Uczenie drzewa decyzyjnego na przykładzie wykrywania okularów na obrazach Olivetti

Karol Działowski

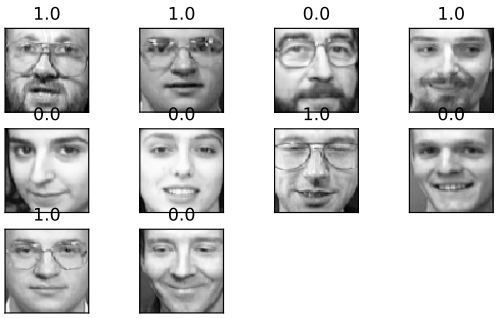
15.05.2019

1 Opis problemu

Celem zadania jest wykrywanie okularów na danych *olivetti faces*. Klasyfikację przeprowadzono przy pomocy CART (classification and regression trees). Dane wejścowe zredukowano za pomocą analizy komponentów składowych (PCA).

W eksperymencie porównano wpływ na dokładność testową:

- liczby wymiarów danych (PCA)
- użytych funkcji nieczystości
- ograniczenia drzewa za pomocą głębokości
- ograniczenia drzewa za pomocą procentu przykładów w liściu
- ograniczenie drzewa za pomocą przycinania
  - przy przeszukiwaniu wyczerpującym
  - przy przeszukiwaniu zachłannym

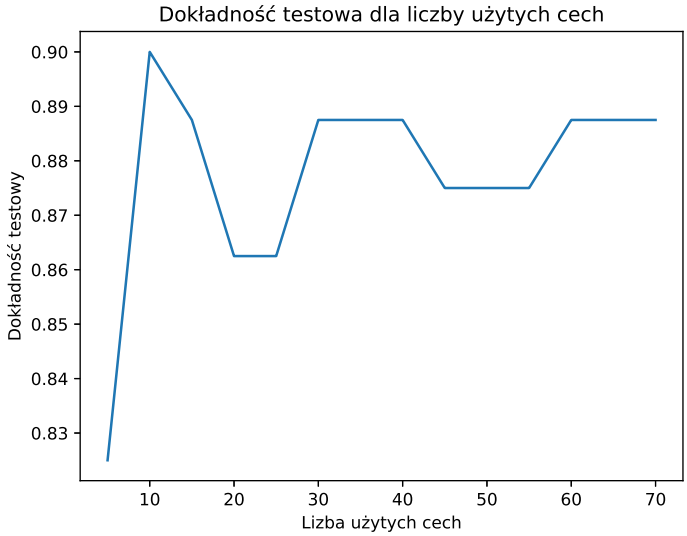


Rysunek 1: Przykład predykcji

2 Wpływ na dokładność testową

2.1 Liczba cech danych

Drzewo CART z funkcją nieczystości typu entropie uczono danymi o wymiarowościach $[15, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70]$ i porównano wpływ na błąd dla danych testowych.



Rysunek 2: Liczba cech danych a dokładność testowa dla CART z funkcją nieczystości typu entropia

Dla 1małej ilości cech — 10 wymiarów algorytm osiąga maksymalną dokładność, na poziomi 90%. Przy zwiększaniu liczby cech do 25 dokładność maleje, by później znowu rosnąć do poziomi 87.5% dla 30 i 40 cech.

2.2 Wykorzystana funkcja nieczystości

Porównano trzy funkcje nieczywości: entropia, indeks Giniego, błąd klasyfikacji.
Błąd klasyfikacji:

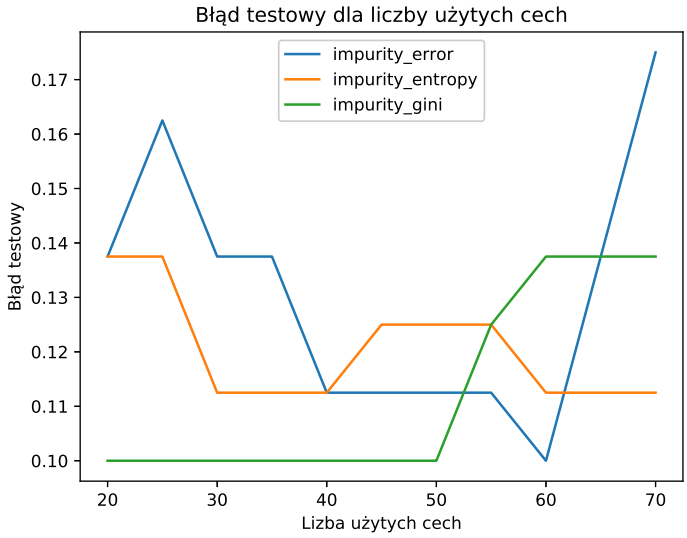
$$i(t) = 1 - \max_y P(y|t) \tag{1}$$

Entropia:

