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

**1. 다음 백분위수에 대한 설명 중 틀린 것은?**

난이도: 하

**정답: 1**

해설: 표본의100% 분위수는 최고값 이고 0% 분위수는 최소값 입니다. 이들 사이의 거리는 1이 아닐 수도 있습니다.

1. 표본의100% 분위수와 0% 분위수 사이의 거리는 항상 1이다.

2. 박스플롯 (상자그림)을 그리면 데이터의 25%, 50%, 75% 분위수를 쉽게 알 수 있다.

3. 50% 분위수 (중위수)는 이상치에 둔감하다.

4. 50% 분위수 (중위수)는 평균과 일치할 수도 있고 일치하지 않을 수도 있다

**2. 다음 확률분포 함수 중 그 성격이 나머지 것들과 다른 것은?**

난이도: 하

**정답: 2**

해설: 카이제곱은 연속형인데 나머지 것들은 이산형 입니다.

1. 주사위.

2. 카이제곱.

3. 이항분포.

4. 동전 (베르누이).

**3. 백열전구의 수명이 평균3000시간이며 표준편차는 1000시간인 정규분포를 따른다고 알려져 있다. 백열전구 중 가장 수명이 긴 상위 5%는 최소 몇시간 이상 사용할 수 있는가를 계산해 주는 Python수식은?**

**"import scipy.stats as st" 를 전제한다.**

난이도: 중

**정답: 1**

해설: 백열전구 수명의 상위 5%위치는 95% 분위수로 계산할 수 있습니다. 분위수 함수는 ppf() 입니다.

1. st.norm.ppf(0.95, 3000, 1000)

2. st.norm.pdf(0.05, 3000, 1000)

3. st.norm.cdf(5, 3000, 1000)

4. st.norm.ppf(0.05, 3000, 1000)

**4. X와 Y라는 두개의 확률변수가 있다. 그런데 이들 사이의 공분산은 -15이다. 또한 X의 분산은 25이고 Y의 분산은 36이다. X와 Y사이의 상관계수는? (단답식)**

난이도: 중

**정답: -0.5**

해설: 상관계수 = 공분산/(표준편차 X \* 표준편차 Y)

= -15/(5 \* 6) = -0.5

**5. 주어진 '입력 픽셀 행렬'과 '합성곱 커널'을 활용하여**

**'출력 픽셀 행렬'을 도출하고,**

**'출력 픽셀 행렬'의 '왼쪽 상단 값' 의 도출 과정을 수식으로 작성하시오.**

**(※편향은 존재하지 않습니다.)**

**(※'출력 픽셀 행렬'의 작성 순서는 최상단에서 시작, 왼쪽에서 오른쪽으로 이어지는 순서로 작성하시오.**

**ex) '합성곱 필터'를 예를 들면 1,0,0,1,1,0,0,0,1 )**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | 5 | 2 | 8 | 1 |
| 9 | 7 | 5 | 4 | 3 |
| 2 | 0 | 6 | 1 | 6 |
| 6 | 3 | 7 | 9 | 2 |
| 1 | 4 | 9 | 5 | 1 |

※입력 픽셀 행렬 ※합성곱 필터

|  |  |  |
| --- | --- | --- |
| 1 | 0 | 0 |
| 1 | 1 | 0 |
| 0 | 0 | 1 |

난이도: 하

정답: 25,18,17,18,22,14,20,15,23 / (3\*1)+(5\*0)+(2\*0)+(9\*1)+(7\*1)+(5\*0)+(2\*0)+(0\*0)+(6\*1)

해설: 수업 자료 참고

5번 (이재화강사님)  
1. 답은 맞는데 수식은 안적은 경우 => 1점  
2. 왼쪽 상단값만 수식과 답을 적은 경우 => 1점  
3. 첫번째 행렬의 수식은 맞았으나 9개 답 중 하나만 틀린 경우 => 0점

