

Assignment #9

머신러닝 이론과 실전

Due: November 15, 2021

1. LDA (linear discriminant analysis) 방법을 이용하여 앙상블 방법을 구현해 보고, 변수중요도를 계산하고자 합니다.
 - A. 데이터는 training 으로 'veh.dat'과 test 로 'vehtest.dat'를 사용합니다.
 - B. 기본 분류함수인 LDA 방법의 함수는 python 의 'sklearn.discriminant_analysis.LinearDiscriminantAnalysis'를 사용합니다.
 - C. 앙상블 방법으로는 Bagging 과 Random feature ensemble 방법을 구현합니다. (직접 코딩 필요함)
 - D. 앙상블내 분류모형의 갯수는 B=101 을 사용합니다.
 - E. Random feature ensemble 을 사용할 때는 $m = \frac{p}{2}$ 을 사용합니다.
 - F. OOB 데이터를 이용한 변수중요도를 계산합니다.
 - G. 정확도 결과물은 test 데이터의 결과만 보입니다.
 - H. 결과물 예시는 아래에 있습니다.

(1) LDA - bagging

```
Variable Importance:
X1: 23
X2: 45
X3: 10

Confusion Matrix (LDA - bagging)
-----
                Predicted Class
                1         2
Actual   1      239      14
Class    2       12     153

Model Summary (LDA - bagging)
-----
Overall accuracy = .793
```

(2) LDA - random feature ensemble

```
Variable Importance:
X1: 23
X2: 45
X3: 10

Confusion Matrix (LDA - random feature)
-----
                Predicted Class
                1         2
Actual   1      239      14
Class    2       12     153

Model Summary (LDA - random feature)
-----
Overall accuracy = .793
```