

글로벌경영학부 IT파이낸스학과

IT파이낸스프로그래밍 졸업시험 문제은행(총 60문제)

시험문제 객관식 20개

1. 파이썬에서 다른 사람이 만든 기능들을 가져다 쓰기 위해 사용하는 기본 단위는 무엇인가?

- ① 모듈(Module)
- ② 스레드(Thread)
- ③ 커널(Kernel)
- ④ 프레임(Frame)

정답: ①

해설: 파이썬에서는 .py 파일 단위의 코드 묶음을 모듈이라 부르며 import로 불러와 재사용한다.

2. 파이썬으로 데이터 분석을 할 때 주로 사용하는 라이브러리 조합으로 가장 적절한 것은?

- ① Hadoop, Spark, Scala, Kafka
- ② Django, Flask, React, Vue
- ③ Numpy, Pandas, Matplotlib, Seaborn
- ④ Word, Excel, PowerPoint, Outlook

정답: ③

해설: 수치 연산은 Numpy, 표 형식 데이터는 Pandas, 시각화는 Matplotlib과 Seaborn이 주로 사용된다.

3. “바퀴를 다시 발명하지 말라(Don't reinvent the wheel)”는 프로그래밍에서 주로 어떤 태도를 의미하는가?

- ① 운영체제를 직접 개발해서 사용한다.
- ② 항상 모든 코드를 직접 처음부터 작성한다.
- ③ 휠 관련 하드웨어를 직접 설계한다.
- ④ 이미 있는 라이브러리와 도구를 적극 활용한다.

정답: ④

해설: 검증된 라이브러리나 도구를 활용해 생산성을 높이고 오류를 줄이자는 의미이다.

4. 파이썬 파일 확장자로 올바른 것은?

- ① .py
- ② .java
- ③ .exe

④ .cpp

정답: ①

해설: 파이썬 소스 코드는 기본적으로 .py 확장자를 사용한다.

5. 데이터 분석에서 “EDA”의 의미로 가장 알맞은 것은?

- ① 암호적 데이터 암호화
- ② 탐색적 데이터 분석
- ③ 자동화된 데이터 접근
- ④ 외부 데이터 승인

정답: ②

해설: EDA는 데이터를 시각화·요약 통계를 통해 탐색하는 탐색적 데이터 분석을 뜻한다.

6. Colaboratory(Colab)의 특징으로 옳은 것은?

- ① SQL 전용 쿼리 도구이다.
- ② 로컬에서만 작동하는 전용 IDE이다.
- ③ 웹 브라우저에서 파이썬 코드를 실행할 수 있다.
- ④ 하드웨어 설계용 CAD 도구이다.

정답: ③

해설: 구글 Colab은 브라우저에서 파이썬 노트북을 실행할 수 있는 클라우드 환경이다.

7. IT-데이터 관련 직무에서 파이썬이 선호되는 이유로 가장 적절한 것은?

- ① 컴파일 없이 실행할 수 없기 때문이다.
- ② 오직 윈도우에서만 실행되기 때문이다.
- ③ GUI 애플리케이션만 만들 수 있기 때문이다.
- ④ 풍부한 데이터 분석·머신러닝 라이브러리를 제공하기 때문이다.

정답: ④

해설: 파이썬은 Numpy, Pandas, scikit-learn 등 데이터 분석과 머신러닝 라이브러리가 풍부하다.

8. 데이터 과학 직무가 매력적인 이유로 가장 알맞은 것은?

- ① 높은 수요와 경쟁력 있는 연봉 전망을 가진다.
- ② 단순 반복 업무 위주라 스트레스가 적다.
- ③ 비IT 산업에서만 활용된다.
- ④ 통계 지식이 거의 필요 없다고 언급된다.

정답: ①

해설: 데이터 관련 직무가 여러 산업에서 수요가 높고 연봉 수준도 경쟁력이 있다.

9. 소스 코드 파일을 실행 가능한 기계어로 바꾸는 프로그램은 무엇인가?

- ① 스위치(Switch)
- ② 브라우저(Browser)
- ③ 컴파일러(Compiler)
- ④ 파서(Parser)

정답: ③

해설: 컴파일러는 고급 언어로 작성된 소스 코드를 기계어(실행 파일)로 번역한다.

10. 소프트웨어 버전 관리 시스템(VCS)의 대표적인 용도는 무엇인가?

- ① 데이터베이스 백업만을 담당한다.
- ② 프로그램 실행 속도를 높여 준다.
- ③ 하드웨어 온도를 제어한다.
- ④ 코드 변경 이력을 관리하고 협업을 돕는다.