**6. CNN은 모델에 비선형성을 주기 위해 컨볼루션 작업이 끝날 때마다 합성곱된 특성에 꺾인 선형 유닛(ReLU) 변형을 적용합니다.**

**로 정의되는 ReLU 함수는 X > 0인 모든 값에 대해 A 를,**

**x  ≤ 0인 모든 값에 대해 B를 반환합니다.**

**여기서 A와 B의 값으로 알맞게 구성된 보기를 선택하시오.**

1. **A :** X, **B :**  0

2. **A :** 0, **B :** X

3, **A :** 1, **B :** 0

4. **A :** 0, **B :** 1

난이도: 중

정답: 1

해설: ReLU 함수는  x   > 0인 모든 값에 대해   x 를,   x   ≤ 0인 모든 값에 대해 0을 반환합니다.

**7. 빈칸에 알맞은 단어로 짝지어진 것을 고르시오**

어떠한 신경망 모델에서나 그렇듯이 컨볼루션 신경망을 학습시킬 때 가장 중요한 관심사는 **'A' 가(이)** 발생하지 않도록 하는 것입니다.

**'A'** 란 모델이 학습 데이터의 특성에 지나치게 맞춰져 새로운 예를 접했을 때 일반화할 수 없는 상태를 말합니다.

CNN을 구축할 때 다음과 같은 두 가지 기법을 사용하여 **'A'** 가(이) 방지할 수 있습니다.

