

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

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

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

- ① 군집(클러스터링) 은(는) 비지도학습의 일종으로 데이터에 대해 따로 분류기준을 주지 않고 모델이 스스로 분류 기준을 찾아 집단을 모으는 기계학습 기법을 의미한다
- ② 지도학습 은(는) 문제와 답을 함께 학습함으로 미지의 문제에 대해 올바른 답을 예측한다
- ③ 데이터의 “특성”이나 “특징”을 나타내는 피처(feature) 은(는) 주로 기계학습의 입력으로 사용된다
- ④ 클래스나 프로세스 데이터 등의 현재 생성된 객체를 우리는 인스턴스 (이)라고 부른다 (힌트: 인간은 클래스, 사람은 이것)
- ⑤ ufunc (universal functions) 은(는) numpy의 핵심 중 하나로, 서로 다른 크기와 차원을 가진 array들의 연산을 정의해 준다.
→ 의도는 ufunc이었으나, 브로드캐스트로 작동하는 것도 맞으므로 브로드캐스트도 인정
- ⑥ 두 개의 데이터 프레임을 merge로 병합할 때 매개변수로 how= “inner” 을(를) 넣으면 두 데이터프레임에 공통된 값만 새 데이터프레임에 기록해준다
→ 이런 방식을 inner join 이라 하므로 inner join이라고 한 경우도 인정
- ⑦ 경사하강법(Gradient Descent) 은 함수의 기울기(경사)를 구하고, 그 경사의 반대 방향으로 계속 이동시키는 것을 극값에 이를 때 까지 반복하는 최적화 방법이다.
- ⑧ 로지스틱 함수는 최종 출력값을 $(0, 1)$ 로 가지며 분류 모델에 쓰이는 로지스틱 회귀에 사용된다.
→ 소프트맥스는 인정. 시그모이드는 출력값이 $(0, 1)$ 이 아닌 함수들도 포함되므로 불인정. 두 개를 서로 다르게 쓴 경우에 둘 중 하나가 맞다면 정답 인정
- ⑨ 정보 엔트로피 는 불확실성(균질성)을 나타내는 척도로, Shannon이 제안하였고, 트리 모델에 쓰인다.
→ Shanon이라는 정보를 이미 제시하였으므로 지니 불인정. 정보엔트로피가 완전한 정답이나, 엔트로피라고만 써도 인정 (사실 정보엔트로피라고 쓴 사람이 몇 없습니다)
- ⑩ 나이브 베이즈 분류기에서 나이브의 의미는 각각의 피처가 모두 독립사건 임을 가정한다는 뜻이다.
→ 독립사건, 독립, 상호 독립적 등 “독립”의 의미가 담기면 일단 인정

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 (5점)
 - z스코어 정규화: 표준편차와 평균을 이용하여 정규화 하는 방법 $z = \frac{x - \mu}{\sigma}$
 - 최댓값-최소값 정규화: 데이터의 최댓값과 최소값을 기준으로 정규화 하는 방법 $x' = \frac{x - \min(x)}{\max(x) - \min(x)}$
 - 두 정규화 각 3점. 2개다 맞추는 경우 5점. 식이 써 있거나, 정확한 정규화 방식이 써 있어야 인정하고, 예를 들어 단순히 “데이터의 최대와 최소값을 사용하는 정규화” 형태는 인정하지 않음.

② 다음 코드의 예상되는 결과값을 구해보세요 (5점)

```
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)
```

답:

```
array([[ 20,  74],
       [ 29, 110],
       [ 38, 146]])
```

➔ test_a 값 1점, test_b 값 1점 shape 1점. test_a.dot(test_b.T) 값 1점, shape 1점. 과정 없이 답만 쓴 경우는 전체 점수 불인정 (0점) (3번 문제 맨 위에 “과정 써 달라” 는 부분 참고)

③ 회귀식의 성능을 측정하기 위해 사용되는 MAE(Mean Absolute Error), MSE (Mean Squared Error), 결정계수(R-squared)에 대해 설명하세요 (5점)

- MAE: 예측값과 실제값 차이의 절대값을 모두 더해 평균을 내어 계산

- MSE: 예측값과 실제값 차이의 제곱을 모두 더해 평균을 내어 계산

- 결정계수: $\frac{SSE}{SST} = 1 - \frac{SSR}{SST}$. 기울기가 0이라는 가설에 비해 현재 가설이 관측값을 얼마나 더 잘 설명해주는지 계산

➔ 각 2점, 3개 모두 맞추면 5점. 결정계수는 Pearson r의 제곱값 이라는 답도 인정. 단순히 한국어 이름만 쓴 경우 (예를 들어 “평균 제곱 오차”)는 불인정

④ 전체-배치 경사하강법(full-batch gradient descent) 방식과 미니-배치 경사하강법(minibatch gradient descent) 방식의 차이점 대해 설명하세요 (5점)

- 전체 배치 경사하강법: 모든 데이터 (full-batch)의 기울기를 한 번에 계산한 후 종합하여 학습한다

- 미니-배치 경사하강법: 데이터를 여러 개로 분할한 미니-배치(mini-batch) 단위로 학습한다

➔ 각 3점. 모두 맞추면 5점. “샘플링”은 미니 배치의 방식과는 다르므로 “분할”의 개념이 들어가야 인정. 무작위 선택 불인정 (이건 SGD, SGD+ mini batch가 가능하긴 하지만 문제는 mini-batch만 질문 함). Epoque등의 설명을 제대로 하는 경우에만 미니배치를 인정

⑤ 다중클래스 분류와 다중레이블 분류의 차이에 대해 설명하고, 각각에 대한 적절한 예시를 들어주세요

다중클래스 분류는 한 객체가 한 개의 클래스만 가질 수 있고 (예시: 동물의 종 등), 다중레이블 분류는 한 객체가 여러 개의 클래스에 속할 수 있다 (예시: 인스타그램 해시태그)

➔ 다중클래스/레이블 차이 제대로 설명하면 3점. 예시 각 1점. 단순히 [0,1,1] 형태의 예시는 설명을 벡터로 반복하는 것 뿐으로 예시로 보지 않고 불인정.

