

기계학습개론 중간 점검

01-1 인공지능과 머신러닝, 딥러닝

다음 중 규칙을 일일이 프로그래밍하지 않아도 자동으로 데이터에서 규칙을 학습하는 알고리즘 연구 분야를 무엇이라 하는가?

- 1. 로봇분야**
- 2. 머신러닝**
- 3. 컴퓨터비전**
- 4. 딥러닝**

01-2 코랩과 주피터 노트북

구글에서 제공하는 웹 브라우저 기반의 파이썬 실행 환경은 무엇인가?

1. 주피터 노트북
2. 코렙
3. 크롬
4. 아나콘다

코랩은 어디에서 실행되는가?

1. 내 컴퓨터
2. 구글 드라이브
3. 구글 클라우드
4. 아마존 웹서비스

01-3 마켓과 머신러닝

머신러닝에서 데이터를 표현하는 하나의 성질을 무엇이라 하는가? (예> 국가 데이터의 경우 인구 수, GDP, 면적 등이 하나의 국가를 나타냄)

1. 특성

2. 특징

3. 개성

4. 요소

가장 가까운 이웃을 참고하여 정답을 예측하는 알고리즘이 구현된 사이킷런 클래스는 무엇인가?

- 1. SGDClassifier**
- 2. LinearRegression**
- 3. RandomForestClassifier**
- 4. KNeighborsClassifier**

사이킷런 모델을 훈련할 때 사용하는 메서드는 무엇인가?

1. `predict()`
2. `fit()`
3. `score()`
4. `transform()`

02-1 훈련 세트와 테스트 세트

**머신러닝 알고리즘의 한 종류로써 샘플의 입력과 타킷 (정답)을 알고 있을 때 사용할 수 있는 학습법을 무엇이
라 하는가?**

- 1. 지도학습**
- 2. 비지도학습**
- 3. 차원축소**
- 4. 강화학습**

훈련 세트와 테스트 세트가 잘못 만들어져 전체 데이터를 대표하지 못하는 현상을 무엇이라 하는가?

1. 샘플링 오류
2. 샘플링 실수
3. 샘플링 편차
4. 샘플링 편향

사이킷런 활용을 위해 입력 데이터의 행, 열을 어떻게 구성해야 하는지 고르시오.

1. 행:특성, 열:샘플
2. 행:샘플, 열:특성
3. 행:특성, 열:타깃
4. 행:타깃, 열:특성

02-2 데이터 전처리

스케일 조정 방식의 하나로 특성값이 0에서 표준편차의 몇 배수만큼 떨어져 있는지를 나타내는 값을 무엇이라 하는가?

- 1. 기본점수**
- 2. 원점수**
- 3. 표준점수**
- 4. 사분위수**

테스트 세트의 스케일을 조정하려고 한다. 다음 중 어떤 데이터 통계값을 사용하는 것이 적절한가?

- 1. 훈련세트**
- 2. 테스트 세트**
- 3. 전체 데이터**
- 4. 변경 안 함**

03-1 k -최근접 이웃 회귀 알고리즘

k-최근접 이웃 회귀에서는 새로운 샘플에 대한 예측을 어떻게 만드는가?

- 1. 이웃 샘플 클래스 중 다수인 클래스**
- 2. 이웃 샘플의 타깃값과 평균**
- 3. 이웃 샘플 중 가장 높은 타깃값**
- 4. 이웃 샘플 중 가장 낮은 타깃값**

03-2 선형 회귀 알고리즘

선형 회귀 모델이 찾은 방정식의 계수를 무엇이라 하는가?

1. 회귀 파라미터
2. 선형 파라미터
3. 학습 파라미터
4. 모델 파라미터

사이킷런에서 다항 회귀 모델을 훈련시킬 수 있는 클래스는 무엇인가?

1. **LinearRegression**
2. **PolynomialRegression**
3. **KNeighborsClassifier**
4. **PolynomialClassifier**

03-3 특성 공학과 규제

특성 a , b , c 로 이루어진 훈련 세트를
`PolynomialFeatures(degree=3)`으로 변환하였다고 가정하
자. 이 때 변환된 데이터에 포함되지 않는 특성은 무엇인가?

1. 1

2. a

3. $a * b$

4. $a * b^3$

다음 중 특성을 표준화하는 사이킷런 변환기 클래스는 무엇인가?

1. Ridge
2. Lasso
3. StandardScaler
4. LinearRegression

다음 중 평가 결과가 과대적합인 모델을 고르시오.

- 1. 훈련세트 : 0.923 테스트 세트 : 0.851**
- 2. 훈련세트 : 0.851 테스트 세트 : 0.923**
- 3. 훈련세트 : 0.923 테스트 세트 : 0.923**
- 4. 훈련세트 : 0.754 테스트 세트 : 0.743**

04-1 로지스틱 회귀

분류해야 할 클래스 수가 3개 이상인 분류 문제를 무엇이라 하는가?

1. 이진 분류
2. 다중 분류
3. 단변량 회귀
4. 다변량 회귀

로지스틱 회귀 알고리즘을 이용하여 다중 분류를 할 때, 결과를 확률로 변환하기 위해 사용하는 함수는 무엇인가?

- 1. 시그모이드 함수**
- 2. 소프트맥스 함수**
- 3. 로그 함수**
- 4. 지수 함수**

04-2 확률적 경사 하강법

**다음 중 표준화 같은 전처리를 처리하지 않아도 되는
방식으로 구현된 클래스는 무엇인가?**

- 1. KNeighborsClassfier**
- 2. LinearRegression**
- 3. Ridge**
- 4. SGDClassifier**

경사 하강법 알고리즘의 하나로써 훈련 세트에서 몇 개의 샘플만을 뽑아서 훈련하는 방식은 무엇인가?

1. 확률적 경사 하강법
2. 배치 경사 하강법
3. 미니배치 경사 하강법
4. 부분배치 경사 하강법