정답: ④

해설: VCS는 변경 이력 기록, 브랜치 관리 등을 통해 협업과 코드 품질 관리를 지원한다.

11. Kaggle 플랫폼의 주요 목적은 무엇인가?

- ① 웹 브라우저 개발
- ② 데이터 분석·머신러닝 대회와 코드 공유
- ③ 운영체제 개발 협업
- ④ 하드웨어 설계 공모전

정답: ②

해설: Kaggle은 데이터 분석·머신러닝 대회, 데이터셋 공유, 노트북 공유를 제공하는 플랫폼이다.

12. Kaggle의 Competitions, Datasets, Notebooks, Discussion 중 “코드 실행 환경과 예제” 중심의 탭은 무엇인가?

- ① Datasets
- ② Competitions
- ③ Notebooks
- ④ Discussion

정답: ③

해설: Notebooks 탭에서는 커널 또는 노트북 형태로 코드를 실행하고 공유할 수 있다.

13. Kaggle 사용자 등급 체계에 포함되는 단계가 아닌 것은?

- ① Grandmaster
- ② Novice
- ③ Master
- ④ Beginner

정답: ④

해설: 자료에는 Novice, Contributor, Expert, Master, Grandmaster와 같은 등급이 소개된다.

14. Dacon이 제공하는 것은 무엇인가?

- ① 전자상거래 결제 서비스
- ② 소프트웨어 설치 관리자
- ③ 클라우드 인프라 제공 업체
- ④ 국내 중심의 데이터 분석·AI 경진대회 플랫폼

정답: ④

해설: Dacon은 국내 기업·기관이 주최하는 데이터 분석·AI 경진대회를 제공하는 플랫폼이다.

15. 지도학습(Supervised Learning)의 특징으로 가장 올바른 것은?

- ① 정답이 없는 데이터에서 군집을 찾는다.
- ② 정답(레이블)이 있는 데이터를 이용해 모델을 학습한다.
- ③ 보상을 통해 행동 전략을 학습한다.
- ④ 데이터를 무작위로 섞는 과정만 포함한다.

정답: ②

해설: 지도학습은 입력 데이터와 정답 레이블을 함께 사용해 입력-출력 관계를 학습한다.

16. 비지도학습(Unsupervised Learning) 알고리즘의 예로 가장 적절한 것은?

- ① K-Means 클러스터링
- ② 선형 회귀
- ③ 로지스틱 회귀
- ④ 의사결정나무 분류

정답: ①

해설: K-Means는 정답 레이블 없이 데이터의 군집 구조를 찾는 비지도학습 알고리즘이다.

17. 강화학습(Reinforcement Learning)에 대한 설명으로 옳은 것은?

- ① 단순 통계 요약만 수행한다.

- ② 항상 정답 레이블이 필요하다.
- ③ 환경과 상호작용하며 보상을 최대화하도록 정책을 학습한다.
- ④ 군집 개수 k 를 미리 정해야 한다.

정답: ③

해설: 강화학습은 에이전트가 환경에서 행동을 선택하고, 장기적인 보상을 최대화하는 정책을 학습한다.

18. 머신러닝에서 “피처(Feature)”에 대한 설명으로 알맞은 것은?

- ① 데이터 저장 형식
- ② 모델이 출력하는 예측값
- ③ 모델 학습에 사용되는 손실 함수
- ④ 모델이 학습에 사용하는 입력 변수

정답: ④

해설: 피처는 모델이 입력으로 사용하는 설명 변수 또는 특성이다.

19. Kaggle-Dacon 대회를 활용하는 학습 전략으로 옳은 것은?

- ① 항상 가장 복잡한 모델만 사용한다.
- ② 대회 순위만 보고 이론 학습은 생략한다.
- ③ 실제 데이터를 통해 이론과 실습을 동시에 경험한다.
- ④ 코드 제출 없이 결과만 올린다.

정답: ③

해설: 실전 데이터와 문제를 다루면서 이론을 적용해 보는 것이 효과적인 학습 전략으로 소개된다.

20. “데이터 to 가치(DATA TO VALUE)”라는 표현이 의미하는 바에 가장 가까운 것은?

- ① 데이터를 단순히 시각화만 한다.
- ② 데이터를 저장만 하고 활용하지 않는다.
- ③ 데이터를 분석해 실제 비즈니스·사회적 가치를 창출한다.
- ④ 데이터를 모두 삭제해 보안을 강화한다.

