

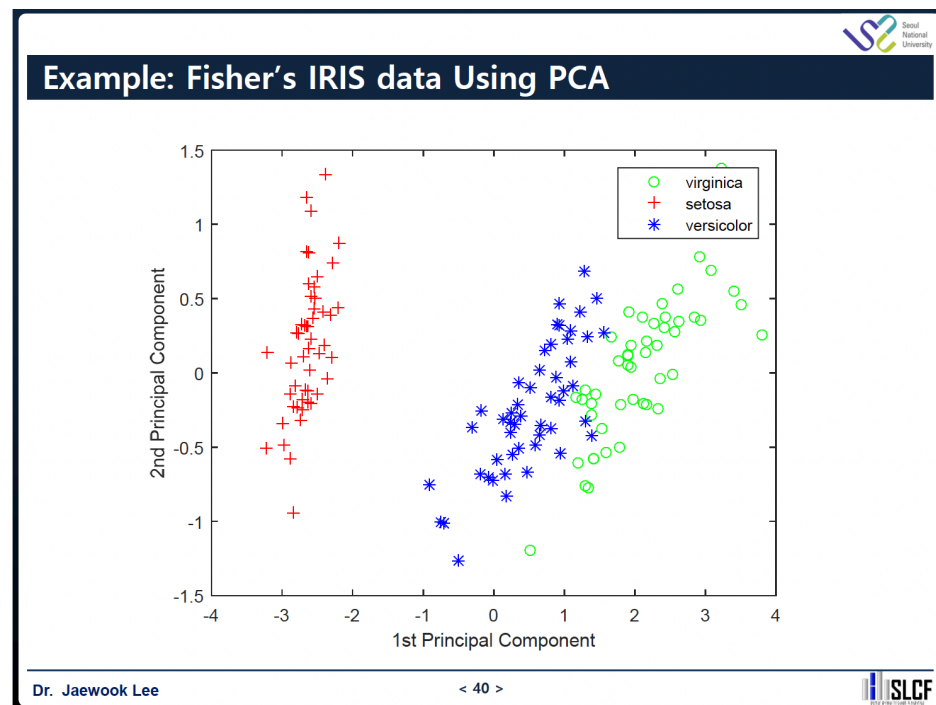
[서울대 AI빅데이터 과정] 이재욱 교수님 프로젝트 2

프로젝트 2

- 수업 시간에 배운 PCA를 실습해보는 프로젝트입니다.
- 조별로 하나의 ipynb 파일을 제출해주시면 되고, 이름은 0조_project2.ipynb 로 jinseong@snu.ac.kr로 제출해주시길 바라며, 기간은 10월 22일(금)까지 제출해주시면 됩니다.

프로젝트 설명

- [https://github.com/SLCFLAB/Math4DS/blob/main/ch05-Covariance and PCA_supp.ipynb](https://github.com/SLCFLAB/Math4DS/blob/main/ch05-Covariance%20and%20PCA_supp.ipynb) 파일을 참고하여 PCA를 python, sklearn을 통해 활용하는 방법을 조별로 배웁니다.
- 그리고 같은 파일의 2. PCA(Principal Component Analysis) 부분(실행 cell 기준 13-23, Logistic Regression Classifier with PCA 이전까지)을 다른 데이터셋(iris dataset)에 대해서 구현해보는 것이 이번 프로젝트의 목표입니다. EDA ppt 40페이지의 그림을 참고해서 이와 같은 결과가 나오는지 비교해주세요.



- covariance matrix, eigenvalue 등을 직접 구하는 방법과 scikit-learn의 PCA()를 활용하는 두 가지 방법 모두 제출하여야 합니다.
- 데이터를 받아오는 부분, 데이터를 간단하게 전처리하는 부분, pca를 통해 나온 값들을 plotting해주시면 됩니다!

프로젝트 참고

- iris dataset은 sklearn에서 제공하는 기본적인 데이터로, pca에 많이 활용되는 꽃과 관련된 데이터셋입니다. (https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html)
- 데이터셋을 loading하는 부분은 다음을 참고해주세요.

```
from sklearn import datasets

# Loading the dataset
iris = datasets.load_iris()
X = iris.data
y = iris.target
```

프로젝트를 하다 궁금한 점이나 해결되지 않는 부분이 있으면 조교 박진성(jinseong@snu.ac.kr)로 보내주세요!