

## Wk12-2 : 의사결정나무 (Decision Tree) II

## 2. 의사결정나무 – rpart 패키지

- 의사결정나무 실행 패키지: rpart, party 패키지 (tree패키지 외 사용)

```
# lec12_2_tree.R
# Decision tree
# use package rpart and party

# other package for tree
install.packages("rpart")
install.packages("party")
library(rpart)
library(party)

#package for confusion matrix
install.packages("caret")
library(caret)
```

```
#decision tree : use rpart package
help("rpart")
```



### Recursive Partitioning and Regression Trees

#### Description

Fit a rpart model

#### Usage

```
rpart(formula, data, weights, subset, na.action = na.rpart, method,
      model = FALSE, x = FALSE, y = TRUE, parms, control, cost, ...)
```

#### Arguments

formula	a <a href="#">formula</a> , with a response but no interaction terms. If this is a data frame, that is taken as the model frame (see <a href="#">model.frame</a> ).
data	an optional data frame in which to interpret the variables named in the formula.
weights	optional case weights.
subset	optional expression saying that only a subset of the rows of the data should be used in the fit.

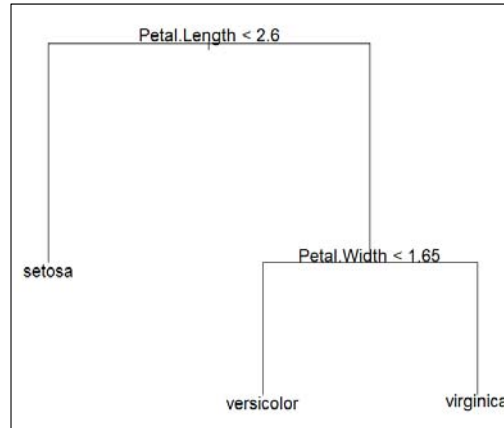
## 2. 의사결정나무 – rpart 패키지

12.2 Decision Tree

- 의사결정나무 함수 : rpart (종속변수~x1+x2+x3+x4, data= )

```
# tree using rpart
c11<-rpart(Species~., data=train)
plot(c11)
text(c11, cex=1)
```

rpart 함수는 가지치기를 해서 나온 결과  
-> 데이터에 따라 추가적인 가지치기가  
필요할 수도 있음



\* tree패키지에서 pruning한 결과와 동일

## 2. 의사결정나무 – rpart 패키지

12.2 Decision Tree

- rpart패키지는 과적합의 우려가 있으므로 pruning을 해줘야 함(iris의 경우 필요없음)
- printcp에서 xerror(cross validation error)의 값이 최소가 되는 마디를 선택.

```
#pruning (cross-validation)-rpart
printcp(c11)
plotcp(c11)
help(printcp)
```

```
> printcp(c11)

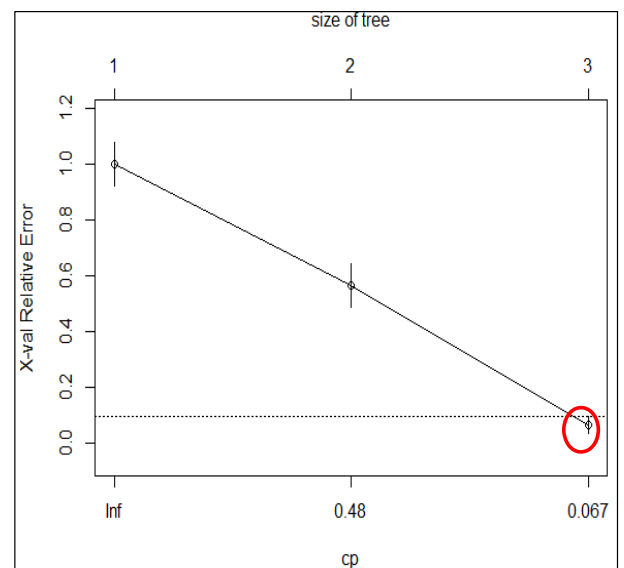
Classification tree:
rpart(formula = Species ~ ., data = train)

Variables actually used in tree construction:
[1] Petal.Length Petal.Width

Root node error: 62/100 = 0.62

n= 100

  CP nsplit rel error  xerror  xstd
1 0.50000    0 1.000000 1.000000 0.078288
2 0.45161    1 0.500000 0.564516 0.076931
3 0.01000    2 0.048387 0.064516 0.031606
```



## 2. 의사결정나무 – rpart 패키지

12.2 Decision Tree

- rpart결과에서 복잡도계수에 기반한 최적 가지치기

```
#pruning (cross-validation)-rpart  
printcp(c11)  
plotcp(c11)  
help(printcp)
```

printcp (rpart) R Documentat

Displays CP table for Fitted Rpart Object

**Description**

Displays the cp table for fitted rpart object.

**Usage**

```
printcp(x, digits = getOption("digits") - 2)
```

**Arguments**

**x** fitted model object of class "rpart". This is assumed to be the result of some function that produces an object with the same named components as that returned by the rpart function.

**digits** the number of digits of numbers to print.

**Details**

Prints a table of optimal prunings based on a complexity parameter.

