

Klasyfikacja i regresja: Testowanie klasyfikatorów

14 listopada 2018

Opis pliku z zadaniami

Wszystkie zadania na zajęciach będą przekazywane w postaci plików `.pdf`, sformatowanych podobnie do tego dokumentu. Zadania będą różnego rodzaju. Za każdym razem będą one odpowiednio oznaczone:

- Zadania do wykonania na zajęciach oznaczone są symbolem \triangle – nie są one punktowane, ale należy je wykonać w czasie zajęć.
- Punktowane zadania do wykonania na zajęciach oznaczone są symbolem \diamond – należy je wykonać na zajęciach i zaprezentować prowadzącemu.
- Zadania do wykonania w domu oznaczone są symbolem \star – są one punktowane, należy je dostarczyć w sposób podany przez prowadzącego i w wyznaczonym terminie (zwykle przed kolejnymi zajęciami).

1 Porównanie klasyfikatorów

10p.◇

Treść

Zadaniem jest porównanie naiwnego klasyfikatora Bayesa (NKB) z innymi klasyfikatorami np. regresją logistyczną. Plan ćwiczenia jest następujący:

1. Zapoznaj się z kodem dostępnym na stronie przedmiotu.
2. Zapoznaj się z metodą `evaluate_classifier`: wykonuje ona prostą ewaluację klasyfikatora pod względem czasu uczenia oraz błędu klasyfikacyjnego i błędu kwadratowego na wybranym zbiorze danych.
3. Na podstawie implementacji metody `evaluate_classifier` i dostarczonego szablonu kodu oraz dokumentacji pakietu `sklearn` zaimplementuj:
 - Liczenie błędu na zbiorze uczącym i na zbiorze testowym oraz czasu uczenia i testowania,
 - Liczenie błędu i czasów dla różnej wielkości zbioru treningowego,
 - Ewaluację klasyfikatorów na różnych zbiorach danych,
 - Wyświetlenie wyników w sposób łatwy do wykreślenia krzywych uczenia (czyli krzywej pokazującej zależność pomiędzy liczbą przykładów uczących i błędem klasyfikatora).
4. Wykonaj eksperymenty dla paru klasyfikatorów (użyj również Twojej implementacji NKB; w miarę konieczności rozszerz go odpowiednio) i zbiorów danych dostępnych na stronie przedmiotu.
5. Zastanów się, jak otrzymać bardziej wiarygodne wyniki.
6. Przedstaw wykresy czasów i błędów jako funkcji liczby przykładów uczących.
7. Zinterpretuj wyniki.

Uwagi do Twojej implementacji NKB:

- Jeżeli nie zabezpieczyłeś swojej implementacji przed pustymi wartościami atrybutów to wprowadź odpowiednie poprawki do swojego kodu.
- Prawdopodobieństwa szacowane z małej liczby przykładów jest niewiarygodne. Dlatego spróbuj ulepszyć swoją implementację NKB stosując np. korektę Laplace'a.