

## 주제분석 3주차 패키지

- 분석 툴은 R/Python 둘 다 가능합니다. 주제분석 1-3주차 패키지 문제의 조건 및 힌트는 Python을 기준으로 하지만, R을 사용해도 무방합니다.
- 제출형식은 HTML, PDF 모두 가능합니다. **.ipynb** 이나 **.R** 등의 **소스코드 파일은 불가능합니다**. 파일은 [psat2009@naver.com](mailto:psat2009@naver.com)으로 보내주세요.
- 패키지 과제 발표는 세미나 쉬는 시간 후에 하게 되며, 랜덤으로 5시 00분에 발표됩니다.
- 제출기한은 **목요일 자정까지** 이고, 지각 시 벌금 5000원, 미제출시 10000원입니다. 패키지 무단 미제출 2회 시 퇴출이니 유의해주세요.

✓ 이번 모델링에서 모든 씨드(Seed)는 2728로 맞추어 주세요

### Chapter 1. SVM

파이썬의 Scikit-learn 라이브러리는 대표적인 머신러닝 라이브러리로 여러가지 모델링을 편하게 할 수 있습니다. 이번 챕터에는 SVM(서포트벡터머신)을 해보겠습니다.

**문제1.** data1.csv항목을 불러와주세요

**문제2.** target의 setosa를 0으로 versicolor를 1로, virginica를 2로 바꾸어 주세요.

- ✓ X는 sepal length와 sepal width만 사용할 예정입니다.
- ✓ SVM은 스케일링을 해야하지만 이번 패키지에서는 뒤에 Decision Boundary를 그릴 것이기 때문에 그냥 넘어가겠습니다.

**문제3.** Test셋과 Train셋을 나눠주세요. (Test셋의 비율은 0.3입니다)

**문제4.** SVM의 하이퍼 파라미터에 대해서 간단히 적어주세요.

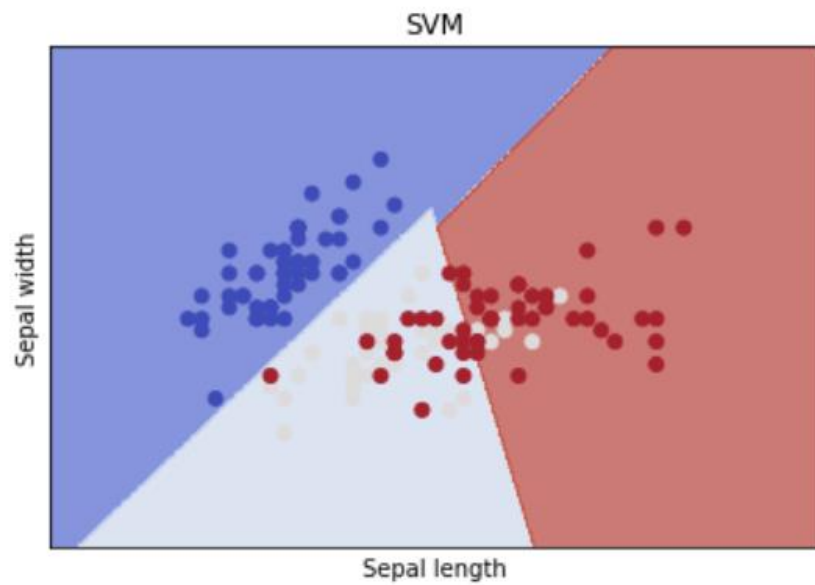
**문제5.** 하이퍼 파라미터 튜닝을 해주세요. (서칭할 하이퍼 파라미터는 자유롭게 정하셔도 됩니다.)

- ✓ 예시 : kernel (linear), C (0.01, 0.1, 1, 10, 30, 50, 70)
- ✓ 5-fold CV
- ✓ 평가지표 : Accuracy
- ✓ GridSearchCV함수사용하면 편합니다.

문제6. 최적 파라미터 조합과 결과를 출력하세요.

문제7. 테스트셋을 예측한 뒤 결과를 출력하세요.

문제8. 다음과 같이 Decision Boundary를 그려보세요.



## Chapter 2 LGBM

LGBM은 Gradient Boosting Model로, tree를 기반으로 학습하는 알고리즘입니다.  
이번에는 LGBM을 모델링해봅시다.

문제1. data2.csv를 불러와주세요.

문제2. Test셋과 Train셋을 나눠주세요. Test셋의 비율은 0.3입니다)

문제3. LGBM 파라미터에 대해서 간단히 적어주세요.

문제4. LGBM 하이퍼 파라미터 튜닝을 해주세요. (서칭할 하이퍼 파라미터는 자유롭게 정하셔도 됩니다.)

- ✓ 예시 : max\_depth (10, 20), min\_child\_samples (20, 40, 60, 80), num\_leaves (60, 70, 80)
- ✓ 5-fold CV
- ✓ 평가지표 : RMSE
- ✓ GridSearchCV 함수사용하면 편합니다

문제5. 최적의 파라미터와 결과를 출력하세요.

문제6. 테스트셋에 대하여 예측한 후 결과를 출력하세요.

## Chapter3 LSTM

이번에 사용할 모델은 LSTM입니다. Keras는 딥러닝 모델을 간편하게 만들고 훈련시킬 수 있는 파이썬 라이브러리입니다. 딥러닝 모델을 간단히 구현해보는 것이 목표입니다.

**문제 1.** Samsung.csv를 불러와주세요.

**문제 2.** 데이터 형식을 확인해보고, 수치형으로 바꾸어 주세요.

- Price, Open, High, Low만 사용할 예정입니다.
- y는 Price입니다.
- 1-step 예측 모델링을 할 예정입니다.

**문제3.** 테스트 데이터를 분리해주세요. ( 테스트 데이터는 최근 60일입니다. )

**문제4.** Train과 Test셋을 모두 스케일링 시켜주세요.

- MinMaxScaler사용

**문제5.** Window사이즈에 맞게 Train과 테스트 셋을 구성해주세요. (아래의 그림을 참고하여 X와 y데이터를 구성해주세요; Window\_size = 20)



- ✓ HINT : X는 ( i:i+window\_size ) 만큼 y는 ( i + window\_size)만큼 잘라 데이터를 구성하는 for문을 작성하세요

**문제5.** LSTM 딥러닝 모델을 구성해주세요. ( Layer수, Unit수 등을 자유롭게 변형시켜보세요! )

**문제6.** 모델을 적합시킨 후 결과를 확인해보세요.

- ✓ 튜닝까지 하실 필요는 없습니다. 한두번 해본 후 잘 나오는 결과로 제출해주세요!

**문제7.** 테스트데이터에 대해서 예측 후 결과를 아래와 같이 시각화 해보세요.

