Machine Learning - 1100-ML0ENG (Ćwiczenia informatyczne Z-23/24)

Boosting - German credit

Dataset

```
gc<-read.csv("germancredit.csv", stringsAsFactors = T)
summary(gc)
gc$credit_risk <- factor(gc$credit_risk, levels = c(0, 1))
levels(gc$credit_risk)<-c("no","yes")
```

```
library(caTools)
set.seed(12345)
split = sample.split(gc$credit_risk, SplitRatio = 0.7)
gc.Train <- subset(gc, split == TRUE)
gc.Test <- subset(gc, split == FALSE)
summary(gc.Test)
```

```
prop.table(table(gc$credit_risk))
prop.table(table(gc.Train$credit_risk))
prop.table(table(gc.Test$credit_risk))
```

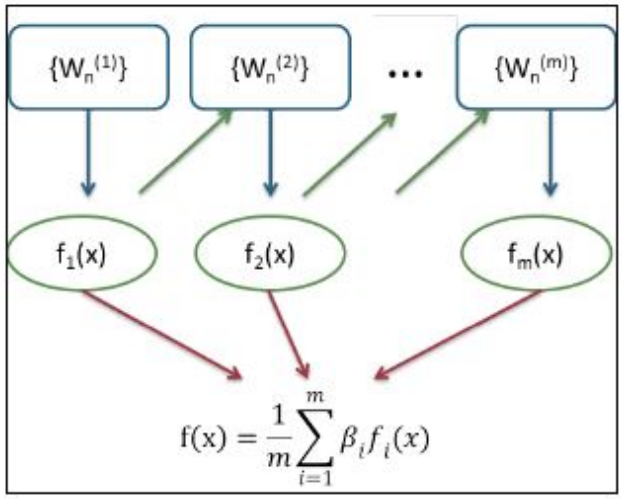
As a reminder, we are building a model of a tree. We will use it to compare with ensemble methods.

```
library(rpart)
gc.rpart <- rpart(credit_risk~., gc.Train)
library(rattle)
fancyRpartPlot(gc.rpart)
plot(gc.rpart)
gc.pred.rpart <- predict(gc.rpart,newdata = gc.Test,type="class")
table(gc.pred.rpart,gc.Test$credit_risk)
acc.rpart=acc(gc.pred.rpart, gc.Test$credit_risk)
acc.rpart
roc.function(gc.pred.rpart, gc.Test$credit_risk)
```

Ada boost model

Adaptive Boosting (in short AdaBoost) follows metadata algorithms of machine learning.

- It is used for weak learner and is adaptive in the sense.
- It is weak for the instances that are misclassified by the previous classifiers.
- The properties of AdaBoost are sensitive for the data that are noisy and outliers.
- Overfitting problem is less susceptible in some problem than other algorithms of learning.
- The learners of individual are weak, but as long as each one's performance is slightly better than guessing randomly, after all the final model will prove the converge to a strong learner.



Function boosting implemented method the AdaBoost.M1 algorithm. AdaBoost.M1 is simple generalizations of AdaBoost for more than two classes.

```
#install.packages('adabag')
library(adabag)
```

```
gc.boo <- boosting(credit_risk~., gc.Train,mfinal=50)
gc.boo$importance
```

```
gc.boo.pred=predict(gc.boo, newdata=gc.Test, type="class")
gc.boo.pred$confusion
gc.boo.pred$class
```

```
table(gc.boo.pred$class,gc.Test$credit_risk)
acc(gc.boo.pred$class,gc.Test$credit_risk)
roc.function(gc.boo.pred$prob[,2],gc.Test$credit_risk)
```

Last modified: środa, 17 stycznia 2024, 9:29

Accessibility settings

Przetwarzanie danych osobowych

Platformą administruje Komisja ds. Doskonalenia Dydaktyki wraz z Centrum Informatyki Uniwersytetu Łódzkiego Więcej

Informacje na temat logowania

Na platformie jest wykorzystywana metoda logowania za pośrednictwem Centralnego Systemu Logowania.

Studentów i pracowników Uniwersytetu Łódzkiego obowiązuje nazwa użytkownika i hasło wykorzystywane podczas logowania się do systemu USOSweb.

Deklaracja dostępności