



작업 표시줄 표시



디스플레이 설정 ▼



슬라이드 쇼 마침

0:00:07



오전 12:03



CUAI BASIC 스터디 1조

2022.03.21

발표자 : 정서현, 김지호



슬라이드 1/21



다음 슬라이드



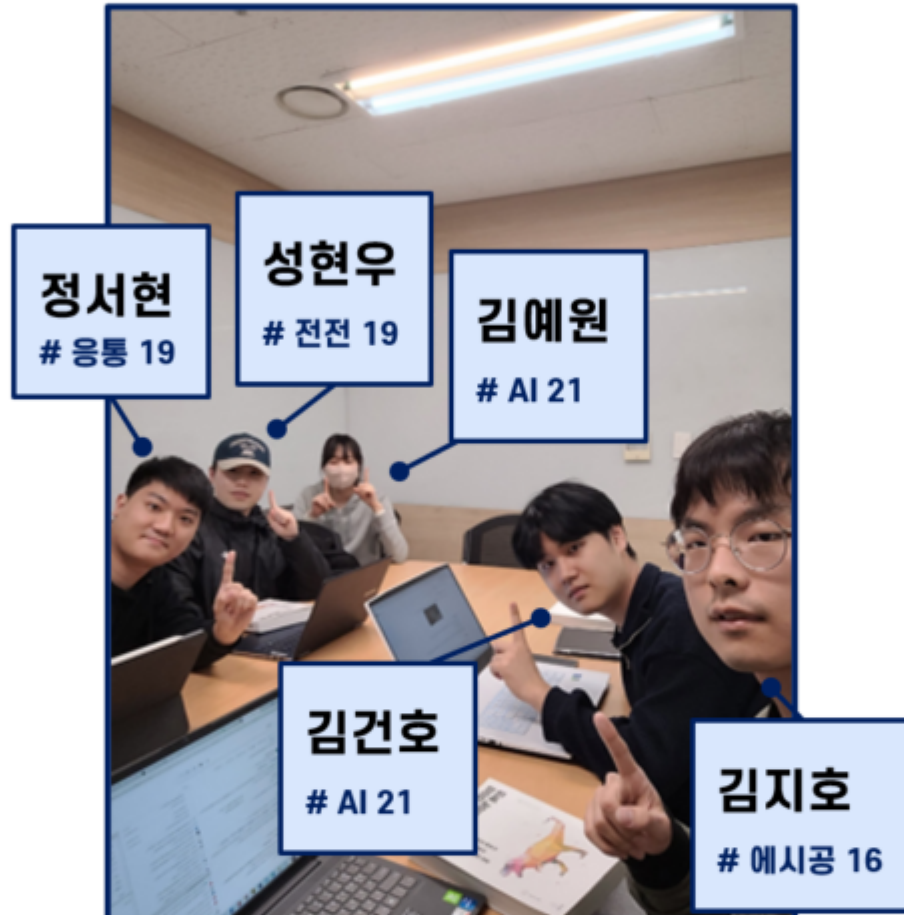
스터디원 소개 및 만남 인증

안녕하세요. 이번 CUAIBASIC 스터디 1조 발표를 맡게 된 정서현이라고 합니다. (스페이스바)

A⁺

A⁺

스터디원 소개 및 만남 인증



순번 1
분관 서울 학술정보원
유형 창의문화 스터디룸
공간 [스터디룸10 \(창의문화\)](#)
동반이용자수 3명
예약일시 2023/03/16 18:00 ~ 20:00
상태 예약됨

수정

예약취소

목차

01. 사이킷런 소개와 특징
02. 붓꽃 품종 예측하기
03. 사이킷런 기반의 프레임워크
04. Model Selection
05. 데이터 전처리
06. 각자 관심사 토의

(스페이스바)저희 1조는 응용통계학과인 저와 전기전자학과의 성현우님, 에너지시스템공학과의 김지호님, ai학과의 김예원님과 김건호님으로 이루어져 있고 (스페이스바) 저희 1조는 매주 목요일 오후 6시부터 8시까지 스터디를 진행하고 있습니다. (스페이스바)



0:00:40 || ↺

오전 12:04

목차

- 01. 사이킷런 소개와 특징
- 02. 붓꽃 품종 예측하기
- 03. 사이킷런 기반의 프레임워크
- 04. Model Selection
- 05. 데이터 전처리
- 06. 각자 관심사 토의