**B** : 기존의 이미지를 임의로 변형함으로써 학습용 예의 다양성과 개수를 인위적으로 늘려 새로운 변형 세트를 만듭니다. **B** 는 데이터 세트가 비교적 작을 때 특히 유용합니다.

**C** : 학습 단계에서 신경망으로부터 유닛을 임의로 삭제합니다.

1. A : 과적합 B : 이미지 보강 C : 드롭아웃

2. A : 지역 최솟값 B : 데이터 변형 C : 레이어 삭제

3. A : 과적합 B : 데이터 보강 C : 드롭아웃

4. A : 기울기 소실 B : 이미지 변형 C : 레이어 삭제

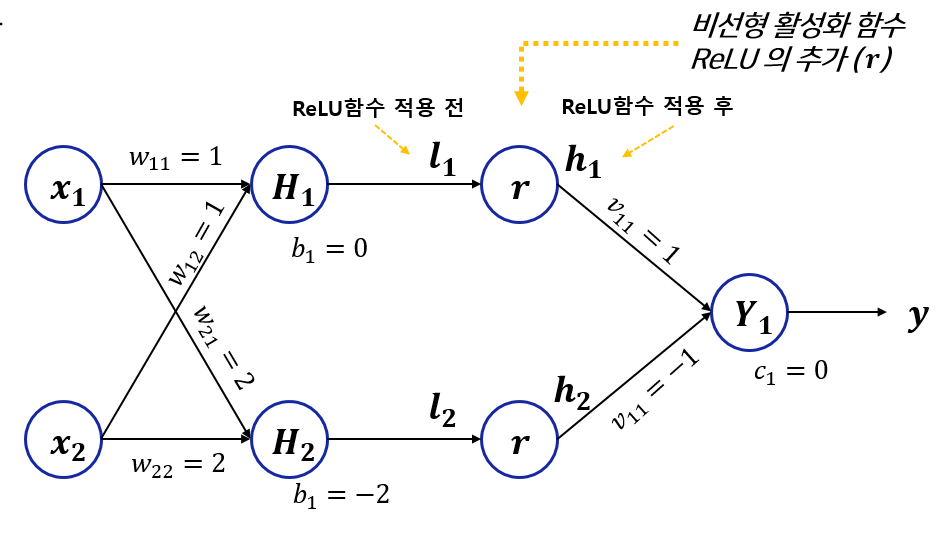
난이도: 중

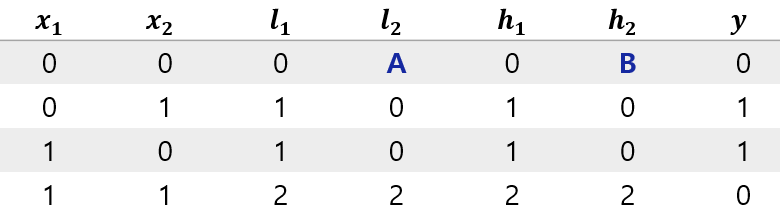
정답: 1,3 둘 중 하나 체크 정답

해설: 수업 참고자료 참고

**8. 본 그림은 민스키의 XOR 문제를 활성화 함수와 다층 퍼셉트론**

**으로 풀어내는 과정이다. 해당 자료를 바탕으로 A와 B에 들어갈**

**알맞은 값들로 짝지어진 것을 고르시오.** 



1. A : -2 B : 0

2. A : 2 B : 0

3. A : -1 B : 1

4. A : -1 B : 1

난이도: 중

정답: 1

해설 : 수업 참고자료 참고

**9. RNN에 대한 설명 중 틀린 것은?**

난이도: 하

정답: 1

해설: RNN이란 Recurrent Neural Network를 말합니다.

1. RNN이란 Recursive Neural Network를 말한다.

2. RNN은 단어 하나를 분석할 때보다 문장을 분석할 때 더 유용하다.

3. 기본 RNN은 데이터가 길어질 수록 앞 정보의 영향력이 점점 감소하게 된다.

4. RNN에는 다음 셀로 넘겨줄 hidden state가 존재한다.

**10. LSTM에 대한 설명 중 틀린 것은?**

난이도: 하

정답: 2

해설: LSTM의 장기 상태는 entrywise product를 통해서 구할 수 있습니다.

1. LSTM에는 장기 상태와 단기 상태가 존재한다.

2. LSTM의 장기 상태는 dot product를 통해서 구할 수 있다.

3. LSTM에는 삭제를 위한 게이트가 존재한다.

4. LSTM은 기본 RNN의 장기 의존성 문제를 대처하고 있다.

**11. CNN에 대한 설명 중 틀린 것은?**

난이도: 하

정답: 3

해설: 커널 수는 채널 수와 관계없습니다.

11번 전원정답

1. 자연어처리에는 1D CNN이 주로 사용된다.

2. Global Pooling에서는 strides, pool size를 지정하지 않아도 된다.

3. 커널 수는 채널 수와 항상 같다.

4. 커널 사이즈란 커널과 연산을 진행할 때 함께 들어갈 단어의 개수를 말한다.

**12. 다중분류를 할 때에 일반적으로 사용하는 출력 레이어 활성함수의 이름과 모델 손실함수의 이름을 작성하시오. <단답식>**

난이도: 하

정답: **softmax, categorical\_crossentropy**

해설: 일반적으로 다중분류에서는 softmax와 categorical\_crossentropy를 사용합니다.

12번 (김준태강사님)  
1. relu softmax를 순서대로 적은 경우  
2. softmax만 맞은 경우  
3. softmax, (크로스)엔트로피라고만 쓴 경우  
=> categorical이 없거나 softmax만 맞는경우 1점

**13. 다음 로지스틱 회귀(Logistic Regression)에 대한 설명 중 틀린 것은?**

난이도: 하

정답: 4

해설: 경사하강법(Gradient Descent) 최적화 알고리즘을 사용하여 선형 모델을 작성하는 것은 선형 분류(Linear Classifier) 입니다.

정답으로 된 4번 내용이 수업에 사용된 ml\_05\_지도학습\_Linear\_Classifier.ipynb 파일의 선형분류에서 SGDClassifier() 함수를 설명할 때 언급했던 문장이기에 SGDClassifier를 떠올리지 않을까 생각했으나 문자그대로 해석하면 틀린 말은 아니기에 답이 없는 것으로 하겠습니다.

