

# 2022년 「대구 AI 스쿨」 5월 월말평가 - 심화과정

**제출 시간 : 15:00:00~15:59:59**

**\* 월말 평가 방법 :**

- 오픈북 테스트
- 16시 정답 공개 후 불합격자에 한하여 1회 재응시 진행 예정

**\* 합격 기준 :**

- 총 15점 만점
- 15점~10점 합격, 9점~0점 불합격
- 시간 내 미제출시 불합격

...

점수: 10/15



1

데이터 시각화 및 선형관계와 관련한 설명 중 옳지 않은 것은?

\*

(0/1점)

- ☐ 데이터 품질 및 분포 특성을 확인하기 위해 사용할 수 있다
- ☐ 모델 분석 결과를 설득력 있게 표현하기 위해 사용할 수 있다
- ☒ category/label 별 attribute의 분포를 확인하기 위해 box plot이나 violin plot을 사용할 수 있다

- ☐ scatter plot을 그리거나 상관관계를 계산함으로써, 두 attribute/variable 사이의 선형 관계를 분석할 수 있다
- ☐ 두 변수 사이의 공분산이 음수인 경우, 두 변수는 양의 선형관계를 가진다 ✓

✕

2

아래와 같은 문제를 검증하기 위한 귀무가설, 대립가설로 적절한 것은? (문제 및 보기는 이미지 참고) \*

(0/1점)

아래와 같은 문제를 검증하기 위한 귀무가설( $H_0$ ), 대립가설( $H_1$ )로 적

기존 그룹(A)과 구분되는 대조군 그룹(B)을 두어, B에 대해 새로운 비료의 효과를 평가하고 싶다. 열매가 클 경우 비료의 효능이 좋다고 판단한다. 열매의 크기를 측정하여 각 그룹에 대해 평균과 분산을 계산하였으며, 그 값은  $\mu_A$ ,  $\mu_B$ 와  $\sigma_A^2$ ,  $\sigma_B^2$ 일 때, 비료로 인한 차이가 있는지 확인하기 위한

- ①  $H_0: \mu_A = \mu_B$ ,  $H_1: \mu_A \neq \mu_B$
- ②  $H_0: \mu_A = \mu_B$ ,  $H_1: \mu_A > \mu_B$
- ③  $H_0: \sigma_A^2 = \sigma_B^2$ ,  $H_1: \sigma_A^2 \neq \sigma_B^2$
- ④  $H_0: \sigma_A^2 = \sigma_B^2$ ,  $H_1: \sigma_A^2 > \sigma_B^2$
- ⑤  $H_0: \sigma_A^2 = \sigma_B^2$ ,  $H_1: \sigma_A^2 < \sigma_B^2$

- ☐ ① ✓
- ☐ ②
- ☐ ③
- ☒ ④
- ☐ ⑤

3

Chi-Square Test와 관련한 설명 중 옳지 않은 것은? \*

(1/1점)

- ☐ 주로 count data / contingency table 형태의 자료에 대해 사용
- ☐ expectation과 실제 observed data 간의 차이를 이용하여 통계량을 계산
- ☐ 통계량은 pearson residual의 제곱합으로 계산됨
- ☐ 두 attribute/variable 사이의 독립성을 평가하는 용도로 사용할 수 있음
- ☒ 자유도(degrees of freedom, df)와는 관련이 없음 ✓

4

분산 분석(ANOVA, Analysis of variance anlysis)과 관련된 정보로 틀린 것은? \*

(1/1점)

- ☐ numeric 타입의 데이터에 대해 사용하며 두 개 이상의 그룹들 사이에 통계적으로 유의미한 차이가 있는지 검정
- ☒ t-statistic을 사용함 ✓
- ☐ 편차의 제곱합과 자유도를 사용하여 통계량을 계산함
- ☐ SSE는 SST와 SSR의 차를 이용하여 계산할 수 있음
- ☐ 그룹들을 나누는 요소가 하나인 경우 one-way ANOVA, 두 개인 경우 two-way ANOVA라고 함