정답: ③

해설: 슬로건은 데이터를 활용해 인사이트와 경제적 가치를 만들어 내는 것을 강조한다.

21. 회귀분석(Regression Analysis)의 주요 목적은 무엇인가?

- ① 그래프를 자동 생성한다.
- ② 범주형 변수를 군집화한다.
- ③ 텍스트를 분류한다.
- ④ 연속형 종속변수를 예측·설명한다.

정답: ④

해설: 회귀분석은 하나 이상의 독립변수로 연속형 종속변수의 변화를 설명하고 예측한다.

22. 단순선형회귀(Simple Linear Regression)는 어떤 경우에 사용되는가?

- ① 독립변수가 1개이고 종속변수가 연속형일 때
- ② 독립변수가 여러 개이고 종속변수가 범주형일 때
- ③ 종속변수가 여러 개일 때
- ④ 독립·종속변수가 모두 범주형일 때

정답: ①

해설: 단순선형회귀는 독립변수 1개와 연속형 종속변수 간 선형 관계를 모델링한다.

23. 다중선형회귀(Multiple Linear Regression)의 특징으로 옳은 것은?

- ① 여러 독립변수가 종속변수에 선형적으로 영향을 미친다고 가정한다.
- ② 독립변수는 항상 하나만 존재해야 한다.
- ③ 종속변수는 범주형이어야 한다.
- ④ 군집 개수를 찾는 데 사용된다.

정답: ①

해설: 다중선형회귀는 여러 독립변수를 동시에 사용해 종속변수를 설명하는 모델이다.

24. 회귀분석에서 독립변수 x 와 종속변수 y 의 관계를 표현할 때 사용하는 기본 가정은 무엇인가?

- ① x 가 y 를 항상 감소시킨다.
- ② y 가 x 와 무관하게 일정하다.
- ③ y 가 x 에 대해 선형으로 변한다.
- ④ x 와 y 는 항상 음의 상관만 가진다.

정답: ③

해설: 선형회귀는 x 와 y 사이가 직선 형태의 선형 관계라는 가정을 기반으로 한다.

25. 회귀분석에서 2010-2020년 기상 데이터(weather.xlsx)를 사용하는 이유로 알맞은 것은?

- ① 범주형 설문 응답을 분류하기 위해
- ② 시간에 따라 변화하는 연속형 변수의 추세를 분석하기 위해
- ③ 이미지 인식을 위해
- ④ 텍스트 요약을 위해

정답: ②

해설: 연도별 기상 데이터를 이용하면 시간에 따른 변수들의 추세와 관계를 회귀로 분석할 수 있다.

26. 회귀분석에서 결정계수 R^2 값이 0.99에 가깝다는 설명은 무엇을 의미하는가?

- ① 표본 수가 99개라는 뜻이다.
- ② 데이터의 99%가 이상치라는 뜻이다.
- ③ 설명력이 매우 낮다는 뜻이다.
- ④ 모델이 종속변수 변동의 약 99%를 설명한다.

정답: ④

해설: R^2 값은 독립변수가 종속변수 변동을 얼마나 설명하는지를 나타내며 0.99면 설명력이 매우 높다는 의미이다.

27. 회귀분석에서 잔차(residual)에 대한 설명으로 가장 적절한 것은?

- ① 입력 변수들의 합
- ② 실제 값과 예측 값의 차이
- ③ 예측 값들의 평균
- ④ 상관계수의 제곱근

정답: ②

해설: 잔차는 관측값에서 회귀선의 예측값을 뺀 값이다.

28. 회귀 모델을 평가할 때 과적합을 확인하는 일반적인 방법은?

- ① 모델을 전혀 학습시키지 않는다.
- ② 학습 데이터만 여러 번 반복해서 본다.
- ③ 피처를 무작정 많이 추가한다.
- ④ 학습 데이터와 시험 데이터의 성능 차이를 비교한다.

정답: ④

해설: 학습 데이터에선 성능이 매우 좋고 시험 데이터에선 나쁘면 과적합 가능성이 크다.

29. 단순선형회귀에서 기울기(slope)의 부호가 음수라는 것은 어떤 의미인가?

- ① x 와 y 가 전혀 관련이 없다.
- ② x 가 증가할수록 y 가 감소하는 경향이 있다.
- ③ x 가 증가할수록 y 도 증가한다.
- ④ y 가 항상 0이 된다.

정답: ②

해설: 음의 기울기는 독립변수가 증가할 때 종속변수가 감소하는 반비례 관계를 의미한다.

