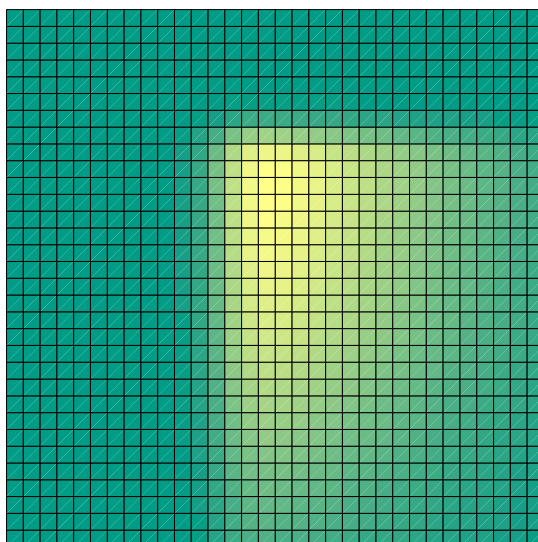


N o r b e r t J a n k o w s k i

Ontogeniczne sieci neuronowe

O sieciach zmieniających swoją
strukturę



Warszawa 2003

Opracowanie książki było wspierane stypendium
Uniwersytetu Mikołaja Kopernika

Spis treści

Wprowadzenie	15
1 Funkcje transferu	21
1.1 Funkcje realizowane przez neuron	24
1.2 Funkcje aktywacji	28
1.2.1 Miary odległości i podobieństwa jako funkcje aktywacji	30
1.2.1.1 Jednorodne miary odległości	31
1.2.1.2 Niejednorodne miary odległości	33
1.2.2 Funkcje aktywacji powstające jako złożenie iloczynu skalarnego i miar podobieństwa	36
1.3 Funkcje wyjścia	37
1.3.1 Funkcje sigmoidalne	39
1.3.2 Funkcje zlokalizowane wokół jednego centrum	40
1.3.3 Funkcje semi-centralne	49
1.4 Funkcje transferu	49
1.4.1 Nielocalne funkcje transferu	50
1.4.2 Lokalne i semi-lokalne funkcje transferu	51
1.4.3 Gaussowska i sigmoidalna funkcja wstępowa	54
1.4.4 Funkcje o gęstościach elipsoidalnych	57
1.4.5 Uniwersalne funkcje transferu	59
1.4.6 Funkcje bicentralne	66
1.4.7 Rozszerzenia funkcji bicentralnych	71
1.4.7.1 Funkcje bicentralne z niezależnymi skosami	71
1.4.7.2 Funkcje bicentralne z rotacją	71
1.4.7.3 Funkcje bicentralne z rotacją i niezależnymi skosami	74
1.4.8 Hierarchia funkcji transferu pod względem ich elastyczności	76
1.4.9 Końcowe porównanie różnych funkcji transferu	77
2 Sieci z radialnymi funkcjami bazowymi (RBF)	81
2.1 Sieci z radialnymi funkcjami bazowymi i regularyzacją	81
2.2 Uogólniona sieć z radialnymi funkcjami bazowymi (GRBF)	85
2.3 Metody inicjalizacji i uczenia bez nadzoru sieci typu RBF	87
2.3.1 Inicjalizacja położenia zbiorem wektorów uczących	87

