

4과목 빅데이터 결과해석

61. 다음 중 ROC 곡선 축을 구성하는 지표로 바르게 짝지어진 것은?

- ① 정확도, 정밀도
- ② 정밀도, 특이도
- ③ 특이도, 민감도
- ④ 재현율, f1-score

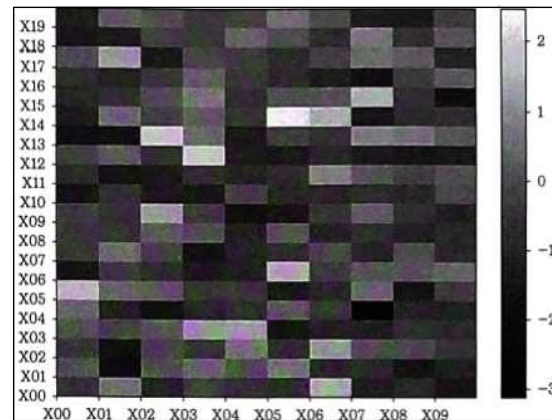
62. 다음 중 교차 검증에 대한 설명으로 옳지 않은 것은?

- ① 훈련 데이터, 검증 데이터, 테스트 데이터의 비율은 보통 2 : 3 : 5의 비율로 구성된다.
- ② 홀드아웃 교차 검증은 데이터를 무작위로 7:3 또는 8:2 비율로, 학습 데이터와 검증 데이터로 나누는 방법이다.
- ③ 교차 검증은 과적합을 방지하기 위해 사용된다.
- ④ 데이터의 수가 적은 경우에 사용될 수 있다.

63. 다음 중 최종 모델을 평가하는 기준으로 옳지 않은 것은?

- ① 평가 지표
- ② 업무 관계자의 의견
- ③ 시스템 구현 가능성
- ④ 표본의 충분성

64. 다음과 같은 차트가 나타내는 시각화 기법은?



- ① 트리맵
- ② 카토그램

- ③ 히트맵
- ④ 산점도 행렬

65. 다음 중 분류 모형 평가에 대한 설명으로 옳지 않은 것은?

- ① F1-Score는 정밀도와 재현율의 조화평균 값이다.
- ② 혼동행렬에서 모델이 참으로 예측한 수치는 TP+FP이다.
- ③ ROC Curve로 혼동행렬을 구할 수 있다.
- ④ AUC 값이 1에 가까울수록 모델의 분류 성능이 좋다.

66. 다음 중 분석 모형 평가 지표에 대한 수식으로 옳지 않은 것은?

- ① MAE :  $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$
- ② MSE :  $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$
- ③ MAPE :  $\frac{100}{n} \times \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$
- ④ RMSE :  $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

67. 다음 중 정규성 검정 기법으로 옳지 않은 것은?

- ① 샤피로-윌크 검정
- ② Q-Q plot
- ③ 콜모고로프-스미르노프 검정
- ④ 카이제곱 검정

68. 다음 중 데이터 시각화에 대한 설명으로 옳지 않은 것은?

- ① 데이터 구조화 단계에서는 시각화를 위한 요건을 정의하고, 사용자에게 따른 시나리오를 작성한다.
- ② 데이터 시각표현 단계에서는 데이터 모델링을 수행한다.
- ③ 데이터 시각화 단계에서는 여러 변수를 비교하여 분석 정보의 시각화를 구현한다.
- ④ 데이터 시각표현 단계에서는 데이터가 목적과 의도에 맞게 시각적으로 잘 표현되었는지 확인한다.

69. 다음 중 비교 시각화 기법이 아닌 것은?

- ① 히스토그램
- ② 스타차트
- ③ 플로팅 바 차트
- ④ 체르노프페이스

70. 다음 중 관계 시각화 기법이 아닌 것은?

