**LAB 2: Related Tree-Based Ensemble Method**

**Objectives:**

**Materials**

|  |  |
| --- | --- |
| **File Name** | **Description** |
| *classification\_tr.csv* | Calibration dataset for classifying 9 land cover classes |
| *classification\_va.csv* | Validation dataset for classifying 9 land cover classes |
| *GOCI\_ocean\_cali.csv* | Calibration dataset for estimating surface seawater *f*CO2 |
| *GOCI\_ocean\_vali.csv* | Validation dataset for estimating surface seawater *f*CO2 |
| *Extra tree in R.txt* | Extra Tree algorithm code in R |
| *Extra tree in Python.py* | Extra Tree algorithm code in Python |
| *rotForest in R\_classificaiton.txt* | Rotation Forest algorithm code in R for classification |

**PART I: Extra Tree in R**

**Task 1. Setup environment**

1. Setup environment

* R을 켜서 CRAN 미러 설정 후 작업 디렉토리 변경

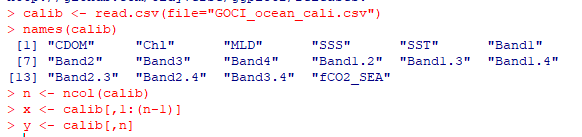
**Task 2. Build Extra Tree in R**

1. Install packages : (##Installing, ##Loading ExtraTrees)

* 사용할 package들을 설치 후 불러온다
* ‘extraTrees’ 패키지 사용을 위해서는 컴퓨터의 Java와 R의 호환성을 맞춰줘야 한다. (ex. 둘 다 64 bit or 32 bit)

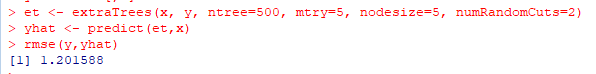
1. Read data file (## Read the data file)

* Data 파일을 읽어온 뒤 input parameter와 target을 설정해준다.



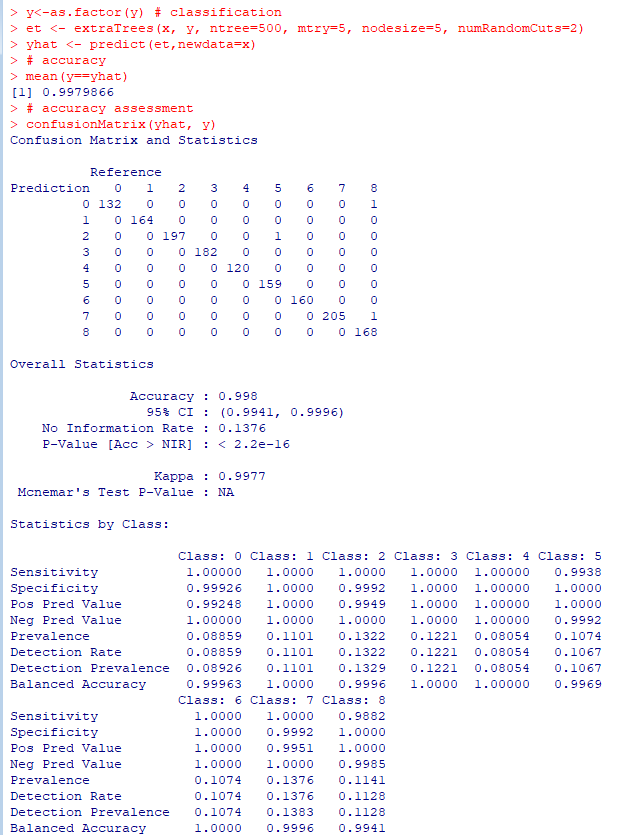
1. Build Extra Tree with regression (## Regression with ExtraTrees)

* 원하는 option 설정 후 regression tree를 설정한다. (변수 설명은 PDF 참고)
* Print RMSE



1. Build Extra Tree with classification (## Classification with ExtraTrees)

* 원하는 option 설정 후 classification tree를 설정한다. (변수 설명은 PDF 참고)
* Print accuracy and accuracy assessment



1. Validation (## Validation)

* Validation 결과 산출

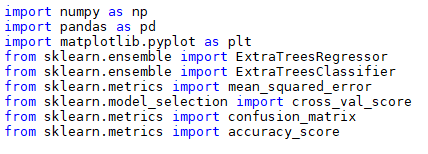
1. Save (## Save results)

* 필요에 따라 모델과 결과를 저장한다.

