Algorytmika praktyczna

Bartosz Ochędowski

w59300

IID

Szacowanie błędu rzeczywistego

**Zadanie 2.**

1. dane <- **read.table**(**file**="https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data", sep=",")
2. dane <- dane[**sample**(**nrow**(dane)),]
3. dane <- **unique**(dane)
4. n <- **nrow**(dane)
5. p <- 4/5 # ile przypadkow do treningowego
6. treningowy <- dane[1:(**floor**(n\*p)),]
7. testowy <- dane[(**floor**(n\*p+1)):n,]
8. drzewo <- C5.0(x=treningowy[,1:4], y=treningowy[,5])
9. **plot**(drzewo)
10. pred <- predict.C5.0(drzewo, testowy[,1:4])
11. comp <- pred == testowy[,5]
12. err <- **length**(comp[comp == FALSE])
13. err <- err/**nrow**(testowy)

**Zadanie 3.**

1. dane <- **read.table**(**file**="https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data", sep=",")
2. dane <- **unique**(dane)
3. n <- **nrow**(dane)
4. dane$id <- **sample**(n) #nowa kolumna
5. pred <- **data.frame**()
6. test <- **data.frame**()
7. **for** (i **in** 1:n){
8. treningowy <- **subset**(dane, id != i)
9. testowy <- dane[dane$id == i,]
10. drzewo <- C5.0(x=treningowy[,1:4], y=treningowy[,5])
11. przew <- predict.C5.0(drzewo, testowy[,1:4])
12. pred <- **rbind**(pred, przew)
13. test <- **rbind**(test, testowy[,5])
14. }
15. err <- 0
16. **for** (i **in** 1:n){
17. **if** (pred[i,1] != test[i,1])
18. err <- err + 1
19. }
20. err <- err/n

**Zadanie 6.**

1. dane <- **read.table**(**file**="https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data", sep=",")
2. dane <- **unique**(dane)
3. rows <- **nrow**(dane)
4. dane <- dane[**sample**(rows),]
5. dane$id <- 1:rows
6. n <- 10
7. div <- rows %/% n
8. mod <- rows %% n
9. # Podział zbioru na "prawie równe" części
10. sampleSizes <- **vector**(**mode** = "integer", **length** = n)
11. **for** (i **in** 1:n){
12. **if**(i<=mod)
13. sampleSizes[i] <- div+1
14. **else**
15. sampleSizes[i] <- div
16. }
17. # 10-krotna walidacja krzyżowa
18. pred <- **data.frame**()
19. test <- **data.frame**()
20. **for** (i **in** 1:n){
21. begin <- **sum**(sampleSizes[0:(i-1)])
22. **end** <- **sum**(sampleSizes[1:i])
23. **cat**(begin, **end**, "**\n**")
24. trainingset <- dane[**c**(0:(begin), (**end**+1):rows),]
25. testset <- dane[(begin+1):**end**,]
26. **cat**(**nrow**(testset), "**\n**")
27. model <- C5.0(x=trainingset[,1:4], y=trainingset[,5])
28. temp <- **as.data.frame**(predict.C5.0(model, testset[,1:4]))
29. pred <- **rbind**(pred, temp)
30. test <- **rbind**(test, **as.data.frame**(testset[,5]))
31. }
32. # obliczenie błędu
33. err <- pred == test
34. err <- **length**(err[err == FALSE])
35. err <- err/rows