

Zaimplementować naiwny klasyfikator Bayesa (Gaussowski). Do eksperymentów wykorzystać zbiór danych dot. jakości wina.

Założenia wstępne

Rodzaj wina – **czerwone**.

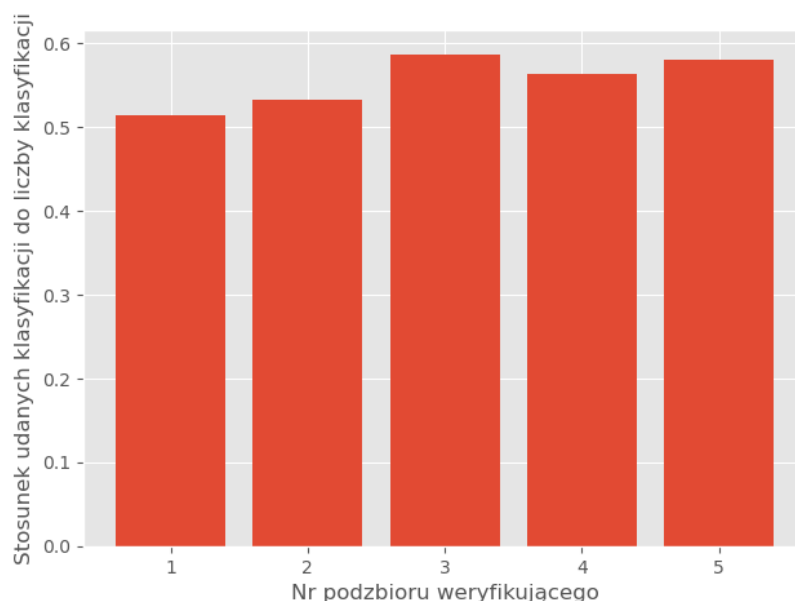
Dokonano podziału danych w sposób losowy na zbiór uczący i weryfikujący w proporcjach 60-40. Przy walidacji krzyżowej dane również podzielono losowo na podzbiory. Jako wskaźnik jakości modelu przyjęto stosunek próbek poprawnie zaklasyfikowanych ze zbioru weryfikującego do liczby prób klasyfikacji (rozmiaru zbioru weryfikującego). Ten wskaźnik jest intuicyjny i w prosty sposób przekazuje informację, czy model nadaje się do użytku.

Wyniki eksperymentów

W tabeli poniżej przedstawiono wskaźniki jakości (procent udanych klasyfikacji) dla dwóch różnych metod walidacji modelu – walidacja krzyżowa 5-krotna i walidacja z podziałem na zbiór weryfikujący i uczący.

Przy walidacji krzyżowej, jako procent udanych klasyfikacji przyjęto średnią arytmetyczną wskaźników jakości każdego z pięciu modeli. Dodatkowo na wykresie przedstawiono wskaźnik jakości dla każdego modelu zbudowanego tą metodą.

Metoda walidacji	Stosunek udanych klasyfikacji
Walidacja 5-krotna krzyżowa	0.555
Walidacja z podziałem na zbiór uczący i weryfikujący	0.565



Z tabeli wynika kilka rzeczy:

1. Modele są dość słabej jakości (choć dużo lepiej niż losowa klasyfikacja). Może być tego wiele przyczyn. Jedną z nich to niska jakość danych. Są one mało zróżnicowane – występuje duża liczba win średnich i bardzo mało win słabych lub dobrych. Z tego względu, jeżeli chcielibyśmy klasyfikować skrajne wina stworzonym modelem, to mogłyby wystąpić problemy.
2. Wybrany wskaźnik jakości nie oddaje tego co dzieje się w modelu w stu procentach. Można wybrać wskaźnik, który uwzględnia jak duży błąd został popełniony przy klasyfikacji. Przykładowo zaklasyfikowanie wina z jakością 3 jako wina z jakością 4 powinno być traktowane inaczej niż zaklasyfikowanie tego samego wina jako wina z jakością 8. Nie zmienia to jednak faktu, że model ze skutecznością około 50% nie nadaje się do użytku.
3. Obie metody walidacji dają bardzo podobne wskaźniki jakości, więc trudne jest określenie, czy któraś z tych metod jest jednoznacznie lepsza.

Odpowiedzi na pytania

Jakiego podzbioru danych (z tych którymi dysponujemy) użyjemy do zbudowania docelowego modelu na potrzeby klasyfikowania nowych próbek (czyli dla tych dla których budujemy klasyfikator)?

Do zbudowania modelu najlepiej będzie użyć całego zbioru danych. Wykres zamieszczony dla walidacji krzyżowej pokazuje, że cały zbiór się do tego nadaje. Dodatkowo w klasyfikatorze Bayesa zjawisko overfittingu do danych nie jest częste.

Jak zinterpretować różnice/brak różnic w wynikach z weryfikacji jakości modelu obu metod (k-krotna walidacja vs zbiór treningowy i testowy)?

Brak różnic w wynikach oznacza, że w tym konkretnym przypadku wybrana metoda klasyfikacji daje słabe rezultaty niezależnie od metody walidacji i wyboru zbioru uczącego. W żadnym z łącznie 6 zbudowanych modeli nie udało się przekroczyć 60% udanych klasyfikacji.