2.3.2	Inicjalizacja położenia poprzez podzbiór zbioru uczącego . . .	88
2.3.3	Inicjalizacja położenia metodą klasteryzacji k -średnich	88
2.3.4	Inicjalizacja za pomocą metody k najbliższych sąsiadów . . .	89
2.3.5	Konstruowanie klastrów za pomocą dendrogramów	90
2.3.6	Inicjalizacja za pomocą histogramów i drzew decyzyjnych	91
2.4	Uczenie z nadzorem sieci RBF	94
2.5	Rozszerzenia sieci RBF	96
2.5.1	Rozszerzenia głównego równania sieci RBF	96
2.5.2	Regularyzacja	97
2.5.3	Inne metody uczenia sieci RBF	99
2.6	Porównanie sieci RBF z sieciami MLP	100
2.7	Probabilistyczne sieci neuronowe	103
3	Support Vector Machines (SVM)	107
3.1	Funkcje jądrowe	108
3.2	Konstrukcja optymalnej hiperpłaszczyzny	110
3.3	Konstrukcja hiperpłaszczyzny dla przypadków nieseparowalnych (C-SVC)	112
3.4	ν -SVC	114
3.5	Problem regresji (ϵ -SVR)	117
3.6	Problem regresji dla ν -SVM (ν -SVR)	118
3.7	Optymalizacja problemów programowania kwadratowego (QP)	119
3.7.1	Dekompozycja	119
3.7.2	Wybór zbioru roboczego dla C-SVM	121
3.7.3	Kryterium stopu	122
3.7.4	Wybór zbioru roboczego dla ν -SVM	122
3.7.5	Kryterium stopu dla ν -SVM	122
3.7.6	Analityczne rozwiązanie problemu dekompozycji	123
3.7.7	Wyznaczenie wartości b i ρ	124
3.7.8	Dalsze sposoby przyspieszenia rozwiązywania problemów QP dla SVM	124
3.8	Zbieżność algorytmów dekompozycji QP	125
3.9	SVM a RBF	125
3.10	Meta-SVM	125
3.10.1	Walidacja skośna stosowana do uczenia	126
3.10.2	Wyniki algorytmu Meta-SVM	128
3.10.3	Podsumowanie	131
4	Ontogeniczne modele sieci neuronowych	133
4.1	Modele zmniejszające strukturę	137
4.1.1	Modele zmniejszające strukturę a regularyzacja	138
4.1.1.1	Rozpad wag	138
4.1.1.2	Eliminacja wag	139
4.1.1.3	MLP2LN	139
4.1.1.4	Lokalna regresja grzbietowa	140

4.1.1.5	Metody współdzielenia wag	140
4.1.2	Usuwanie wag metodami Optimal Brain Damage (OBD) i Optimal Brain Surgeon (OBS)	140
4.1.3	Statystyczne i inne metody zmniejszania struktury sieci neuronowych	143
4.2	Modele o strukturach rozrastających się	145
4.2.1	Algorytm kafelkowania	145
4.2.1.0.1	Algorytm kieszonkowy	146
4.2.1.0.2	Sieć kafelkowa dla problemów wieloklasowych	147
4.2.2	Algorytm wieża i piramida	147
4.2.3	Upstart	148
4.2.4	Algorytm budowania sieci kaskadowych przez analizę dychotomii	149
4.2.5	Algorytm korelacji kaskadowej	152
4.2.6	Kaskadowa sieć perceptronowa	153
4.2.7	Feature Space Mapping (FSM)	153
4.2.8	Sieć RAN z przydziałem zasobów	155
4.2.8.1	Uczenie sekwencyjne	155
4.2.8.2	Geometryczne Kryterium Rozrostu	157
4.2.8.3	Adaptacja sieci RAN	160
4.3	Sieć IncNet ze statystyczną kontrolą złożoności sieci	160
4.3.1	Struktura sieci i funkcje transferu	162
4.3.2	Rozszerzony filtr Kalmana	164
4.3.3	Szybka wersja rozszerzonego filtru Kalmana	166
4.3.4	Kryterium wystarczalności modelu	167
4.3.5	Usuwanie neuronów	169
4.3.6	Łączenie neuronów	171
4.3.7	Wykorzystanie sieci IncNet w klasyfikacji	173
4.3.8	Charakterystyka parametrów kontroli procesu adaptacji sieci IncNet	175
4.4	Sieć neuronowa optymalnych funkcji transferu	177
4.4.1	Sieć optymalnych funkcji transferu (OTFN)	178
4.4.1.1	Usuwanie neuronów	178
4.4.1.2	Statystyczne kryterium usuwania neuronów	179
4.4.1.3	Kryterium wystarczalności sieci	179
4.4.2	Sieć optymalnych funkcji transferu typu II	180
4.4.3	Przykłady działania sieci optymalnych funkcji transferu	182
4.4.3.1	Problem parzystości	182
4.4.3.2	Problem półsfery i półprzestrzeni	183
4.4.3.3	Problem trójkąta	184

