

Przykład użycia CMIM

Daniel Ponikowski

28 kwietnia 2019

Wywołanie Funkcji:

Dokładny opis opisany jest w książce: Cover, T. M. and Thomas, J. A. (1990). *Elements of Information Theory*. John Wiley, New York.

Użyj znanego zbioru danych *BreastCancer* z pakietu *mlbench*. Potrzebujemy danych bez braków danych oraz zmienną objaśnianą jako zmienną kategorię.

```
data("BreastCancer")
X <- BreastCancer %>% na.omit %>% "["(,-c(1,11))
y <- BreastCancer$Class[as.numeric(row.names(X))]
```

Przykład użycia

```
CMIM_selection(X,y,5)
```

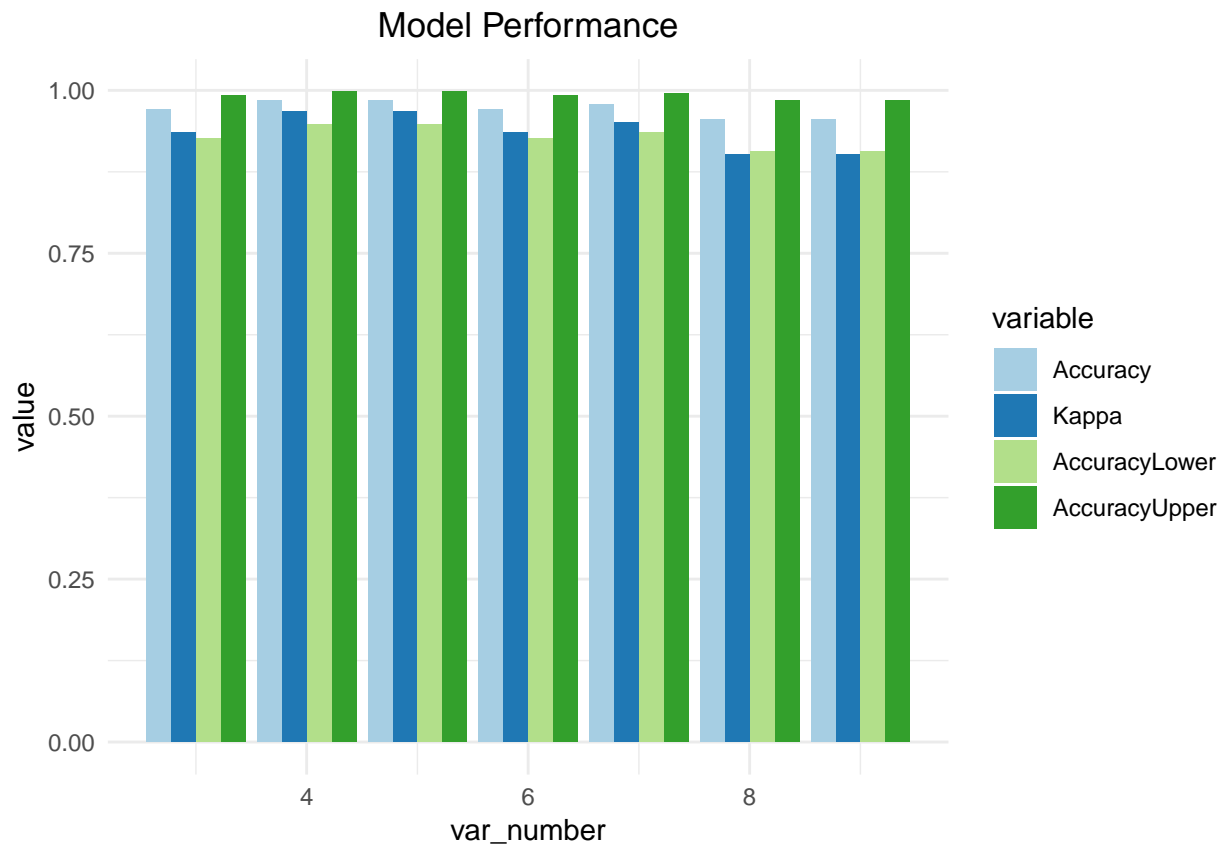
```
## $S
## [1] 2 6 1 8 7
##
## $score
## [1] 0.48681994 0.09995591 0.07688931 0.07076222 0.06307534
```

Sprawdzenie działania algorytmu na zmiennych wybranych przez algorytm CMIM.

Dla 3,4,...,9 wybranych zmiennych zastosuj algorytm regresji logistycznej na tych danych i oceń jego działanie.

