|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1문제 | 2문제 | 3문제 | 4문제 | 5문제 | 6문제 | 7문제 | 8문제 | 9문제 | 10문제 | 11문제 | 12문제 |
| 3 | 2 | 3 | 주관식 | 2 | 1 | 1 | 주관식 | 2 | 1 | 2 | 주관식 |
| 13문제 | 14문제 | 15문제 | 16문제 | 17문제 | 18문제 | 19문제 | 20문제 | 21문제 | 22문제 | 23문제 | 24문제 |
| 4 | 4 | 3 | 주관식 | 3 | 4 | 4 | 주관식 | 2 | 4 | 1 | 주관식 |

1. Pandas의 시리즈에 대한 설명 중 틀린 것은?

난이도: 하

정답: 3

해설: columns는 데이터 프레임의 속성입니다. 시리즈에는 없는 속성입니다.

1. 데이터 프레임의 컬럼 하나를 시리즈 객체로 가져올 수 있다.

2. 시리즈에는 index라는 속성이 있다.

3. 시리즈에는 columns라는 속성이 있다.

4. 여러 통계 함수가 메서드로 제공된다.

2. 다음 Numpy 배열에 대한 설명 중 틀린 것은?

난이도: 하

정답: 2

해설: shape라는 속성은 배열의 차원을 tuple로 보여줍니다.

1. size라는 속성은 성분 (원소)의 개수를 나타낸다.

2. shape라는 속성은 배열의 차원을 하나의 수치로 나타낸다.

3. reshape()라는 메서드로 shape를 바꾸어 줄 수 있다.

4. '\*' 연산자는 두 배열의 원소를 서로 곱해서 새로운 배열을 만들어 준다.

3. Pandas의 시리즈에 대한 설명 중 틀린 것은?

난이도: 중

정답: 3

해설: 시리즈의 자료형은 통일되어 있어야 합니다.

1. 시리즈 객체의 속성 values는 시리즈의 값을 numpy 배열로 변환시켜 준다.

2. 시리즈에는 name이라는 속성이 있다.

3. 시리즈의 자료형은 혼재해 있어도 된다.

4. pandas.Series()로 새롭게 시리즈 객체를 생성할 수 있다.

4. 주머니 안에 흰색 공 4개 빨간색 공 6개가 들어있다. 공을 한 개씩 두 번 꺼낼 때, 둘 다 모두 빨간 공일 확률은? 단 처음 꺼낸 공은 주머니에 다시 넣지 않는다. <단답식>

난이도: 중

정답: 1/3

해설: 첫번째 공이 빨간색일 확률 x 첫번째 공이 빨간색인 것을 전제하고 두번째 공도 빨간색일 확률 = 6/10 x 5/9 = 30/90 = 1/3.

5. 다음 보기는 어떠한 개념에 대한 답을 얻기 위한 질문이며,

질문자는 어떠한 것을 파악하기 위하여 질문을 하였는지 선택하시오.

보기

질문자: “ 양성으로 식별된 사례 중 실제로 양성이었던 사례의 비율은 어느 정도인가요? ”

난이도: 하

정답: 2

해설: 수업자료 참고

1. 재현율

2. 정밀도

3. F1 score

4. 정확도

6. 다음 보기와 같이 4가지 지표에 대한 값이 주어진 경우,

올바르게 구해진 재현율(recall)을 선택하시오.

난이도: 하

정답: 1

해설: 재현율 도출 공식 참고

보기.

TP : 1 / FP : 1 / FN : 8 / TN : 90

1. 0.11

2. 0.21

3. 0.31

4. 0.41

7. 단층 퍼셉트론을 구축하기 위해 정의한 함수 중에서, '직접적인 학습'이 이뤄지는 함수를 고르시오.

1. backprop\_neuralnet()

2. forward\_neuralnet()

3. backprop\_postproc()

4. forward\_postproc()

난이도: 중

정답: 1

해설: 수업자료 참고

8. 신경망의 역전파 수행과정에 대한 설명 중, 빈칸에 알맞은 함수를 채워 넣으시오.