$$i(t) = -\sum_y \frac{P(y|t) \cdot log_2 P(y|t)}{1} \tag{2}$$

Indeks Giniego:

$$i(t) = 1 - \sum_y P^2(y|t) \tag{3}$$

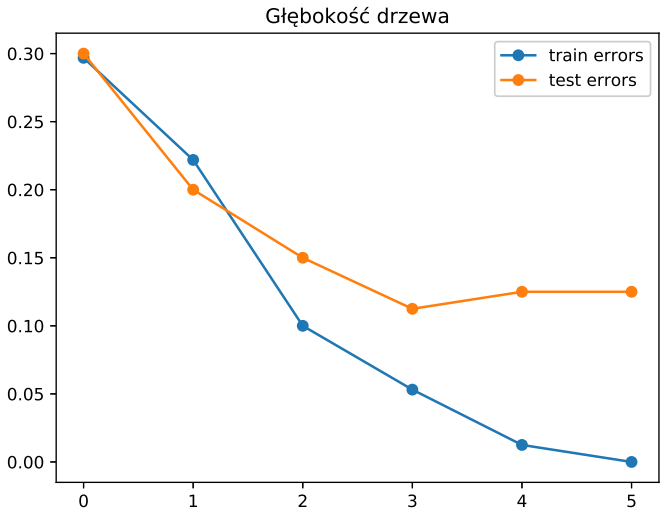


Rysunek 3: Porównanie funkcji nieczystości

Najlepsze wyniki osiąga indeks Giniego, który dla osiąga minimalny błąd rzędu 10%. Wyniki eksperymentu nie są przekonujące i nie można wysnąć oczywistych wniosków.

2.3 Ograniczenie głębokości

Wykonany testy ograniczenia głębokości drzewa z funkcją nieczystości typu Entropia dla danych wejściowych o 50 cechach. Porównano wszystkie możliwe głębokości dla zbudowanego drzewa.

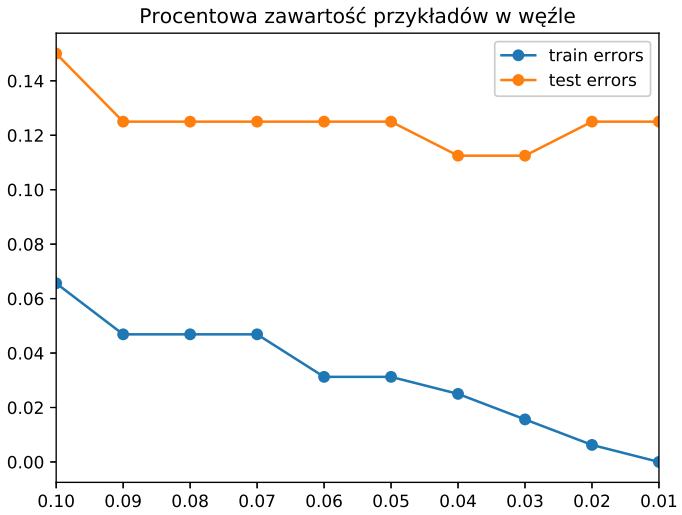


Rysunek 4: Porównanie ograniczenia głębokości drzewa

Błąd testowy maleje wraz z wzrostem liczby poziomów drzewa i osiąga minimum dla głębokości równej 3. Dla głębokości 4 i 5, wyniki testów są minimalnie gorsze od minimum. Warto takie drzewo przyciąć na głębokość 3.

2.4 Ograniczenie drzewa stosując procent przykładów

Usunięto te węzły, które nie zawierały w sobie danego poziomu ogółu przykładów. Testowano drzewa z liściami które mają 10% i mniej wszystkich przykładów. Porównanie przeprowadzono dla drzewa korzystającego z funkcji nieczystości typu entropia dla danych o 50 cechach.