1. 로지스틱 회귀는 희소한 데이터 세트 분류에도 뛰어난 성능을 보여서 텍스트 분류에서도 자주 사용됩니다.

2. 직선 대신 S 모양 곡선을 사용한다는 사실 때문에 데이터를 그룹으로 나누는 데 적합 합니다.

3. 선형 회귀 방식을 기반으로 하되 시그모이드 함수를 이용해 분류를 수행하는 회귀입니다.

4. 경사하강법(Gradient Descent) 최적화 알고리즘을 사용하여 선형 모델을 작성합니다.

**14. 다음 사이킷런 LogisticRegression 클래스의 주요 하이퍼 파라미터 penalty와 C에 대한 설명 중 틀린 것은?**

난이도: 중

정답: 2

해설: penalty의 기본은 'l2' 입니다.

1. C는 규제 강도를 조정하는 alpha값의 역수입니다.

2. penalty는 규제(Regularization)의 유형을 설정하며 'l2'로 설정시 L2 규제를, 'l1'으로 설정 시 L1 규제를 뜻합니다. 기본은 'l1'입니다.

3. C 값이 작을 수록 규제 강도가 큽니다.

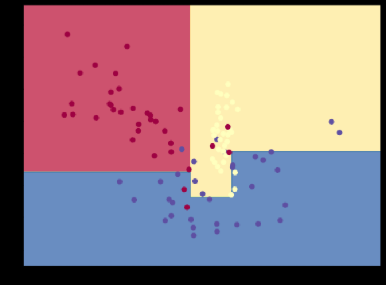
4. C가 크면 개개의 데이터 포인를 정확히 분류하려고 하고, C가 작으면 데이터 포인트 중에서 다수에 맞추려고 합니다.

**15. 결정 트리(Logistic Regression)에 결정기준 경계를 시각화한 결과이다. 바르게 설명된 것은?**

난이도: 중

정답: 4

해설: 1,2,3는 이상치(Outlier)를 학습한 과적합된 경우를 설명하고 있습니다.



1. 일부 이상치(Outlier) 데이터까지 분류하기 위해 분할이 자주 일어나서 결정 기준 경계가 매우 많아졌습니다.

2. 리프 노드 안에 데이터가 모두 균일하거나 하나만 존재해야 하는 엄격한 분할 기준으로 인해 결정 기준 경계가 많아지고 복잡해졌습니다.

3. 이렇게 복잡한 모델은 학습 데이터 세트의 특성과 약간만 다른 형태의 데이터 세트를 예측하면 예측 정확도가 떨어지게 됩니다.

4. 이상치에 크게 반응하지 않으면서 좀더 일반화된 분류 규칙에 따라 분류 되었습니다.

**16. 다음 밑줄에 들어갈 알맞은 결정 트리 파라미터를 쓰시오.**

난이도: 중

정답: min\_samples\_split

해설: 과적합을 제어기 위해 사용하며 1로 설정할 경우 오히려 분할되는 노드가 많아져서 과적합 가능성이 증가합니다.

\_\_\_\_\_\_\_\_\_\_ 파라미터는 노드를 분할하기 위한 최소한의 샘플 데이터 수로 과적합을 제어하는 데 사용됩니다. 디폴트는 2이고 작게 설정할수록 분할되는 노드가 많아져서 과적합 가능성이 증가합니다.

**17. box\_container라는 클래스를 가진 요소를 가운데 정렬하기 위한 코드를 flex를 활용하여 CSS 속성 부분만 작성하시오.**

난이도: 하

정답: .box\_container {

display : flex;

justify-content : center;

align-items : center;

}

17번 (최주원 강사님)  
1. 선택자를 쓰지 않은 경우 =>1점  
2. display, justify-content, align-items는 모두 맞히셨으나   
다른 속성도 쓴 경우 => flex관련속성이 아니면 정답

**18. flex-direction : column; 인 상황에서 가로축을 정렬하기 위한 방법으로 옳은 것을 고르시오.**

난이도: 하

정답: 2

해설: column 상태에서는 main-axis가 바뀌기 때문에, 가로축을 정렬하기 위해서 justify-content가 아닌, cross-axis 정렬 방법인 align-items 를 사용해야한다.