난이도: 상

정답: np.ones, np.prod

해설: 수업자료 참고

회귀분석의 순전파 과정에 대한 역전파를 수행하기 위하여 ∂L/∂Y 를 구하기 위한 과정인 '평균제곱오차'에 대한 역전파 처리 과정 중,

'평균'과 '제곱'의 과정을 하나의 과정으로 가정 했을 때, 이에 대한 '제곱'에 대한 '손실 기울기'를 구하는 단계를 파이썬으로 구현하고자 한다. 이때 사용할 수 있는 넘파이 함수는 \_\_\_\_\_\_ 과 \_\_\_\_\_\_ 가 있다.

9. 정규표현식에 대한 설명 중 틀린 것은?

난이도: 하

정답: 2

해설: [\W] 란 문자나 숫자에 해당하지 않는 것을 표현하는 정규표현식입니다.

1. [^\w] 란 문자나 숫자가 아닌 것을 표현하는 정규표현식이다.

2. [\W] 란 문자나 숫자에 해당하는 것을 표현하는 정규표현식이다.

3. \1 은 만든 그룹을 재참조할 때에 쓰인다.

4. \s 는 띄어쓰기에 해당하는 것을 표현하는 정규표현식이다.

10. 문장의 전처리 과정에 대한 설명 중 틀린 것은?

난이도: 하

정답: 1

해설: 토큰화의 기준은 사용자가 형태소, 단어, 문장 등의 기준 중에서 정하는 것입니다.

1. 토큰화란 단어를 기준으로 문장을 나누는 것을 말한다.

2. nltk 토크나이저는 split 함수에 비해 수행속도가 느리다.

3. 전처리 과정에서 사용할 토크나이저는 데이터의 특성 및 분석 목적에 따라 다르게하여야 한다.

4. nltk stopwords를 사용하지 않아도 stopwords를 제거할 수 있다.

11. TF-IDF에 대한 설명 중 틀린 것은?

난이도: 하

정답: 2

해설: 특정 단어 t의 IDF는 전체 문서가 n일때, log(n/(1+df(t)))로 표현할 수 있습니다.

1. DF란 특정 단어가 포함된 문서들의 빈도수를 뜻한다.

2. 특정 단어 t의 IDF는 전체 문서가 n일때, exp(n/(1+df(t)))로 표현할 수 있다.

3. TF-IDF란 특정 문서, 특정 단어의 TF와 IDF를 곱한 것을 뜻한다.

4. TF-IDF 행렬은 DTM 행렬과 똑같은 차원을 가진다.

12. [‘나는 밥을 먹었다’, ‘어제의 밥은 정말 맛이 없었다’, ‘어떻게 하면 맛있는 밥을 먹을 수 있을까’]

위 세 문장을 split 함수를 사용하여 토큰화를 한다고 가정할 때, ‘밥을’의 idf 값을 구하시오. <단답식>

난이도: 중

정답: 0

해설: log(3/2+1) = 0

13. 다음 중 머신러닝과 딥러닝의 구분 중 틀린 것은?

난이도: 하

정답: 4

해설: 딥러닝은 사람의 개입이 적고 알고리즘이 수행하기에 해석에 어려움이 있습니다.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 구분 | 머신러닝 | 딥러닝 |
| 1. | 훈련 데이터 크기 | 작음 | 큼 |
| 2. | 시스템 성능 | 저 사양 | 고 사양 |
| 3. | feature 선택 | 전문가 (사람) | 알고리즘 |
| 4. | 해석력 | 해석 어려움 | 해석 가능 |

14. 훈련-테스트 세트에 대한 설명 중 틀린 것은?

난이도: 중

정답: 4

해설: 훈련 세트에서 잘 동작하지만 테스트 세트에서 성능이 나쁘다면 테스트 데이터에 더 가깝게 되도록 훈련 데이터를 개선해야 합니다.