30. 회귀분석 실습에서 사용하는 weather.xlsx 파일의 데이터 구조를 이해하는 데 가장 유용한 Pandas

메서드는?

- ① .groupby()
- ② .drop()
- ③ .head()
- ④ .info()

정답: ④

해설: info() 메서드는 행 수, 열 수, 데이터 타입, 결측치 여부 등을 한 번에 보여 준다.

31. 상관계수(correlation coefficient)의 범위로 옳은 것은?

- ① -100에서 100 사이
- ② 0에서 100 사이
- ③ -1에서 1 사이
- ④ 1 이상

정답: ③

해설: 피어슨 상관계수는 -1에서 1 사이 값을 가지며, 절댓값이 클수록 선형 관계가 강하다.

32. 상관계수 Heatmap을 그릴 때 주로 사용하는 Python 라이브러리는?

- ① Flask
- ② TensorFlow
- ③ Seaborn
- ④ BeautifulSoup

정답: ③

해설: Seaborn의 heatmap 함수는 상관계수 행렬을 색상으로 시각화하는 데 자주 사용된다.

33. 군집화(Clustering)의 대표적인 목적은 무엇인가?

- ① 시계열 분해를 수행하는 것
- ② 회귀 계수를 추정하는 것
- ③ 텍스트를 번역하는 것
- ④ 비슷한 특성을 가진 데이터들을 그룹으로 묶는 것

정답: ④

해설: 군집화는 유사한 데이터 포인트들을 군집으로 묶어 데이터 구조를 파악하는 비지도학습 기법이다.

34. K-Means 알고리즘에서 k는 무엇을 의미하는가?

- ① 데이터 차원 수
- ② 반복 횟수

- ③ 만들고자 하는 군집의 개수
- ④ 학습률

정답: ③

해설: K-Means의 k는 형성할 군집의 개수를 의미하며 사전에 지정해야 한다.

35. scikit-learn에서 K-Means를 사용하기 위한 클래스 이름으로 옳은 것은?

- ① sklearn.cluster.KMeans
- ② sklearn.cluster.KNN
- ③ sklearn.linear_model.KMeans
- ④ sklearn.neighbors.KMeans

정답: ①

해설: K-Means 군집화는 sklearn.cluster 모듈의 KMeans 클래스로 구현된다.

36. inertia 값은 K-Means에서 무엇을 나타내는가?

- ① 각 데이터와 해당 군집 중심 간 거리 제곱합
- ② 모델의 정확도
- ③ 데이터 개수
- ④ 학습률

정답: ①

해설: inertia는 모든 포인트에서 자신이 속한 클러스터 중심까지의 거리 제곱합으로 군집 응집도를 나타낸다.

37. 엘보우(Elbow) 방법의 기본 아이디어는 무엇인가?

- ① 항상 k=2가 최적이라고 가정한다.
- ② 가장 큰 k를 선택해야 한다.
- ③ k 증가에 따른 inertia 감소 곡선에서 꺾이는 지점을 최적 군집 수로 본다.
- ④ 정확도가 최대인 k를 택한다.

정답: ③

해설: inertia 감소 폭이 급격히 줄어드는 지점을 ‘팔꿈치’로 보고 적정 k로 선택하는 방법이다.

38. yellowbrick 라이브러리의 KElbowVisualizer를 사용하는 주된 목적은?

- ① 데이터를 정규화하기 위해
- ② 회귀선의 기울기를 구하기 위해
- ③ 시계열의 계절성을 분석하기 위해
- ④ 시각적으로 적정 군집 수 k를 찾기 위해

정답: ④

해설: KElbowVisualizer는 다양한 k에 대한 점수 변화를 시각화해 적절한 군집 수를 고르는 데 도움을 준다.

39. 계층적 군집(Hierarchical clustering)에서 데이터 간 결합 순서를 시각화한 트리를 무엇이라 부르는가?

- ① 히스토그램(histogram)
- ② 덴드로그램(dendrogram)
- ③ 박스플롯(boxplot)
- ④ 파레토 차트

정답: ②

해설: 덴드로그램은 병합 또는 분할 과정을 계층 구조로 표현한 트리형 도표이다.

40. scipy.cluster.hierarchy.linkage 함수의 주요 역할은 무엇인가?

- ① 회귀 계수 추정
- ② 계층적 군집을 위한 클러스터 병합 정보 계산
- ③ KNN 분류 수행
- ④ PCA 차원 축소

정답: ②

해설: linkage 함수는 데이터 간 거리와 병합 방식을 기준으로 계층적 군집 구조를 계산한다.