5

아래 선형 회귀분석 결과표를 바르게 해석한 것은? \*

(1/1점)

OLS Regression Results							
Dep. Variable:	sales		R-squared:	0.900			
Model:	OLS		Adj. R-squared:	0.899			
Method:	Least Squares		F-statistic:	503.7			
Date:	Mon, 09 May 2022		Prob (F-statistic):	2.09e-83			
Time:	13:47:34		Log-Likelihood:	-359.99			
No. Observations:	171		AIC:	728.0			
Df Residuals:	167		BIC:	740.5			
Df Model:	3						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	3.5059	0.400	8.772	0.000	2.717	4.295	
youtube	0.0452	0.002	29.830	0.000	0.042	0.048	
facebook	0.1884	0.009	20.895	0.000	0.171	0.206	
newspaper	0.0043	0.007	0.658	0.511	-0.009	0.017	
Omnibus:	57.934	Durbin-Watson:	2.065				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	162.792				
Skew:	-1.388	Prob(JB):	4.47e-36				
Kurtosis:	6.891	Cond. No.	544.				

- ☐ 설명 계수가 나와있지 않아 모델 성능을 평가하기 어렵다
- ☐ 독립변수 'newspaper'는 유의한 p-value를 가지고 있다
- ☒ 독립변수 'facebook'이 종속변수에 대해 가장 큰 영향력을 가지고 있다 ✓
- ☐ 독립변수 'youtube'의 계수는 0.002이다.
- ☐ 회귀 방정식의 intercept는 8.772이다.

6

아래 혼동 행렬(confusion matrix)과 지표와 관련하여 틀린 것은

★

(1/1점)

		Predicted condition	
Total population = P + N		Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

$$Precision = \frac{TP}{TP+FP} \quad Recall, Sensitivity = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP} \quad Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- ☐ 정밀도(precision)는 모델이 양성으로 평가한 것 중 실제 양성의 비율이다
- ☐ 재현율(recall)은 실제 양성 중 모델이 양성으로 평가한 것의 비율이다
- ☐ 특이도(specificity)는 실제 음성 중 모델이 음성으로 평가한 것의 비율이다
- ☐ 정확도(accuracy)는 모델이 실제로 옳게 평가한 것의 비율이다
- ☒ 모델 평가 시 정확도만 고려하면 충분하다 ✓

7

다른 값과의 유사도를 이용하는 머신러닝 기법을 바르게 묶은 것은?

\*  
(1/1점)

ㄱ. K-Means    ㄴ. PCA    ㄷ. KNN    ㄹ. Tree    ㅁ. Bayesian

- ☐ ㄱ, ㄴ
- ☒ ㄱ, ㄷ ✓
- ☐ ㄱ, ㅁ
- ☐ ㄴ, ㄷ

☐ 리, □

×

8

주성분 분석(Principal Components Analysis)과 관련된 설명으로 옳지 않은 것은?

\*

(0/1점)

- ☐ 주성분 분석은 변수 여러 개를 선형 방정식 형태로 조합하여 주요 변수로 만드는, 변수 개수를 줄이기 위한 분석 방법임
- ☐ 독립 변수를 사용하지 않는 비지도 학습의 일종 ✓
- ☐ 주성분을 구성하는 변수들의 계수를 통해 주성분의 의미를 분석할 수 있음
- ☒ 주성분 분석 결과로 도출된 변수들을 이용하여, 머신러닝 분석을 동일하게 수행할 수 있음
- ☐ 변수들의 공분산 행렬에 대한 고유값과 고유벡터를 통해 주성분의 설명력과 계수를 계산할 수 있음

×

9

K-means 알고리즘과 관련된 설명으로 옳지 않은 것은?