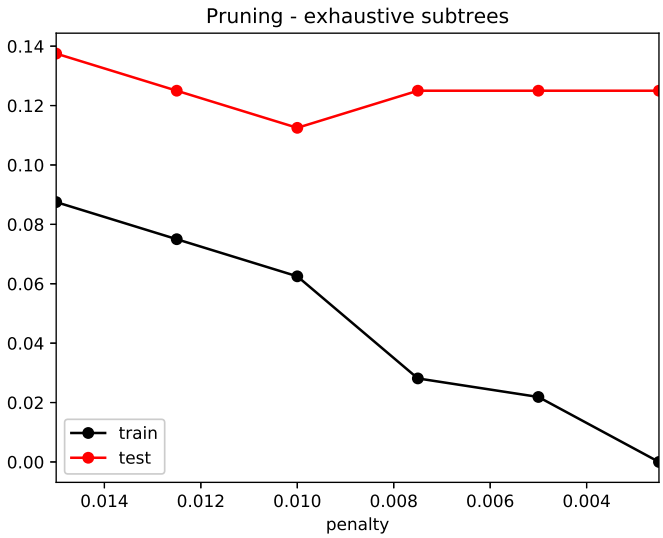
Rysunek 5: Porównanie ograniczenia minimalnego procentu przykładów w liściach

Osiągnięto minimum błędu testowego dla minimum 4% i 3% ogółu przykładów w węźle osiągając błąd 11.25%. Osiągane wyniki są lepsze niż dla ograniczenia głębokości drzewa.

2.5 Przycinanie drzewa

2.5.1 Przeszukiwanie wyczerpujące

Porównano wpływ kar przy użyciu wyczerpującego przeglądania drzewa. Drzewo zbudowano dla danych o 50 cechach i przy użyciu funkcji nieczystości typu entropia.



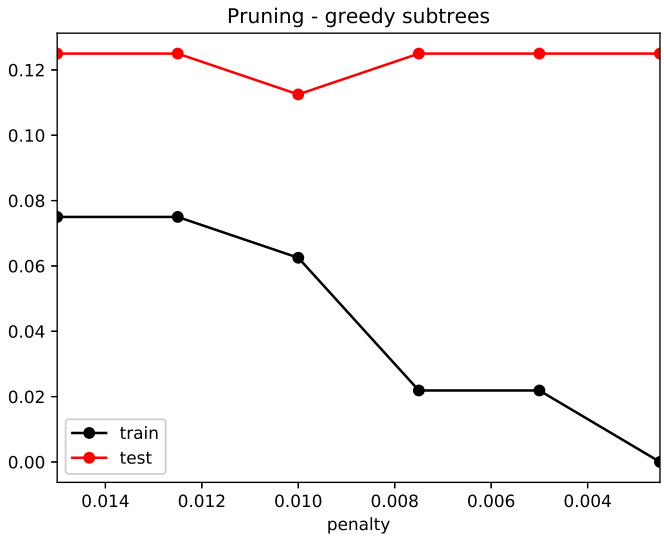
Rysunek 6: Porównanie wpływu kar dla przycinania z przeglądaniem wyczerpującym

Zmiejszając karę za każdy liść błąd na danych uczących się zmniejsza, w pewnym momencie prowadząc do przeuczenia. Dla kary równej 0.10 błąd na danych testowych jest najmniejszy i wynosi 11.25%, czyli osiągamy podobny wynik jak przy ograniczeniu stosując procent przykładów.

Średni czas budowania drzewa wynosił 6.1 sekundy.

2.5.2 Przeszukiwanie zachłanne

Porównano wpływ kar przy użyciu zachłannego przeglądania drzewa. Drzewo zbudowano dla danych o 50 cechach i przy użyciu funkcji nieczystości typu entropia.



Rysunek 7: Porównanie wpływu kar dla przycinania z przeglądaniem wyczerpującym

Zmiejszając karę za każdy liść błąd na danych uczących się zmniejsza, w pewnym momencie prowadząc do przeuczenia. Dla kary równej 0.10 błąd na danych testowych jest najmniejszy i wynosi 11.25%, czyli osiągamy podobny wynik jak przy ograniczeniu stosując procent przykładów.

Średni czas budowania drzewa wynosił 4.5 sekundy. Zachłanne przeszukiwanie daje rezultaty zbliżone do przeszukiwania wyczerpującego (w naszym eksperymencie takie same) przy mniejszej złożoności obliczeniowej.

3 Wnioski

W eksperymencie porównano różne sposoby budowania drzewa CART na przykładzie danych *olivetti faces*, których wymiarowość została zredukowana przy pomocy PCA. Przy odpowiednim stworzeniu drzewa można uzyskać dokładność na poziomie 90% poprawnie przewidzianych klas na danych testowych.