3. Unsupervised Learning exercise

Unsupervised Learning Challenge

- due date : **Sat. 10/12 11:59pm**
- 이 과제는 2학기 프로젝트 조 편성 시 반영됩니다.

이 사이트들을 참고하세요

- Train Test split
  https://rpubs.com/ID_Tech/S1
- Hierarchical clustering
  https://www.r-bloggers.com/how-to-perform-hierarchical-clustering-using-r/
- Kmeans
  https://rpubs.com/jmhome/K-means
  https://stackoverflow.com/questions/29605911/r-k-means-algorithm-custom-centers
  https://stackoverflow.com/questions/49016343/what-package-to-use-in-r-for-kmeans-prediction

---

Data setting

```
set.seed(2946)
```

- iris data

```
data(iris)
summary(iris)
```

```
head(iris)
```

- Normalize the non-target variables
- split train, test set

Agglomerative clustering

- 여러 linkage를 활용하여 dendrogram을 그리시오.
- Single Linkage, Complete Linkage, Average Linkage, Centroid Linkage, Ward's method

```
# Dissimilarity matrix
d <- dist(train[,-5], method="euclidean")
```

- Ward's method 결과의 misclassification error를 구하시오.

```
# Get misclassification error
ward_error = 0
print(paste("ward error : ", round(ward_error, 4), sep=""))
```

- Visualize the result

K-means Clustering

- get initial centroid from ward's method
- build model with initial centroid
- compare two centroids
- Get misclassification error
- Visualize clustering result

K-means prediction

```
predict.kmeans <- function(object, newdata){
  centers <- object$centers
  n_centers <- nrow(centers)
  dist_mat <- as.matrix(dist(rbind(centers, newdata)))
  dist_mat <- dist_mat[-seq(n_centers), seq(n_centers)]
  max.col(-dist_mat)
}
```

- make crosstable between predictino label and real label
- get misclassification error of test data
- visualize clustering result

END