ROB LAB 2 – Adrian Wiśniewski

Implementacja

- preprocess(data, drop_percentage) – Funkcja filtrująca dane. Iteruje po wszystkich atrybutach i dla każdego odrzuca (drop_percentage)% wartości skrajnych.
- bayes(train, classes, test, method) – Klasyfikator będący tematem laboratorium. Parametr method przyjmuje wartości od 1 do 3 odpowiadające kolejnym wariantom modelu.
- testbayes(train, test, method, columns) – Funkcja testująca klasyfikator.
- bestbayes(train, test, method) – Funkcja wybierająca optymalny zestaw atrybutów.

Wnioski

- Najlepszy wynik dla dwóch atrybutów zapewnia wariant 3. klasyfikatora – aproksymacja oknem Parzena (h1 = 0.006). Najsłabiej wypadł wariant 1. Wyniki to kolejno: [2 7] - 85.3%, [2 4] - 96.1% i [3 4] - 97.3%.
- Wybranie 1/2 zbioru uczącego dla metody 2. podnosi skuteczność klasyfokacji do 98.3%. Przy odpowiednio dobranych atrybutach (3 i 4) wzięcie co dziesiątego rekordu podnosi skuteczność nawet do 99.1%. Zjawisko to można wytłumaczyć nadmiernym dopasowaniem modelu, jednak losowe dopieranie zbioru trenującego może dać równie losowy rezultat.
- Najlepszy klasyfikator uzyskano dla najmniejszej wartości h1=0.0001 osiągając 99.78% skuteczności. Mniejszy rozmiar okna faworyzuje składniki należące do najbliżej leżących próbek.
- Najlepsze zestawy cech dla wszystkich trzech wariantów klasyfikatora to: [2 3 4 5] – 86.8%, [2 3 4 5] – 99.89% i [3 4 5] – 99.83%.
- W zależności od wylosowanej próby zbioru testowego skuteczność klasyfikacji zmienia się w przedziałach od 90.9%, 99.27%, 99.27% do 91.00%, 100%, 99.85%. Z wyjątkiem pierwszego klasyfikatora, jakość modelu się pogorszyła.
- Klasyfikator z pierwszego ćwiczenia osiągał najwyższą skuteczność 94%. Klasyfikator ten wielokrotnie częściej się mylił w porównaniu z klasyfikatorem Bayesa z drugiego ćwiczenia (około 54 razy częściej przyjmując najlepszy wynik 99.89%).