ANN

KWONHYUNJIN

2018년 9월 20일

#nnet: ANN 패키지

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

```
##   NA. age tumor.size inv.nodes deg.malig 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

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

```
##       NA. age tumor.size inv.nodes deg.malig 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.malig 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        1       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

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

#size = 2는 hidden layer가 2층 이라는 뜻 #decay = 5e-04는 overfitting을 피하기 위해 사용하는 weight decay parametem, default 값은 0 #maxit = 200은 200번 반복

```
model <- nnet(Class ~ ., data = bc_train, size = 2, decay = 5e-04, maxit = 200)
```

```
## # weights:  39
## initial  value 157.490006
## iter  10 value 133.289573
## iter  20 value 126.193855
## iter  30 value 122.276419
## iter  40 value 121.593436
## iter  50 value 121.248434
## iter  60 value 120.622451
## iter  70 value 120.480613
## iter  80 value 120.412198
## iter  90 value 120.358022
## iter 100 value 120.355621
## iter 110 value 120.326474
## iter 120 value 119.871678
## iter 130 value 119.735313
## iter 140 value 119.734936
## iter 150 value 119.734663
## iter 160 value 119.734548
## final  value 119.734492
## converged
```

#분류를 얻을때는 type="class"를 지정해야하지만, 기본 값이 class이므로 생략 가능

```
pred <- predict(model,bc_test,type="class")
table(pred,bc_test$Class)
```

```
##
## pred      norecur recur
##   norecur      13     4
##   recur         2     9
```

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

```
## Confusion Matrix and Statistics
##
##
## pred      norecur recur
##   norecur      13     4
##   recur         2     9
##
##                Accuracy : 0.7857
##                  95% CI : (0.5905, 0.917)
##     No Information Rate : 0.5357
##     P-Value [Acc > NIR] : 0.0057
##
##                   Kappa : 0.5648
##  Mcnemar's Test P-Value : 0.6831
##
##             Sensitivity : 0.8667
##             Specificity : 0.6923
##          Pos Pred Value : 0.7647
##          Neg Pred Value : 0.8182
##              Prevalence : 0.5357
##          Detection Rate : 0.4643
##    Detection Prevalence : 0.6071
##       Balanced Accuracy : 0.7795
##
##        'Positive' Class : norecur
##
```