다음 슬라이드

01. 사이킷런 소개와 특징

이번 발표의 목차는 다음과 같습니다. (스페이스바)



슬라이드 3/21



A⁺

A⁺

01. 사이킷런 소개와 특징




다음 슬라이드

02. 붓꽃 품종 예측하기

우선, 사이킷런에 대해 간략히 소개해드리겠습니다.
(스페이스바) 사이킷런은 파이썬 기반의 머신러닝을 위한
가장 쉽고 효율적인 개발 라이브러리 중 하나입니다. 아래
그림에서 확인할 수 있듯이, 이 라이브러리는 대표적인
머신러닝 라이브러리 중 하나입니다. (스페이스바)

02. 붓꽃 품종 예측하기


Iris 데이터 세트



Fisher's Iris Data [hide]

Dataset Order	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	<i>I. setosa</i>
2	4.9	3.0	1.4	0.2	<i>I. setosa</i>
3	4.7	3.2	1.3	0.2	<i>I. setosa</i>
4	4.6	3.1	1.5	0.2	<i>I. setosa</i>
...					
50	5.0	3.3	1.4	0.2	<i>I. setosa</i>
51	7.0	3.2	4.7	1.4	<i>I. versicolor</i>
52	6.4	3.2	4.5	1.5	<i>I. versicolor</i>
53	6.9	3.1	4.9	1.5	<i>I. versicolor</i>
54	5.5	2.3	4.0	1.3	<i>I. versicolor</i>
...					
100	5.7	2.8	4.1	1.3	<i>I. versicolor</i>
101	6.3	3.3	6.0	2.5	<i>I. virginica</i>
102	5.8	2.7	5.1	1.9	<i>I. virginica</i>
103	7.1	3.0	5.9	2.1	<i>I. virginica</i>
104	6.3	2.9	5.6	1.8	<i>I. virginica</i>
...					