⑥ One-vs-One과 One-Vs-Rest (혹은 One-Vs-All)의 차이에 대해 설명하세요. N개의 분류가 있다면 각각 몇 개의 분류기를 만들어야 하나요?

OVO분류는 N개의 클래스에 대해서 두 개 클래스씩 뽑아 모든 조합에 대한 분류기를 만드므로 $\frac{N(N-1)}{2}$ 개의 분류기가 필요하고, OVR은 1개의 정답 클래스 이외의 것을 모두 오답으로 가증하므로 N개의 분류기가 필요하다.

➔ 분류기 설명에 3점, 분류기의 수 각 1점 (총 2점). 분류기 수만 쓰는 경우는 2점만 인정.

⑦ 다음 통계를 기반으로 학생이 도서관에 가지 않았을 때, 학사경고를 받을 확률은 얼마인가?

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.

답: H: 학교, L: 도서관

$$\begin{aligned} P(H|\sim L) &= \frac{P(\sim L|H)P(H)}{P(\sim L)} \\ &= \frac{P(\sim L|H)P(H)}{P(\sim L|H)P(H) + P(\sim L|\sim H)P(\sim H)} \\ &= \frac{0.9 \times 0.1}{0.9 \times 0.1 + 0.9 \times 0.1} = 0.5 \end{aligned}$$

➔ 베이시안을 의도했으나 다른 형태 풀이도 인정. 과정 없으면 무조건 0점. 과정이 있고 답이 맞았는데 과정이 정답을 이끌어내는데 불충분하다 판단되는 경우 2점만 인정. 베이시안이 아닌 노가다에서 정답이 틀리면 0점. 과정 없으면 무조건 0점.

아래의 ⑧⑨⑩세 문제는. 표로 주어진 데이터를 가지고 풀어주세요

x1	0	1	0	1	0	1	0	1
x2	0	0	0	0	1	1	1	1
x3	0	0	1	1	0	0	1	1
y	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

$$h(D) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = +\frac{1}{2} + \frac{1}{2} = 1$$

➔ 식이 이미 문제에 주어졌으므로 답이 완전히 구해 져야만 5점 인정. $\log_2 = 1$ 을 사용하지 않고 \log 를 남기고 끝낸 경우도 불인정. 과정 없으면 무조건 0점.

- ⑨ x1, x2, x3 각각으로 분할하는 경우 Information Gain을 구하세요.

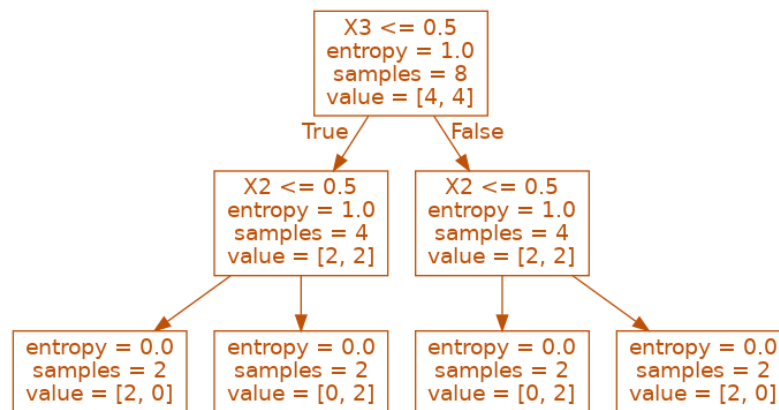
i. X₁분할시 [2,2], [2,..2]로 분할됨: $h_{x_1}(D) = \frac{1}{2} \times 1 + \frac{1}{2} \times 1 = 1, IG = 0$

ii. X₂분할시 [2,2], [2,..2]로 분할됨: $h_{x_2}(D) = \frac{1}{2} \times 1 + \frac{1}{2} \times 1 = 1, IG = 0$

iii. X₃분할시 [2,2], [2,..2]로 분할됨: $h_{x_3}(D) = \frac{1}{2} \times 1 + \frac{1}{2} \times 1 = 1, IG = 0$

➔ 각 2점. 다 구하면 5점. 세 계산이 모두 같은 형태이므로, 계산을 실수했으나 IG의 개념을 알면 2점 인정. 단순히 X_i 분할의 하위 엔트로피만 구하는 것은 불인정 (문제의 의도는 IG가 무엇인지 이해하는 것임). 과정 없으면 무조건 0점.

- ⑩ 정보 이득을 기준으로 완전한 트리를 구하세요. Information Gain이 동률인 경우 X3, X2, X1의 우선순위로 분할한다 가정합니다.



➔ 분할에 대한 IG 계산 혹은 IG에 대한 설명이 없으면 0점 (즉, 과정 없으면 무조건 0점). 만약 분할 시 IG값이 틀렸으나, 이 오류의 원인이 9번이고 X_i 분할단위의 IG가 모두 같다면 정답 인정. 트리는 제대로 그렸으나 분할하지 않아야 depth를 추가로 분할하거나 덜 분할되거나 하는 경우는 2점만 인정