

파이썬 설치와 빅데이터 분석

by 김안정

2024.09.13

포 159

목표

- 파이썬 프로그램의 구조와 특징을 이해하고 직접 설치해본다.
- 주피터 노트북의 개념이 이해하고 코랩을 활용할 줄 안다.
- 파이썬으로 통계분석을 하기 위한 기초 작업에 대해 이해한다.

강의자료 및 실습 데이터:

파이썬이란?

파이썬은 웹 애플리케이션과 소프트웨어 개발, 데이터 사이언스와 머신 러닝을 위한 도구도 제공하는 프로그래밍 언어

- 프로그래밍 언어 중에서 이해하기 쉬움
- 다양한 운영체제에서 작동함
- 코드를 한줄씩 실행한다. (인터프리터 방식)

라이브러리

라이브러리는 코드를 처음부터 작성할 필요가 없도록 파이썬 프로그램에 포함시킬 수 있는 자주 사용되는 코드 모음

데이터 분석과 머신 러닝에 필요한 라이브러리

- Matplotlib - 데이터 시각화를 도움
- Pandas - 데이터 구조를 다루는 작업에 최적화
- Numpy
- Keras

파이썬 vs R

파이썬 설치하기 1 -

윈도우 환경에서 파이썬 설치 방법

1. exe 파일을 설치한 후 IDE(통합 개발 환경) 프로그램으로 실행
 - ide란 한 곳에서 코드 작성, 편집, 테스트, 디버깅 등 필요한 도구를 사용자에게 제공하는 별도의 소프트웨어
 - Python의 경우 PyCharm, VScode 등을 사용
2. 아나콘다 설치 후 아나콘다, 주피터 노트북으로 실행
3. 구글 Colab 이용하기

파이썬 설치하기 2 - 아나콘다 설치하기

아나콘다란?

Python 통합 패키지, 파이썬을 포함한 많은 모듈을 포함

- 주피터 노트북 제공됨

설치하기

1. 공식 홈페이지에서 다운로드 <https://www.anaconda.com/download>
2. 설치하기

파이썬 설치하기 - 구글 코랩 이용하기 3

- 구글 Colab
- 주피터 노트북 같은 인터프리터와 실행 기능을 온라인에서 제공함
- Colab의 구조

파이썬 이용하기

구글 코랩에서 다음을 입력하고 실행해 봅니다.

```
print("Hello World")
```

파이썬 적응하기 : 변수

변수란?

숫자 변수

```
a=1  
print(a)
```

숫자 변수의 연산

```
a=1  
b=2  
print(a+b)
```

```
c=3
```

배열

```
aa=[1,2,3]  
print(aa)
```

```
print(aa[1])
```

```
bb=[[1,2,3],[4,5,6]]  
print(bb)
```

array는 뭘까요?

```
cc=np.array(bb)  
print(cc)
```

라이브러리란?

데이터로 실습해보기

파일 불러오기

데이터 프레임이란? (배열과 어떤 점이 다를까요?)

```
import pandas as pd

x1=np.ones((1,100))
print(x1)
x1=x1.reshape(100,1)
df=pd.DataFrame(x1)
print(df)
```

데이터 프레임 생성해보기

불러온 데이터 확인하기

```
print(df)
df.head(10)
```

기초 통계량

시각화

```
x=df.loc[:, 'SAT']
y=df.loc[:, 'GPA']
print(x)
```

파이썬에서의 OLS

```
from statsmodels.formula.api import ols # statsmodel 라이브러리 불러오기

model= ols(formula='SAT~GPA',data=df).fit()
print(model.summary())
```

```
print('parameters',model.params)
```

```
print('R2',model.rsquared)
```


머신러닝을 해보자

Thank you!