5	Komitety modeli	185
5.1	<i>K</i> -klasyfikatorów	187
5.2	K^2 -klasyfikatorów	188
5.3	Maszyna liniowa	189
5.4	Sposoby podejmowania decyzji przez komitet	190
5.5	<u>Bootstrap Aggregating</u> (Bagging)	191
5.6	Boosting i AdaBoost	192
5.7	Inne komitety	194
5.7.1	Arcing	195
5.7.2	RegionBoost	195
5.7.3	Stacking	195
5.7.4	Grading	197
5.7.5	Mixture of local experts	197
5.8	Komitety heterogeniczne	199
5.9	Komitety z lokalną kompetencją	199
6	Wstępne i końcowe przetwarzanie danych	201
6.1	Transformacje danych	201
6.2	Wartości nietypowe i brakujące	203
6.2.1	Wartości nietypowe	203
6.2.2	Wartości brakujące	204
6.3	Metody selekcji i ważenia cech	205
6.3.1	Ważenie i selekcja cech dyskretną metodą quasi-gradientową	206
6.3.1.1	Algorytm ważenia cech	207
6.3.1.1.1	Uczenie z wykorzystaniem walidacji skrojonej	207
6.3.1.1.2	Estymacja wag końcowych	208
6.3.1.1.3	Procedura <i>FindWeights</i>	209
6.3.1.2	Eliminacja cech	210
6.3.1.3	Przykłady rezultatów dla ważenia cech	210
6.3.1.3.1	Baza danych tarczycy	212
6.3.1.3.2	Dane wyrostka robaczkowego (<i>Appendicitis</i>)	213
6.3.1.3.3	Dane <i>Australian credit</i>	213
6.3.1.3.4	Dane opisujące flagi narodowe (<i>Flags</i>)	213
6.3.1.3.5	Dane raka piersi	218
6.3.1.3.6	Zbiór danych <i>Glass</i>	218
6.3.1.3.7	Dane chorób serca	218
6.3.1.3.8	Dane opisujące gatunki win	219
6.3.1.4	Podsumowanie	219
6.4	Regularyzacja danych	219
6.4.1	Odcienie szarości	222
6.4.2	Eliminacja złych wektorów i przeetykietowanie klas	222
6.4.3	Przykłady użycia regularyzacji danych	223
6.4.4	Podsumowanie	223

6.5	Przedziały ufności jako narzędzie analizy danych i wizualizacji wyników	226
6.5.1	Przedziały ufności i probabilistyczne przedziały ufności a reguły logiczne	230
7	Zastosowanie sieci neuronowych	235
7.1	Techniki porównywania różnych modeli	235
7.2	Medyczne zastosowania sieci IncNet	238
7.2.1	Klasyfikacja i analiza danych psychometrycznych	238
7.2.1.1	Opis problemu	238
7.2.1.2	Dane	239
7.2.1.3	Proces uczenia	240
7.2.1.4	Porównanie i analiza wyników	244
7.2.2	Typowe medyczne dane porównawcze	263
7.2.2.1	Zapalenie wyrostka robaczkowego	263
7.2.2.2	Dane dotyczące raka piersi	265
7.2.2.3	Dane dotyczące zapalenia wątroby	266
7.2.2.4	Dane dotyczące cukrzycy	267
7.2.2.5	Choroby tarczycy	268
7.3	Aproksymacja	271
7.3.1	Funkcja Hermita	271
7.3.2	Funkcja Gabora i Girosiego	272
7.3.3	Funkcja Sugeno	273
8	Zakończenie	277
	Bibliografia	279
	Skorowidz	298
	Ilustracje kolorowe	305