- ① 산점도 행렬                      ② 누적막대그래프
- ③ 네트워크그래프                ④ 버블차트

71. 민감도가 0.6, 정밀도가 0.4인 경우 F1-Score는 얼마인가?

- ① 0.24                      ② 0.48
- ③ 0.5                        ④ 0.6

72. 다음 앙상블 모형에 대한 설명 중 옳은 것을 모두 고른 것은?

가. 랜덤 포레스트가 대표적인 앙상블 모델이다.  
나. 배깅은 부트스트랩 샘플을 사용한다.  
다. 부스팅은 정답에 더 높은 가중치를 적용하여 모델의 성능을 높이는 방법이다.

- ① 가                      ② 가, 나, 다
- ③ 나, 다                ④ 가, 나

73. 다음 중 신경망 모델에서 발생하는 기울기 소실 문제에 대한 설명으로 옳은 것은?

- ① 오차 역전파 과정에서 기울기가 감소하여 가중치가 업데이트되지 않은 현상을 말한다.
- ② 은닉층의 활성화 함수로 시그모이드 함수를 사용하면 문제가 완화된다.
- ③ 그래디언트 클리핑을 하면 문제가 완화된다.
- ④ 신경망 학습 과정에서 기울기가 점차 커지다가 발산하는 현상이다.

74. 다음 중 재현율 공식으로 옳은 것은?

- ① 가지분할은 의사결정나무에서 나무의 가지를 생성하는 과정이다.
- ② 의사결정나무의 해석이 어려운 이유는 계산 결과가 의사결정나무에 직접적으로 나타나지 않기 때문이다.
- ③ 의사결정나무는 전체 자료를 몇몇의 소집단으로 분류하거나 예측하는 방법이다.
- ④ 연속적으로 발생하는 의사결정 문제를 시각화해서 의사결정이 이루어지는 시점과 성과 파악이 쉽다.

75. 의사결정나무에 대한 설명으로 옳지 않은 것은?

- ①  $\frac{TP}{(TP + FN)}$                       ②  $\frac{FP}{(FP + TN)}$
- ③  $\frac{TP}{(TP + FP)}$                       ④  $\frac{FN}{(FN + TN)}$

76. 다음 중 인포그래픽에 대한 설명으로 옳지 않은 것은?

- ① 인포그래픽은 정보 제공자가 전달하고자 하는 주요한 정보를 하나의 그래픽으로 표현하여 보는 사람들이 쉽고 빠르게 정보를 이해할 수 있도록 만든 시각화 방법이다.
- ② 인포그래픽은 정보를 SNS 상에 쉽고 빠르게 전달할 수 있다.
- ③ 전문 분야에 대한 데이터를 전달하기 위해서는 전문적 용어를 위주로 사용하여 표현한다.
- ④ 복잡한 데이터를 쉽게 이해할 수 있도록 그래픽과 텍스트를 적절하게 조합하여 표현한다.

77. 다음 중 특정 기준에 따라 회귀계수에 별점을 부여하여 모형의 복잡도를 낮추는 분석 기법은?

- ① 랜덤 포레스트
- ③ 로지스틱 회귀
- ② 별점화 회귀
- ④ 다항선형 회귀

78. 다음 중 과대적합에 대한 설명으로 옳지 않은 것은?

- ① 비선형 모형은 선형 모형보다 과대적합 발생 가능성이 낮다.
- ② 과대적합은 모형이 과도하게 복잡해진 상태이다.
- ③ 과대적합 모형은 분산이 크다.
- ④ 과대적합 모형은 일반화 성능이 낮다.

79. 다음 중 데이터 시각화 절차에 해당하는 요소가 아닌 것은?

- ① 구조화
- ② 시각화
- ③ 시각표현
- ④ 데이터 보충

80. 다음 중 설명력이 가장 좋은 ROC 곡선은?

- ① AUC : -0.95
- ② AUC : 0.88
- ③ AUC : 0.77
- ④ AUC : 0.5