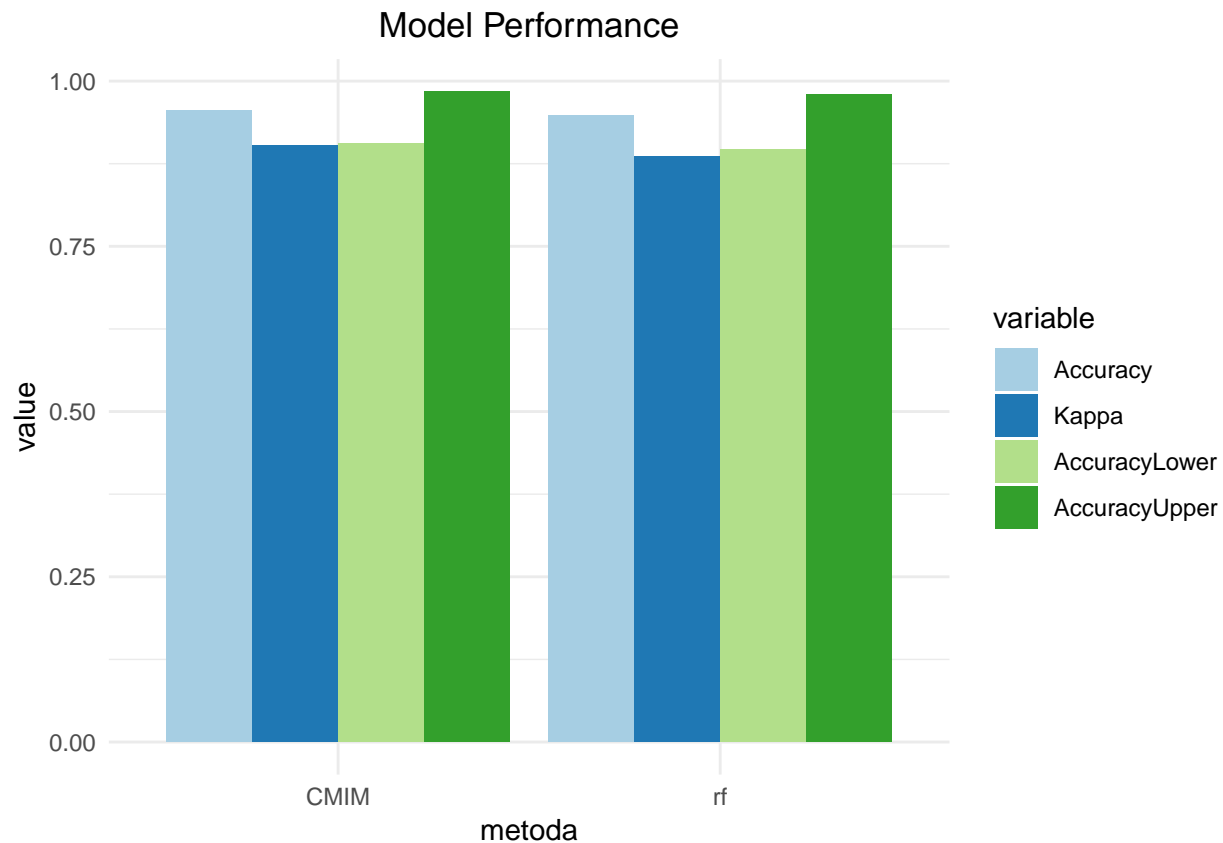
Podział na zbiór treningowy i testowy

```
data <- BreastCancer[, -1] %>% na.omit()
train_num <- createDataPartition(y = data$Class, p = 0.8, list = FALSE)
train_dataset <- data[train_num,]
test <- data[-train_num,]
```



Roznice w kolejnosci

Sprawdzmy także różnice w kolejności ważności zmiennych w porównaniu do wbudowanej w pakiet *caret* funkcji *varImp*, używając modelu lasu losowego (zbudowanego na wszystkich zmiennych). Wybiore 5 najlepszych zmiennych dla obu metod i zbuduje klasyfikator na nich, oraz je porównam (będzie to pojedyncze drzewo, ponieważ jest to dość prosty zbiór, wybrałem słabszy klasyfikator).



```
## [1] "Zmienne wybrane przez algorytm CMIM: Cell.size, Bare.nuclei, Cl.thickness, Normal.nucleoli, Bl.
```

```
## [1] "Zmienne wybrane przez algorytm rf: Cell.size, Cell.shape, Epith.c.size, Bare.nuclei, Normal.nuc.
```

Więc można powiedzieć że lepsze zmienne zostały wybrane przez metode *CMIM*, roznicie nie są duże, ponieważ zbior danych nie jest trudny. Jednak pokazuje to skuteczność metody *CMIM*.