**PART II: Extra Tree in Python**

**Task 1. Setup environment**

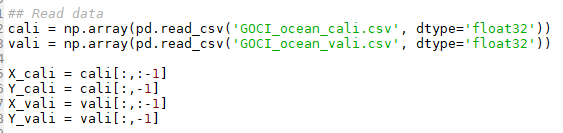
1. Setup environment



**Task 2. Build Extra Tree in Python**

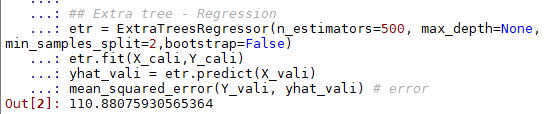
1. Read data:

* Data 읽어오기



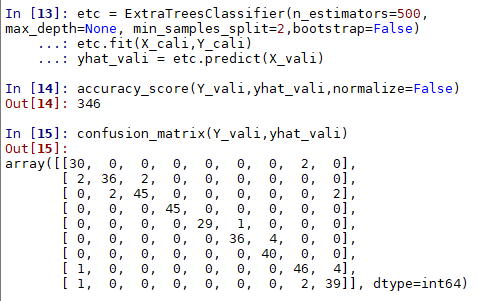
1. Build Extra Tree with regression

* 원하는 option 설정 후 regression tree를 설정한다. (변수 설명은 PDF 참고)
* Print error



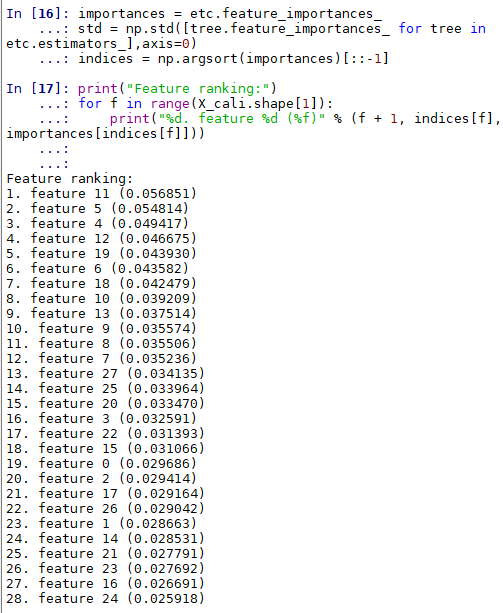
1. Build Extra Tree with classification

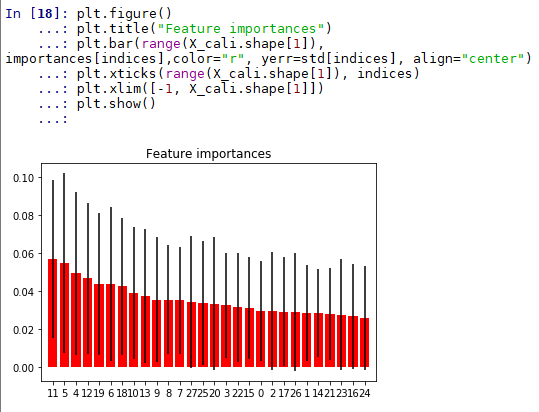
* 원하는 option 설정 후 classification tree를 설정한다. (변수 설명은 PDF 참고)
* Print error



1. Feature importance

* 변수 중요도 출력
* 빨강색 bar는 feature importance, 검은선은 inter-tree variability를 의미





**PART Ⅲ: Rotation Forest in R**

**Task 1. Setup environment**

1. Setup environment

* R을 켜서 CRAN 미러 설정 후 작업 디렉토리 변경

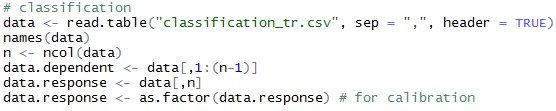
**Task 2. Build Extra Tree in R**

1. Install packages : (## 1, ## 2)

* 사용할 package들을 설치 후 불러온다

1. Read data file (## 3)

* Data 파일을 읽어온 뒤 input parameter (data.dependent)와 target (data.response)을 설정해준다.



1. Build Rotation Forest with classification (## 4, ## 5)

* 원하는 option 설정 후 rotation forest를 실행한다
* rotationForest(x, y, K, L):
* K: # subsets of attributes, L: # trees

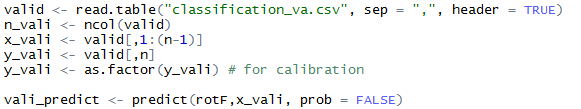


* Print overall accuracy and confusion matrix



1. Validation (## 6)

* Validation 결과 산출



* Print overall accuracy and confusion matrix of validation dataset



**Assignment**

1. PART I , PART II , PART Ⅲ에 나온 코드들을 실행해보고, 각 모델 파라미터가 주는 영향을 파악하기
2. 지난주에 배운 Decision Tree, Random Forests와 결과 비교해보기