

Laboratorium Rozpoznawania Obrazów – Ćwiczenie #5 & #6

Rozpoznawanie cyfr z wykorzystaniem sieci neuronowych

Termin oddawania: **21.05.2012, 10.05.2012, 28.05.2012, 31.05.2012**

W kolejnym ćwiczeniu pozostaniemy przy rozpoznawaniu cyfr, zmieniając jedynie typ klasyfikatora używanego do tego zadania. Dla kompletności powtórzę w kolejnych akapitach informacje o zestawie danych, które będą podstawą do konstrukcji klasyfikatora.

Do pierwszej fazy zadania – przygotowania klasyfikatora – wykorzystamy dane z bazy danych MNIST (<http://yann.lecun.com/exdb/mnist>). Zbiory uczący i testowy można pobrać spod podanego wyżej adresu. Dodatkowo są dostępne w laboratorium 313 oraz na serwerze Galera.

Pracę klasyfikatora „u klienta” będzie można przećwiczyć korzystając z drugiego zbioru testowego, zawierającego cyfry pisane przez studentów WEiTl. Ten zbiór, także w formacie MNIST, jest dostępny w laboratorium 313 oraz na serwerze Galera.

Warto wiedzieć, że obrazy są normalizowane po zeskanowaniu w następujący sposób:

1. Prostokąt zawierający czarno-biały obraz znaku zeskanowanego w rozdzielczości 300 dpi jest skalowany proporcjonalnie do prostokąta o większym z wymiarów równym 20. W trakcie skalowania obraz jest zamieniany na skalę szarości (proporcjonalnie do liczby oryginalnych pikseli pierwszego planu przypadających na jeden piksel obrazu po przeskalowaniu).
2. Jest wyznaczany środek ciężkości przeskalowanego znaku, a znak jest umieszczany w obrazie 28x28 pikseli tak, żeby środek ciężkości znalazł się na środku tego większego obrazu.

Drugi zestaw znaków został poddany dodatkowo usuwaniu przekosu.

Łączy do plików na serwerze Galera są następujące:

<http://galera.ii.pw.edu.pl/~rkz/mnist.zip> (oryginalne dane MNIST)

<http://galera.ii.pw.edu.pl/~rkz/testdigits.zip> („polski” zbiór testowy w formacie MNIST)

<http://galera.ii.pw.edu.pl/~rkz/reference.zip> (rozwiązanie referencyjne - warto w nim sprawdzić, jak odczytywać dane – lub po prostu przejąć z niego funkcję odczytu danych)

Państwa zadaniem, jest przygotowanie sieci neuronowej klasyfikującej cyfry. Oprócz jakości klasyfikacji na zbiorach testowych (zbiorze pochodzącym z bazy danych MNIST i „rzeczywistym” zbiorze testowym) należy podać macierze pomyłek. Dodatkowe zadanie polega na zaproponowaniu sposobu wyznaczania „ufności” do decyzji klasyfikatora (chodzi o wartość z przedziału 0..1 opisującą pewność, że decyzja klasyfikatora jest poprawna).

Rozwiązaniem referencyjnym, jest klasyfikator najbliższego sąsiedztwa (w wersji 5-NN z głosowaniem absolutną większością) działający na cechach otrzymanych bezpośrednio z obrazów znaków metodą składowych głównych (PCA). W testach użyłem 40 składowych głównych. Wyniki klasyfikacji podsumowuje poniższa tabela.

	Zbiór testowy MNIST			„Polski” zbiór testowy		
	OK.	Błąd	Odrzucenie	OK.	Błąd	Odrzucenie
Jakość klasyfikacji	97.21%	2.02%	0.77%	79,96%	15,70%	4,34%

Jako podstawowe metody klasyfikacji cyfr pisanych odręcznie widzę:

1. Wielowarstwową sieć neuronową uczoną algorytmem wstecznej propagacji błędu. Zasadniczym problemem jest tutaj ustalenie architektury sieci, tzn. dobranie liczby warstw i liczby neuronów w poszczególnych warstwach. I oczywiście opracowanie metody uczenia sieci.

W tym przypadku dość istotne jest unikanie przeuczenia sieci, czyli zbytniego dopasowania do danych uczących. Zgodnie z zasadami sztuki, najpewniejszym sposobem uniknięcia tego zjawiska byłoby wydzielenie ze zbioru uczącego tzw. zbioru walidującego, który posłuży do przerywania procedury uczenia (zwykle uczenie przerywa się, kiedy wyniki klasyfikacji na zbiorze walidującym pogarszają się w dwóch, trzech kolejnych iteracjach uczenia).

2. Klasyfikator LIRA (zmodyfikowany perceptron). W tym przypadku problemem może być długi czas uczenia – warto pomyśleć o optymalizacji tego procesu.

Przy tym klasyfikatorze możliwość zbytniego dopasowania klasyfikatora do danych wydaje się mniej groźna (co nie znaczy, że problem nie może się ujawnić). Ze względu na czas uczenia przypuszczam, że wystarczy ograniczenie z góry liczby wykonywanych iteracji uczenia (przejsć po całym zbiorze uczącym).

Co ważne, wybieracie Państwo tylko **jedną** z powyższych pozycji, chyba że ktoś z Was chciałby spróbować jakiejś innej metody klasyfikacji (poza klasyfikatorem NN i liniowym). W takim przypadku proszę o kontakt – przed wpuszczeniem się w dłuższe prace warto przedyskutować sprawę.

W sprawozdaniu proszę zamieścić:

1. Opis metody klasyfikacji znaków.
2. Opis algorytmu uczenia klasyfikatora (proszę podać także informacje o czasie uczenia).
3. Dane dotyczące jakości klasyfikacji znaków.
4. Analizę macierzy pomyłek i wyniki sprawdzenia wyznaczonego poziomu ufności z poprawnymi i błędnymi decyzjami klasyfikatora.