1. 모델을 실전에 투입했을 때 사용될 데이터와 가능한 최대로 가까워야 합니다.

2. 훈련 세트의 일부에서 모델을 훈련하고 테스트 세트에서 평가합니다.

3. 모델이 훈련 세트에서 잘 동작하지만 테스트 세트에서 나쁜 성능을 낸다면 아마도 훈련 세트에 과대적합되었을 가능성이 높습니다.

4. 훈련 세트에서 잘 동작하지만 테스트 세트에서 성능이 나쁘다면 훈련 데이터에 더 가깝게 되도록 테스트 데이터를 개선해야 합니다.

15. 다음 설명에 해당하는 모델은?

난이도: 중

정답: 3

해설: 라쏘회귀는 L1규제를 사용하여 특정 계수들이 0이 되며, 특성선택이 자동으로 이루어 집니다.

Regularization을 사용하여 가중치를 완전히 0으로 만드는 경향이 있습니다. 이는 가장 중요한 가중치를 제외하고는 모두 0이 되는 희소한 모델을 만듭니다. 또한 자동으로 특성 선택의 효과를 가지므로 단지 몇 개의 특성만 실제 유용할 것이라고 의심될 때 사용하면 좋습니다.

1. 선형회귀

2. 릿지회귀

3. 라쏘회귀

4. 엘라스틱넷

16. 다음 밑줄에 들어갈 알맞은 시각화 방법을 쓰시오.

난이도: 하

정답: 산점도 행렬

해설: 산점도 행렬은 3개 이상의 특성을 비교 분석할 때 좋은 시각화 방법입니다.

시각화는 데이터를 조사하는 아주 좋은 방법입니다. 산점도는 데이터에서 한 특성을 x 축에 놓고 다른 하나는 y 축에 놓아 각 데이터 포인트를 하나의 점으로 나타내는 그래프입니다. 컴퓨터 화면은 2차원이라 한 번에 2개의 특성만 그릴 수 있습니다. 3개 이상의 특성을 표현하기 위해 모든 특성을 짝지어 만드는 \_\_\_\_\_\_\_\_\_\_\_을 사용할 수 있습니다.

17. git init 을 한 후 작업 한 파일 hello.py 를 staging area 에 올리는 방법으로 옳은 것을 고르시오.

난이도: 하

정답: 3

해설: staging area 에 올리기 위해서는 git add 파일명을 해야합니다.

1. git commit hello.py

2. git status

3. git add hello.py

4. git log

18. 현재 철수의 컴퓨터에는 데스크탑 디렉토리 안에 인공지능\_사관학교, 컴퓨터\_사이언스, 코테\_준비 디렉토리가 있다. 인공지능\_사관학교 디렉토리의 하위 디렉토리로 깃\_공부 디렉토리가 있다. 철수는 깃\_공부 디렉토리 안에서 git\_tutorial.txt 파일을 작성하며 공부하였다. 철수가 공부한 것을 깃헙에 업로드 하고 싶다고 할 때 git init을 할 위치(ㄱ)와 git remote add origin <https://github.com/chulsu/git_study> 를 할 위치(ㄴ)로 모두 옳은 것을 고르시오.

난이도: 하

정답: 4

해설: git init과 git remote add origin은 올리고 싶은 디렉토리 ‘안’에 들어가서 해야한다.

1. ㄱ. 터미널에서 ls 명령어를 입력했을 때 인공지능\_사관학교가 보이는 위치

ㄴ. 터미널에서 ls 명령어를 입력했을 때 코테\_준비가 보이는 위치

1. ㄱ. 터미널에서 ls 명령어를 입력했을 때 깃\_공부 디렉토리가 보이 는 위치

ㄴ. 터미널에서 ls 명령어를 입력했을 때 깃\_공부 디렉토리가 보이는 위치

3. ㄱ. 터미널에서 ls 명령어를 입력했을 때 git\_tutorial.txt가 보이는 위치

ㄴ. 터미널에서 ls 명령어를 입력했을 때 깃\_공부 디렉토리가 보이는 위치

4. ㄱ. 터미널에서 ls 명령어를 입력했을 때 git\_tutorial.txt 가 보이는 위치

ㄴ. 터미널에서 ls 명령어를 입력했을 때 git\_tutorial.txt 가 보이는 위치

19. 다음 중 position : absolute ; 의 설명으로 옳은 것을 고르시오. (부모는 하나로 가정)

