

### Kolejne etapy realizacji projektu:

1. Zapoznać się z algorytmami diagnostycznymi (algorytmami klasyfikacji), określonymi w temacie:
  - a) sieci neuronowe – sieć jednokierunkowa, z jedną warstwą ukrytą, uczona metodą wstecznej propagacji błędów (z członem momentum i bez członu momentum) (literatura: 3, 7, 8, 9)
  - b) algorytmy minimalno-odległościowe: algorytm najbliższa średnia (NM – nearest mean) i algorytm k- najbliższych sąsiadów (k-NN – k- nearest neighbors) (literatura: 1, 2, 4, 5)
  - c) empiryczny algorytm bayesowski: algorytm Bayesa z prawdopodobieństwami *a priori* oraz prawdopodobieństwami cech w poszczególnych klasach (dla cech dyskretnych) szacowanymi (estymowanymi) ze zbioru uczącego metodą częstościową i/lub warunkowymi gęstościami cech w klasach (dla cech ciągłych) szacowanymi metodą histogramu, metodą empirycznej dystrybucyjności lub metodami jądrowymi (*kernel methods*) (literatura: 1, 2, 4, 5, 6)
2. Zapoznać się z materiałem empirycznym – zdefiniować problem rozpoznawania (klasyfikacji) – określić liczbę i znaczenie klas, liczbę i znaczenie cech oraz charakter cech (ciągłe, wielowartościowe, binarne, itp.)
3. Wyznaczyć ranking cech pod względem ich przydatności do klasyfikacji) korzystając z dowolnej miary (kryterium) jakości cech stosowanych w selekcji cech (proponuję kryterium Kołmogorowa, gdyż jest bardzo proste rachunkowo) – (literatura: 1, 2, 10)
4. Zaplanować badania eksperymentalne dla następujących założeń
  - a) Trenowanie i testowanie zastosowanych klasyfikatorów z wykorzystaniem 5 razy powtarzanej metody 2-krotnej walidacji krzyżowej (5x2cv). Metoda 2cv polega na tym, że dostępny zbiór danych dzielimy (losowo) na dwie jednakowe części (jedna część stanowi zbiór uczący, a druga zbiór testujący) i wykonujemy badania. Następnie powtarzamy badania dla odwrotnej sytuacji – zbiór, który był uczącym teraz jest testującym, a testujący –uczącym. 5xcv oznacza, że procedurę powtarzamy 5 razy (każdorazowo na nowo losujemy dwie połówki) i wyniki uśredniamy. Jakość klasyfikacji (poprawność diagnozy) należy mierzyć częstością poprawnych rozpoznań (diagnoz) na zbiorze testującym.
  - b) Przeprowadzić badania dla różnej liczby cech (poczynając od jednej –najlepszej wg. wyznaczonego rankingu, a następnie dokładać kolejno po jednej (również według wyznaczonego rankingu) tak długo, aż zostanie znaleziona najlepsza liczba cech. Dodawanie cech powinno poprawiać jakość klasyfikacji, ale do pewnego momentu – dalsze dodawanie cech jakość pogorszy. Trzeba znaleźć optimum.
  - c) Przeprowadzić badania dla następujących algorytmów:
    - Dla sztucznych sieci neuronowych – sieć jednokierunkowa z 1 warstwą ukrytą dla 3 różnych liczb neuronów w warstwie ukrytej oraz dla uczenia metodą propagacji wstecznej z momentum i bez momentum.
    - Dla metod minimalno-odległościowych – algorytmy NM, 1-NN, k-NN (dla 2 różnych wartości k), dla dwóch różnych miar odległości (np. Euklidesowa, Manhattan) oraz z normalizacją i bez normalizacji
    - Dla empirycznego algorytmu bayesowskiego: algorytm dla 0-1 funkcji strat oraz algorytm dla funkcji strat z wartościami  $L(i,j)=|i-j|$ , algorytm dla założenia, że

cechy są niezależne oraz bez tego założenia (w tym przypadku należy estymować prawdopodobieństwa (funkcje gęstości) cech w klasach dla możliwie licznych grup cech – ta liczba zależy od rozmiaru zbioru uczącego)

- d) Dla każdego pojedynczego eksperymentu (pojedynczy eksperyment to doświadczalne wyznaczenie jakości klasyfikacji (patrz punkt 4a) dla danego algorytmu, danych wartości parametrów algorytmu i dla danej liczby cech) należy przedstawić wyniki (jakości klasyfikacji) w formie uśrednionej (względem 5 powtórzeń metody 2-krotnej walidacji krzyżowej) oraz dla jednej (dowolnie wybranej) macierzy konfuzji.
- 5. Zaimplementować algorytmy diagnostyczne (klasyfikacji) (środowisko implementacji dowolne), tak aby można było przeprowadzić badania eksperymentalne według przedstawionych założeń
- 6. Zrealizować badania eksperymentalne według przedstawionych w punkcie 4 założeń
- 7. Przeprowadzić dyskusję wyników i przedstawić wnioski

#### **Zawartość sprawozdania:**

1. Charakterystyka analizowanego problemu (jako zadania rozpoznawania)
2. Opis stosowanych algorytmów
3. Informacja o środowisku implementacyjnym
4. Opis badań eksperymentalnych
5. Wyniki (w formie graficznej, tabeli, itp.)
6. Dyskusja wyników i wnioski wynikające z badań.

#### **Literatura:**

1. M. Kurzyński, Rozpoznawanie obiektów – metody statystyczne, Oficyna Wydawnicza Pol. Wr., Wrocław 1997
2. K. Stąpor, Metody klasyfikacji obiektów w wizji komputerowej, PWN, Warszawa 2011
3. L. Rutkowski, Metody i techniki sztucznej inteligencji, PWN, Warszawa 2005
4. R. Tadeusiewicz, M. Flasiński, Rozpoznawanie obiektów, PWN, Warszawa 1991
5. J. Koronacki, J. Cwik, Statystyczne systemy uczące się, WNT, Warszawa 2005
6. M. Krzyśko, Systemy uczące się, WNT, Warszawa 2008
7. J. Łęski, Systemy neuronowo-rozmyte, PWN, Warszawa 2004
8. S. Osowski, Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Pol. Warszawskiej, Warszawa 2006
9. Sieci Neuronowe, seria: Biocybernetyka i Inżynieria Biomedyczna, tom 6, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000
10. W. Sobczak, W. Malina, Metody redukcji i selekcji informacji, WNT, Warszawa 1987