

## 2024년 가을 송실대학교 AI융합학부 머신러닝 기말고사

주의사항: 1. 문제지는 양면입니다. 뒷면에도 문제가 있습니다.

2. 모든 답은 답안지에 작성해야 합니다 (문제지는 채점하지 않습니다). 답안지는 앞 페이지에만 작성해야 합니다 (답안지 뒷면 채점하지 않습니다). 답안지 모든 페이지에 학번, 이름, 분반을 써 주세요

### 1. 다음 문장이 맞는지 틀리는지 O/X로 답하시오 (각 2점, 작성했으나 틀리면 -1점)

- ① \_\_\_\_\_ 비지도학습은 학습 시에 정답 셋이 필요하지 않다
- ② \_\_\_\_\_ 로지스틱 함수의 최종 출력 값은 [0,1] 이다
- ③ \_\_\_\_\_ 경사 하강법은 문제의 최적 해를 항상 찾아준다
- ④ \_\_\_\_\_ 결정 트리 모델의 깊이는 깊어질수록 좋다
- ⑤ \_\_\_\_\_ 모든 결정 트리는 항상 이진 트리 구조를 유지한다
- ⑥ \_\_\_\_\_ 이상치가 많을 때는 RMSE (혹은 MSE)가 MAE보다 왜곡의 정도를 민감하게 표현한다
- ⑦ \_\_\_\_\_  $\hat{w} = (X^T X)^{-1} X^T y$ 의 최소 자승법으로 선형회귀의 답을 항상 구할 수 있다.
- ⑧ \_\_\_\_\_ 배깅에 비해 부스팅은 학습 시간이 오래 걸린다
- ⑨ \_\_\_\_\_ L1 정규화는 파라미터 값의 제곱을 사용하는 정규화이다.
- ⑩ \_\_\_\_\_ numpy의 array에 scalar 값을 더하는 것이 가능하다.

### 2. 아래 문제의 빈 칸을 채우시오 (답안지에 답하시오, 각 3점)

- ① \_\_\_\_\_은(는) 비지도학습의 일종으로 데이터에 대해 따로 분류기준을 주지 않고 모델이 스스로 분류 기준을 찾아 집단을 모으는 기계학습 기법을 의미한다
- ② \_\_\_\_\_은(는) 문제와 답을 함께 학습함으로 미지의 문제에 대해 올바른 답을 예측한다
- ③ 데이터의 "특성"이나 "특징"을 나타내는 \_\_\_\_\_은(는) 주로 기계학습의 입력으로 사용된다
- ④ 클래스나 프로세스 데이터 등의 현재 생성된 객체를 우리는 \_\_\_\_\_ (이)라고 부른다 (힌트: 인간은 클래스, 한 사람은 (예를 들어 윤진혁 교수는) 이것)
- ⑤ \_\_\_\_\_은(는) numpy의 핵심 중 하나로, 서로 다른 크기와 차원을 가진 array들의 연산을 정의해준다.
- ⑥ 두 개의 데이터 프레임을 merge로 병합할 때 매개변수로 how=\_\_\_\_\_을(를) 넣으면 두 데이터프레임에 공통된 값만 새 데이터프레임에 기록해준다

- ⑦ \_\_\_\_\_은(는) 함수의 기울기(경사)를 구하고, 그 경사의 반대 방향으로 계속 이동시키는 것을 (일정 수준 오차 이하로) 극값에 이를 때까지 반복하는 최적화 방법이다.
- ⑧ \_\_\_\_\_함수는 최종 출력값을 [0, 1]로 가지며 분류 모델에 쓰이는 \_\_\_\_\_회귀에 사용된다.
- ⑨ \_\_\_\_\_은(는) 불확실성(균질성)을 나타내는 척도로, Shannon이 제안하였고, 트리 모델에 쓰인다.
- ⑩ 나이브 베이즈 분류기에서 나이브의 의미는 각각의 피쳐가 모두 \_\_\_\_\_임을 가정한다는 뜻이다.

### 3. 아래 주관식 문제를 푸세요 (과정이 있다면 과정도 같이 써 주세요, 각 5점)

- ① 전처리 스케일링 전략 중 z-score 정규화와 최댓값-최소값 정규화에 대해 설명하세요.
- ② 다음 코드의 예상되는 결과값을 구해보세요.

```
import numpy as np

test_a = np.arange(5,14).reshape(3,3)

test_b = np.arange(6).reshape(2,3)

test_a.dot(test_b.T)
```

- ③ 회귀식의 성능을 측정하기 위해 사용되는 MAE(Mean Absolute Error), MSE (Mean Squared Error), 결정계수(R-squared)에 대해 설명하세요.
- ④ 전체-배치 경사하강법(full-batch gradient descent) 방식과 미니-배치 경사하강법(minibatch gradient descent) 방식의 차이점 대해 설명하세요.
- ⑤ 다중클래스 분류와 다중레이블 분류의 차이에 대해 설명하고, 각각에 대한 적절한 예시를 들어주세요
- ⑥ One-vs-One과 One-Vs-Rest (혹은 One-Vs-All)의 차이에 대해 설명하세요. N개의 분류가 있다면 각각 몇 개의 분류기를 만들어야 하나요?
- ⑦ 다음 통계를 기반으로 학생이 도서관에 가지 않았을 때, 학사경고를 받을 확률은 얼마인가요?

모든 학생 중 10%가 학사경고를 받는다.

학사경고를 받은 학생 중 90%가 도서관에 간적이 없다.

학사경고를 받지 않은 학생 중 10%가 도서관에 간적이 없다.

아래의 ⑧⑨⑩세 문제는. 표로 주어진 데이터를 가지고 풀어주세요

<b>x1</b>	0	1	0	1	0	1	0	1
<b>x2</b>	0	0	0	0	1	1	1	1

<b>x3</b>	0	0	1	1	0	0	1	1
<b>y</b>	0	0	1	1	1	1	0	0

- ⑧ 정보 엔트로피의 식은  $h(D) = -\sum_i^m p_i \log_2 p_i$  로 주어집니다. y의 정보 엔트로피를 구하세요
- ⑨ x1, x2, x3 각각으로 분할하는 경우 Information Gain을 구하세요.
- ⑩ 정보 이득을 기준으로 완전한 트리를 구하세요 (Information Gain이 동률인 경우 X3, X2, X1의 우선순위로 분할)

한 학기 동안 고생 많으셨습니다. 지금의 시국에 시험을 보러 오라고 하는 것이 참 죄송할 따름입니다. 고생하신 만큼 좋은 결과 있으셨으면 좋겠습니다. 즐거운

방학 보내세요