

# Week 5 복습과제

## 1. Linear Discriminant Analysis

### LDA (Linear Discriminant Analysis)

- direct하게 확률을 계산하는 방법
- $G$ : class variable.  $k = 1, \dots, K$   
설명변수  $X$
- $P(X|G=k)$  가 다양한 정보들을  
포함하는 가정
- 필요한 확률 =  $P(G=k|X)$   
데이터가 들어왔을 때 이 데이터의 classification  
부터  $k$ 까지 확률 계산  $\rightarrow$  가장 큰 값으로 분류
- Bayes Rule을 이용해서 확률계산이 가능하다!

from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis

lda = LinearDiscriminantAnalysis()

## 2. ExtraTreesClassifier

### ExtraTreesClassifier

- RandomForest의 변종:  
각 후보특성을 무작위로 분할  $\rightarrow$  무작위성 증가  
splitter = 'best' 가 아니라 splitter = 'random' 사용
- RandomForest 보다 특성의 중요도 높게 평가  
= 더 특성은 시각으로 평가한다

from sklearn.ensemble import ExtraClassifier