난이도: 하

정답: 4

해설: position : absolute ; 는 부모의 포지션을 기준으로 상황이 나뉘는데, 부모가 position : static; 이라면 body 태그를 기준으로 이동합니다.

1. static 일 때의 위치를 기준으로 이동합니다.

2. 부모가 position : relative; 라면, static 일 때의 위치를 기준으로 이동합니다.

3. body 태그를 기준으로 이동합니다.

4. 부모가 position : static 이라면, body 태그를 기준으로 이동 합니다.

20. HTML 의 기본 margin 을 0으로 만들고 싶을때 작성해야 하는 CSS 코드를 작성하시오. <단답식>

난이도: 하

정답: \* {

margin : 0 ;

}

또는,

body {

margin : 0 ;

}

해설: \* 선택자 또는 태그 선택자로 전체 margin을 수정할 수 있습니다.

21. 다음 코드의 시간복잡도를 빅오표기법으로 올바르게 표기한 것을 고르시오.

val = int(input())

result = 1

for i in range(val, 0, -1):

result \*= i

print(result)

1. O(1)
2. O(N)
3. O(N2)
4. O(N log N)

난이도: 하

정답: 2

22. MySQL은 세계에서 가장 많이 쓰이는 오픈 소스 RDBMS(Relational Database Managment System) 이다. B-Tree 인덱스 알고리즘은 MySQL을 비롯해 데이터베이스의 인덱스(Index) 알고리즘 중 보편적인 알고리즘 중 하나이다. 해당 알고리즘의 시간복잡도를 빅오표기법으로 올바르게 표기한 것을 고르시오.

1. O(1)
2. O(N)
3. O(N2)
4. O(N log N)

난이도: 상

정답: 4

해설: B-Tree의 시간복잡도는 O(N log N) 이다.

23. 이 정렬은 느린 정렬 알고리즘으로 기본적인 정렬 알고리즘 중 하나이다. 배열의 모든 요소를 이미 정렬된 배열 부분, 즉 자신의 바로 앞 인덱스까지의 구간과 비교하는 방식이며 각각의 요소를 정렬된 배열 사이에 올바른 위치를 찾아 삽입하는 정렬 방식이다. 파이썬에서는 이 정렬방식과 병합 정렬을 Hybrid 방식으로 사용한다고도 하며, 최선의 시간복잡도는 O(N), 최악의 시간 복잡도는 O(N2) 을 가지는 이 정렬 알고리즘의 이름은 무엇인가?

1. 삽입 정렬(Insertion Sort)
2. 선택 정렬(Selection Sort)
3. 거품 정렬(Bubble Sort)
4. 퀵 정렬(Quick Sort)

난이도: 중

정답: 1

해설: 삽입 정렬에 대한 개념을 설명하고 있다.

24. 스택이라는 자료구조는 후입선출(First In Last Out, Last In First Out) 구조를, 큐라는 자료구조는 선입선출 (First In First Out) 구조를 가진다. 비어있는 스택 자료구조가 있을 경우 아래의 작업을 모두 마친 후, 출력되는 수들의 합은 얼마인가? (스택에 존재하지 않는 출력값을 요청하는 경우 -1이 출력된다고 간주한다.)

1을 PUSH => 2를 PUSH => POP 실행 => 3을 PUSH

=> EMPTY 실행 => 4를 PUSH => POP 실행 => TOP 실행

난이도: 중

정답: 9

해설: 첫번째 출력 값인 POP 실행은 2, EMPTY 실행은 0, 다음 POP 실행은 4, 마지막 TOP 실행은 3의 값을 가진다.