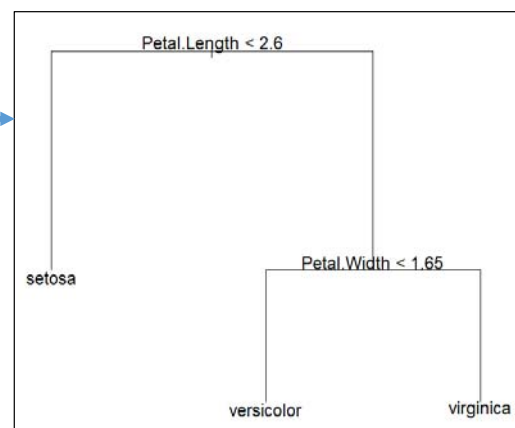
## 2. 의사결정나무 – rpart 패키지

12.2 Decision Tree

- rpart를 사용한 최종 tree모형 (iris data)

```
# final tree model  
pc11<-prune(c11, cp=c11$cptable[which.min(c11$cptable[, "xerror"]), "CP"])  
plot(pc11)  
text(pc11)
```

tree함수를 이용한 최종모형과 동일한 tree결과



## 2. 의사결정나무 – rpart 패키지

12.2 Decision Tree

### • 의사결정나무결과 정확도 : test data에 대한 정확도

```
#measure accuracy(rpart)
pred2<- predict(c11,test, type='class')
confusionMatrix(pred2,test$Species)
```



```
> pred2<- predict(c11,test, type='class')
> confusionMatrix(pred2,test$Species)
```

Confusion Matrix and Statistics

	Reference		
Prediction	setosa	versicolor	virginica
setosa	19	0	0
versicolor	0	17	1
virginica	0	2	11

Overall Statistics

Accuracy : 0.94

## 3. 의사결정나무 – party 패키지

12.2 Decision Tree

### ▪ 의사결정나무 실행 패키지: party 패키지 (tree패키지 외 사용)

```
help(ctree)
```

Conditional Inference Trees {party}

R Documentation

#### Conditional Inference Trees

##### Description

Recursive partitioning for continuous, censored, ordered, nominal and multivariate response variables in a conditional inference framework.

##### Usage

```
ctree(formula, data, subset = NULL, weights = NULL,
      controls = ctree_control(), xtrafo = ptrrafo, ytrafo = ptrrafo,
      scores = NULL)
```

##### Arguments

**formula** a symbolic description of the model to be fit. Note that symbols like : and - will not work and the tree will make use of all variables listed on the rhs of formula.

**data** a data frame containing the variables in the model.

**subset** an optional vector specifying a subset of observations to be used in the fitting process.

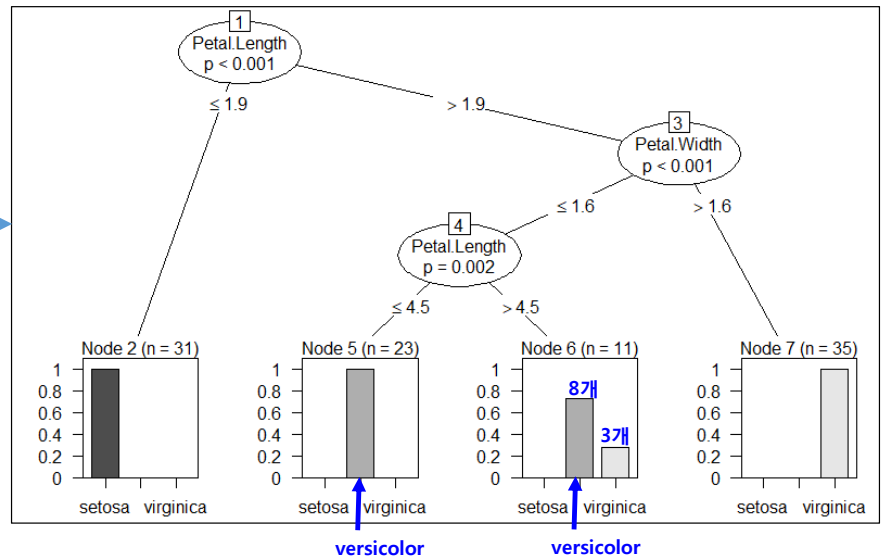
**weights** an optional vector of weights to be used in the fitting process. Only non-negative integer valued weights are allowed.

### 3. 의사결정나무 – party 패키지

12.2 Decision Tree

- 의사결정나무 함수 : `ctree` (종속변수~x1+x2+x3+x4, data= )

```
# decision tree using ctree
partymod<-ctree(Species~.,data=train)
plot(partymod)
```



### 3. 의사결정나무 – party 패키지

12.2 Decision Tree

- party 패키지를 이용한 결과

```
partymod<-ctree(Species~.,data=train)
plot(partymod)
# to see the result using party
partymod
```

```
> partymod

Conditional inference tree with 4 terminal nodes

Response: Species
Inputs: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width
Number of observations: 100

1) Petal.Length <= 1.9; criterion = 1, statistic = 92.568
2)* weights = 31
1) Petal.Length > 1.9
3) Petal.Width <= 1.6; criterion = 1, statistic = 46.625
4) Petal.Length <= 4.5; criterion = 0.998, statistic = 12.436
5)* weights = 23
4) Petal.Length > 4.5
6)* weights = 11
3) Petal.Width > 1.6
7)* weights = 35
```

### 3. 의사결정나무 – party 패키지

- 의사결정나무결과 정확도 : test data에 대한 정확도 (party 패키지 사용)

```
#measuring accuracy(party)
party_pred<-predict(party_mod,test)
confusionMatrix(party_pred,test$Species)
```

```
> party_pred<-predict(party_mod,test)
> confusionMatrix(party_pred,test$Species)
Confusion Matrix and Statistics
```

	Reference		
Prediction	setosa	versicolor	virginica
setosa	19	0	0
versicolor	0	17	1
virginica	0	2	11

Overall Statistics

Accuracy : 0.94