\*

(0/1점)

- ☐ 군집화(clustering)에 사용되는 알고리즘이다
- ☒ sklearn의 KMeans함수에서, n\_cluster 파라미터로 군집의 개수(k)를 설정할 수 있다
- ☐ 군집의 개수는 3으로 설정하는 것이 최적이다 ✓
- ☐ 모델 학습 시 혼동을 방지하기 위해 scaling으로 데이터를 전처리하는 것이 좋다
- ☐ 데이터 사이의 거리를 계산하여 진행하는 알고리즘이다

10

앙상블(ensemble)과 관련된 설명으로 옳지 않은 것은

\*

(1/1점)

- ☐ 앙상블은 모델 여러 개의 결과물을 종합하여 결론을 내는 방식의 기법으로 하나의 데이터셋에 대해 모델들을 학습시켜 사용함
- ☒ bagging은 데이터셋에 대해 복원 추출한 sample 데이터셋 각각에 대해 모델을 학습시켜 사용함 ✓
- ☐ Random forest tree는 bagging을 적용한 decision tree 모델임
- ☐ boosting은 이전 모델의 학습 결과가 다음 모델 학습에 영향을 주는 방식의 앙상블 기법임
- ☐ boosting으로는 adaboost와 gradient boost 등이 있음

11

인공지능 학습에서 사용하는 정보를 기록하는 annotation에 대한 설명으로 옳지 않은 것은?

\*

(1/1점)

- ☐ 파일의 확장자는 .json, .xml, .txt를 주로 사용한다
- ☐ json 파일을 읽고 쓰기 위해 json 라이브러리를 사용한다
- ☐ xml 파일을 읽고 쓰기 위해 xml 라이브러리를 사용한다
- ☒ txt 파일을 읽고 쓰기 위해 txt 라이브러리를 사용한다 ✓
- ☐ detection용 bbox나 segmentation용 polygon 정보를 기록한다

12

custom Dataset을 위해 작성해야 하는 함수들로 바르게 짰지은 것은?

\*

(1/1점)

ㄱ. `__getitem__`    ㄴ. `__init__`    ㄷ. `__dict__`    ㄹ. `__str__`    ㅁ. `__le`

- ☐ Optimizer: ㄱ, ㄴ, ㄹ
- ☒ Optimizer: ㄱ, ㄴ, ㅁ ✓
- ☐ Optimizer: ㄱ, ㄷ, ㄹ
- ☐ Optimizer: ㄴ, ㄷ, ㄹ
- ☐ Optimizer: ㄴ, ㄹ, ㅁ

13

Classification 문제를 위한 클래스 및 라벨 설정과 모델의 수행결과가 다음과 같을 때, 데이터 객체는 어떤 클래스로 분류되는가?

\*

(1/1점)

(클래스 및 라벨 설정)

```
CLASS_NAME = {'cat': 0, 'dog': 1, 'chick': 2, 'bear': 3, 'horse': 4}
```

(모델 수행 결과)

```
[-13.6,       -10.8,       -0.3,       -8.7,       -15.2]
```

- ☐ cat
- ☐ dog
- ☒ chick ✓
- ☐ bear
- ☐ horse



14

인공지능 학습에서 실제값(Ground Truth, GT)과 예측값의 차이를 계산/수치화하는 함수의 명칭은 무엇인가?

\*

(1/1점)

- ☐ Regression
- ☐ Optimizer
- ☐ Crawling
- ☒ Loss Function ✓
- ☐ Gradient Descent



15

다음 중 Optimizer를 설정할 때, 들어가는 파라미터는?

\*

(0/1점)

- ☒ Epoch
- ☐ Kernel
- ☐ Padding
- ☐ Learning rate ✓
- ☐ Pooling

이 내용은 양식 소유자에 의해 작성됩니다. 제출하는 데이터는 양식 소유자에게 전송됩니다. Microsoft는 이 양식 소유자를 포함한 고객의 개인 정보 보호 또는 보안 관행에 대해 책임을 지지 않습니다. 절대 비밀번호를 알려주지 마세요.

Microsoft Forms에서 제공됨 | [개인정보처리방침](#) | [사용 약관](#)