1. justify-content : center;

2. align-items : center;

3. flex-wrap : wrap;

4. justify-items: center;

**19. BeautifulSoup 에서 요소를 배열로 만들어서 리턴해주는 함수를 적으시오. (주관식)**

난이도: 하

정답: select 또는 find\_all

해설 : 배열로 만들어주는 함수는 select 와 find\_all 입니다.

**20. 해당 페이지에 요청을 보낼 때, Request Headers에 담아서 보내는 것으로, 이전 페이지의 정보를 담는 옵션으로 옳은 것을 고르시오.**

난이도: 하

정답: 2

1. cache-control

2. Referer

3. accept

4. user-agent

**21. 다음 코드는 파이썬 내장 함수를 통해 N진법 수 A를 10진수로 변환하는 코드이다. 빈 칸에 들어갈 알맞은 코드를 고르시오.**

N = int(input()) # 16

A = input() # ‘1BF’

result = **[ ??? ]**(A, N)

print(result) # 447

난이도: 하

정답: 2

1. divmod

2. int

3. decimal

4. n\_conv

**22. 다음 설명은 덱(deque)에 대한 설명이다. 이 자료구조는 양쪽 끝에서 삽입과 삭제가 가능하며 큐와 스택을 합친 형태로 간주할 수 있다. 왼쪽(front)에서 삽입을 발생시키기 위해서는 호출해야하는 함수와 시간복잡도가 알맞게 짝지어진 것을 고르시오.**

난이도: 하

정답: 2

해설: appendleft, append, pop, popleft의 시간복잡도는 모두 O(1)이며

덱의 왼쪽(front)에 요소를 삽입하는 함수는 appendleft,

오른쪽(back)에 삽입하는 함수는 append이다.

1. append, O(1)

2. appendleft, O(1)

3. append, O(N)

4. appendleft, O(N)

**23. 아래의 설명에 해당하는 알고리즘을 쓰시오. 지역적으로 최적이면서 전역적으로 최적인 문제를 풀 때 보편적으로 사용하며 이전 단계의 선택이 이후의 선택에 영향을 주지 않고, 문제에 대한 최적해가 sub-problem에 대해서도 최적해라는 조건을 가지는 경우에 해당 알고리즘을 사용하여 적절히 문제를 해결할 수 있다. 이 알고리즘은 무엇인가? (강의에서 배운 용어 안에서 답하시오)**

난이도: 중하

정답: 그리디, greedy, 욕심쟁이, 탐욕을 포함한 답변은 정답으로 인정한다. (그리디 알고리즘, greedy algorithm, 욕심쟁이 방법 등.)

**24. 아래 코드는 이진 탐색을 재귀로 구현한 코드의 일부이다. 코드의 빈 칸을 순서대로 알맞게 짝지어진 것을 고르시오.**

def b\_search(total\_list, find\_target\_value, start\_idx, end\_idx):

target\_idx = **[1번 빈칸]**

if start\_idx > end\_idx:

return -1

if find\_target\_value < total\_list[target\_idx]:

return b\_search(total\_list, find\_target\_value, start\_idx, **[2번 빈칸]**)

elif find\_target\_value > total\_list[target\_idx]:

return b\_search(total\_list, find\_target\_value, **[3번 빈칸]**, end\_idx)

return target\_idx

...

난이도: 중

정답: 3

1. int((start\_idx + end\_idx) / 2), target\_idx+1, target\_idx-1

2. int(start\_idx / 2), target\_idx+1, target\_idx-1

3. int((start\_idx + end\_idx) / 2), target\_idx-1, target\_idx+1

4. int(end\_idx / 2), target\_idx-1, target\_idx+1