02. 붓꽃 품종 예측하기

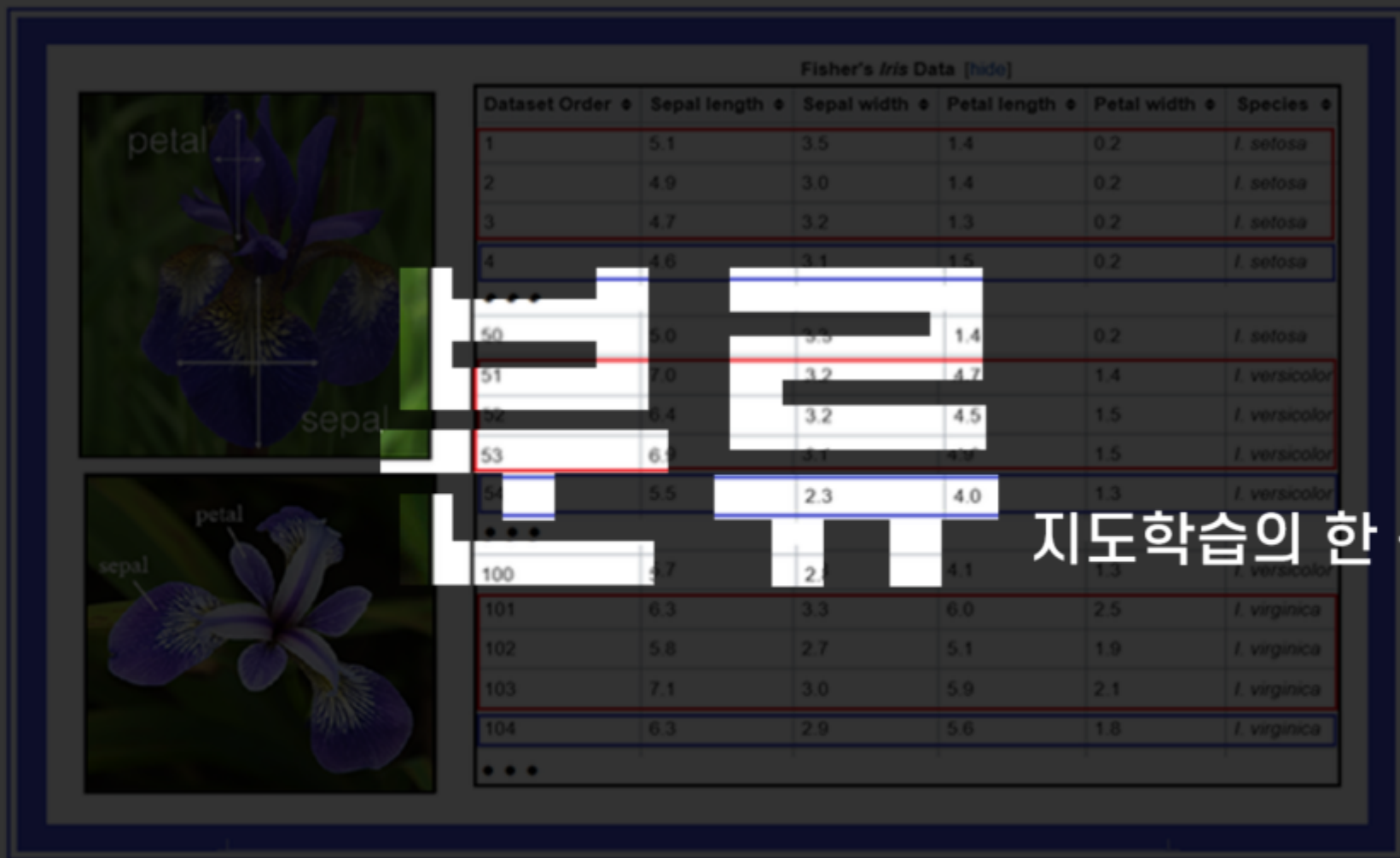


Fisher's Iris Data [hide]

Dataset Order	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	<i>I. setosa</i>
2	4.9	3.0	1.4	0.2	<i>I. setosa</i>
3	4.7	3.2	1.3	0.2	<i>I. setosa</i>
4	4.6	3.1	1.5	0.2	<i>I. setosa</i>
...					
50	5.0	3.3	1.4	0.2	<i>I. setosa</i>
51	7.0	3.2	4.7	1.4	<i>I. versicolor</i>
52	6.4	3.2	4.5	1.5	<i>I. versicolor</i>
53	6.9	3.1	4.9	1.5	<i>I. versicolor</i>
54	5.5	2.3	4.0	1.3	<i>I. versicolor</i>
...					
100	5.7	2.8	4.1	1.3	<i>I. versicolor</i>
101	6.3	3.3	6.0	2.5	<i>I. virginica</i>
102	5.8	2.7	5.1	1.9	<i>I. virginica</i>
103	7.1	3.0	5.9	2.1	<i>I. virginica</i>
104	6.3	2.9	5.6	1.8	<i>I. virginica</i>
...					

사이킷런을 사용하여 붓꽃의 종류를 예측해보았습니다. (스페이스바) 붓꽃 데이터 세트는 iris 데이터 세트를 이용해 불러올 수 있습니다. (스페이스바) 이 데이터 세트는 꽃잎의 길이와 너비, 그리고 꽃받침의 길이와 너비로 이루어져 있습니다. (스페이스바) (스페이스바)

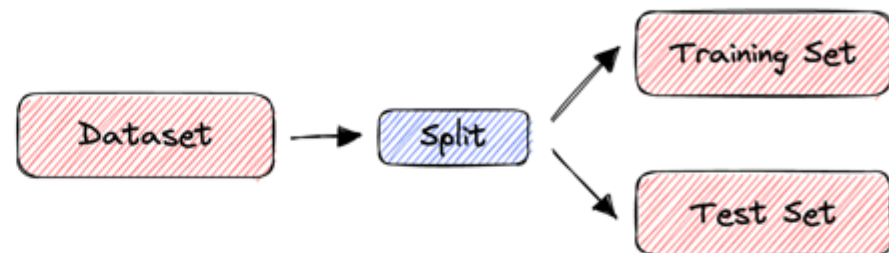
02. 붓꽃 품종 예측하기



Dataset Order	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	<i>I. setosa</i>
2	4.9	3.0	1.4	0.2	<i>I. setosa</i>
3	4.7	3.2	1.3	0.2	<i>I. setosa</i>
4	4.6	3.1	1.5	0.2	<i>I. setosa</i>
...
50	5.0	3.5	1.4	0.2	<i>I. setosa</i>
51	7.0	3.2	4.7	1.4	<i>I. versicolor</i>
52	6.4	3.2	4.5	1.5	<i>I. versicolor</i>
53	6.9	3.1	4.9	1.5	<i>I. versicolor</i>
54	5.5	2.3	4.0	1.3	<i>I. versicolor</i>
...
100	4.7	2.1	4.1	1.3	<i>I. versicolor</i>
101	6.3	3.3	6.0	2.5	<i>I. virginica</i>
102	5.8	2.7	5.1	1.9	<i>I. virginica</i>
103	7.1	3.0	5.9	2.1	<i>I. virginica</i>
104	6.3	2.9	5.6	1.8	<i>I. virginica</i>
...

