

ksvm

KWONHYUNJIN

2018년 9월 21일

#e1071: svm 을 위한 패키지

```
bc <- read.xlsx("New_version_breast_cancer.xlsx",1)
head(bc)
```

```
##      NA. age tumor.size inv.nodes deg.malign irradiat   Class lt40 ge40
## 1    1   4         4         1       3         0   recur    0   0
## 2    2   5         4         1       1         0  norecur    0   1
## 3    3   5         8         1       2         0   recur    0   1
## 4    4   4         8         1       3         1  norecur    0   0
## 5    5   4         7         2       2         0   recur    0   0
## 6    6   5         6         2       2         1  norecur    0   0
##      premeno node.capse no.node.capse breast.left breast.right quad.cen
## 1          1         1         0         0         1         0
## 2          0         0         1         0         1         1
## 3          0         0         1         1         0         0
## 4          1         1         0         0         1         0
## 5          1         1         0         1         0         0
## 6          1         0         1         0         1         0
##      quad.Ll quad.Lu quad.Rl quad.Ru
## 1          0         1         0         0
## 2          0         0         0         0
## 3          1         0         0         0
## 4          1         0         0         0
## 5          0         0         0         1
## 6          0         1         0         0
```

#set.seed란 랜덤한 값을 시작하기 전에 사용하면 이후에도 같은 값으로 랜덤값을 갖는다. #동일한 랜덤값을 계속해서 받기위해 #과연 필요할지에 대해 고찰

```
set.seed(123)
bc_shuffle <- bc[sample(nrow(bc)), ]
head(bc_shuffle)
```

```
##      NA. age tumor.size inv.nodes deg.malign irradiat   Class lt40 ge40
## 80    80   3         1         1       2         0  norecur    0   0
## 218   218  3         5         1       3         1   recur    0   0
## 113   113  4         8         4       2         1  norecur    0   0
## 242   242  5         6         1       1         0  norecur    0   0
## 257   257  5         1         1       2         0  norecur    0   1
## 13    13   5         7         1       1         0  norecur    0   1
##      premeno node.capse no.node.capse breast.left breast.right quad.cen
## 80          1         0         1         0         1         1
## 218         1         0         1         1         0         0
## 113         1         1         0         0         1         0
## 242         1         0         1         1         0         0
## 257         0         0         1         1         0         1
## 13         0         0         1         0         1         1
##      quad.Ll quad.Lu quad.Rl quad.Ru
## 80          0         0         0         0
## 218         0         1         0         0
## 113         0         1         0         0
## 242         1         0         0         0
## 257         0         0         0         0
## 13         0         0         0         0
```

#현재 bc 데이터 프레임에 id라는 컬럼이 필요없다. #bc_shuffle에서 1번컬럼은 제외하고 나머지 컬럼을 bc2에 할당한다.

```
bc2 <- bc_shuffle[-1]
head(bc2)
```

```
##      age tumor.size inv.nodes deg.malign irradiat   Class lt40 ge40 premeno
## 80     3         1         1       2         0  norecur    0   0         1
## 218    3         5         1       3         1   recur    0   0         1
## 113    4         8         4       2         1  norecur    0   0         1
## 242    5         6         1       1         0  norecur    0   0         1
## 257    5         1         1       2         0  norecur    0   1         0
## 13     5         7         1       1         0  norecur    0   1         0
##      node.capse no.node.capse breast.left breast.right quad.cen quad.Ll
## 80            0         1         0         1         1         0
## 218            0         1         1         0         0         0
## 113            1         0         0         1         0         0
## 242            0         1         1         0         0         1
## 257            0         1         1         0         1         0
## 13            0         1         0         1         1         0
##      quad.Lu quad.Rl quad.Ru
## 80          0         0         0
## 218          1         0         0
## 113          1         0         0
## 242          0         0         0
## 257          0         0         0
## 13          0         0         0
```

#Train 과 Test 를 9:1로 나눈다. #bc_n은 shuffle 을 한 bc2 값에서 label 값을 빼고 정규화 시킨값

```
train_num<-round(0.9*nrow(bc2),0)
bc_train<-bc2[1:train_num,]
bc_test<-bc2[(train_num+1):nrow(bc2),]
```

#ksvm()함수는 가우시안 커널을 기본으로 하며 **vanilladot** 은 특별한 변환이 내적 계산

```
Class_classifier <-ksvm(Class~.,data=bc_train,kernel="vanilladot")
```

```
## Setting default kernel parameters
```

```
Class_classifier
```

```
## Support Vector Machine object of class "ksvm"
##
## SV type: C-svc (classification)
## parameter : cost C = 1
##
## Linear (vanilla) kernel function.
##
## Number of Support Vectors : 151
##
## Objective Function Value : -132.8858
## Training error : 0.261044
```

```
Class_predictions <- predict(Class_classifier,bc_test)
```

```
confusionMatrix(table(Class_predictions, bc_test$Class))
```

```
## Confusion Matrix and Statistics
##
##
## Class_predictions norecur recur
##           norecur      14      10
##           recur       1       3
##
##              Accuracy : 0.6071
##              95% CI : (0.4058, 0.785)
##      No Information Rate : 0.5357
##      P-Value [Acc > NIR] : 0.28637
##
##              Kappa : 0.172
##  Mcnemar's Test P-Value : 0.01586
##
##              Sensitivity : 0.9333
##              Specificity : 0.2308
##              Pos Pred Value : 0.5833
##              Neg Pred Value : 0.7500
##              Prevalence : 0.5357
##              Detection Rate : 0.5000
##      Detection Prevalence : 0.8571
##              Balanced Accuracy : 0.5821
##
##              'Positive' Class : norecur
##
```