지도학습의 한 종류

02. 붓꽃 품종 예측하기



(스페이스바)이 데이터를 활용하여 붓꽃의 종류를 분류하고자 합니다. 이는 대표적인 지도 학습 방법 중 하나인데, 미지의 정답을 예측하기 위해 먼저 학습 데이터를 이용하여 모델을 학습시키는 방식을 일컫습니다. (스페이스바)

02. 붓꽃 품종 예측하기

train_test_split(
arrays, test_size=None, train_size=None,
random_state = None, shuffle=True,
stratify=None)

Arrays

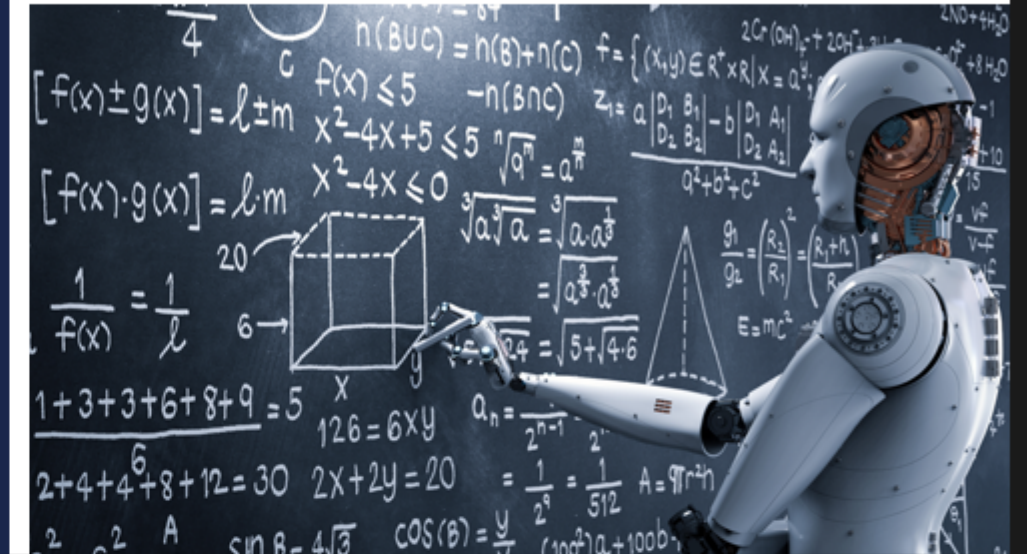
Test_size

Random_state

```
X_train, X_test, y_train, y_test = train_test_split(iris_data, iris_label,  
                                                    test_size=0.2, random_state=11)
```

다음 슬라이드

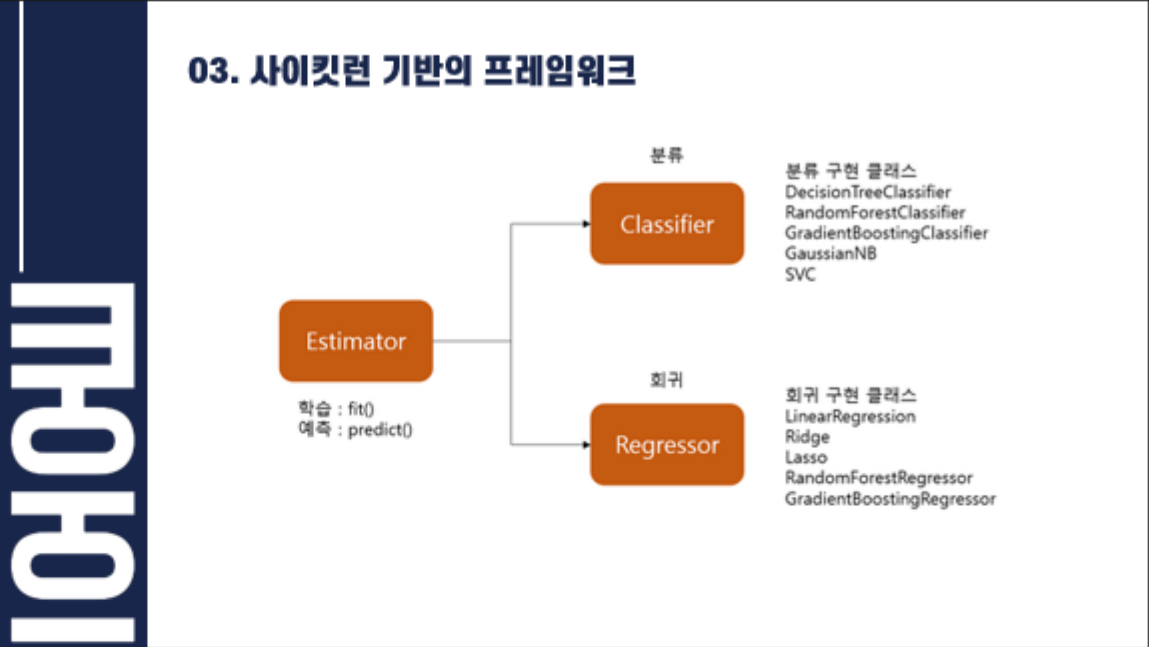
03. 사이킷런 기반의 프레임워크



모델을 학습시키기 위해서는 데이터 세트를 학습 데이터 세트와 테스트 데이터 세트로 나누어야 합니다. (스페이스바) 이는 학습된 모델의 성능을 평가하기 위해 테스트 데이터 세트가 필요하기 때문입니다. (스페이스바) 이를 위해 사용되는 함수가 train_test_split() 함수입니다. (스페이스바) 이 함수는 입력으로 피쳐 데이터 세트와 레이블 데이터 세트로 이루어진 array를 받으며, 전체 데이터 세트 중 테스트 데이터 세트의 비율을 나타내는 test_size와/ 같은 데이터 세트를 재현하기 위한 난수 발생값인 random_state를 인자로 받습니다. (스페이스바)

03. 사이킷런 기반의 프레임워크

Predict() 메서드



이처럼 데이터 세트를 나눈 뒤에는 학습 데이터 세트를 기반으로 모델을 학습시켜야 합니다. (스페이스바)
사이킷런은 머신러닝 모델 학습을 위해서 (스페이스바) `fit()` 메서드를, (스페이스바) 학습된 모델의 예측을 위해 (스페이스바) `predict()` 메서드를 제공합니다. (스페이스바)

03. 사이킷런 기반의 프레임워크

비고	모듈명	설명
데이터셋	sklearn.datasets	사이킷런에서 제공하는 데이터셋
데이터타입	sklearn.utils.Bunch	사이킷런에서 제공하는 데이터셋의 데이터 타입 (자료형)
데이터 전처리	sklearn.preprocessing	데이터 전처리(정규화, 인코딩, 스케일리등)
데이터 분리	sklearn.model_selection.train_test_split	학습용/테스트용 데이터셋 분리
평가	sklearn.metrics	분류, 회귀, 클러스터링 알고리즘의 성능을 측정하는 함수를 제공
머신러닝 알고리즘 (모델)	sklearn.ensemble	앙상블관련 머신러닝 알고리즘 - 랜덤 포레스트, 에이다 부스팅, 그래디언트 부스팅
	sklearn.linear_model	선형 머신러닝 알고리즘 - 릿지, 라쏘, SGD 등
	sklearn.naive_bayes	나이브 베이즈 관련 머신러닝 알고리즘
	sklearn.neighbors	최근접 이웃 모델 관련 - 릿지, 라쏘, SGD 등
	sklearn.svm	SVM관련 머신러닝 알고리즘
	sklearn.tree	트리 관련 머신러닝 알고리즘 - 의사결정 트리 등
	sklearn.cluster	군집관련 머신러닝 알고리즘

04. Model Selection

사이킷런에서는 분류와 회귀, 이 두 가지의 주요한 지도 학습 방법에 대한 다양한 알고리즘을 구현한 모든 클래스가, 단순히 `fit()`과 `predict()` 메서드만을 이용하여 간편하게 학습과 예측 결과를 반환할 수 있습니다. 분류 알고리즘을 구현한 클래스는 `Classifier`로, 회귀 알고리즘을 구현한 클래스는 `Regressor`로 지칭됩니다. 이 두 가지 클래스를 합쳐서 `Estimator` 클래스라고 부르며, 사이킷런은 매우 다양한 유형의 `Classifier`와 `Regressor` 클래스를 제공합니다. (스페이스바) 이들 `Classifier`와 `Regressor`를 포함한 사이킷런의 주요 모듈은 다음과 같습니다. (스페이스바)