

< CUI 4기 BASIC 트랙 야금야금 머신러닝 2회차 >

공통 교재인 '파이썬 머신러닝 완벽 가이드' 책을 통해 자율적으로 학습하시고,
개념에 대한 질문을 토대로 본인의 답변을 작성해주세요.

야금야금 머신러닝의 모든 질문은 공통 교재로부터 출제됩니다.

답변을 작성하는 과정에서 책을 참고해도 좋고 구글링을 통해 알아보셔도 좋습니다.
다른 Basic 부원분들과 협동해서 풀어도 좋습니다.

다만 답변을 작성하면서 머신러닝 개념들을 본인의 것으로 꼭 만들어 주세요!

이름	(이 글을 지우고 이름을 검은 글씨로 작성해주세요)
학과	(이 글을 지우고 학과를 검은 글씨로 작성해주세요)

파일명은 (야금야금 머신러닝 2회차 Basic_홍길동)으로 제출해주세요!

1) 성능 평가 지표(Evaluation Metric)은 무엇에 따라 나뉘나요?

2) 분류의 성능 평가 지표에는 무엇이 있나요?

3) 정확도인 accuracy는 어떻게 정의되나요?

4) 정확도를 적용할 경우 단점은 무엇인가요?

5) 오차행렬(confusion matrix)이란 무엇인가요?

6) 오차행렬의 TP는 어떻게 정의되나요?

7) 사이킷런에서는 오차행렬을 구하기 위해 어떤 API를 제공하나요?

8) 오차행렬 내부에서 정확도는 어떻게 계산되나요?

9) 오차행렬 내부에서 정밀도는 어떻게 계산되나요?

10) 오차행렬 내부에서 재현율은 어떻게 계산되나요?

11) 재현율이 중요 지표인 경우는 어떤 경우인가요?

12) 정밀도가 중요 지표인 경우는 어떤 경우인가요?

13) 정밀도/재현율의 트레이드오프(Trade-off)란 무엇인가요?

14) 정밀도/재현율의 성능 수치를 상호보완하기 위해 사용하는 것은 무엇인가요?

15) 정밀도와 재현율을 통해 F1 스코어는 어떻게 계산되나요?

16) ROC 곡선 그래프에서 실질적인 성능 지표로 사용되는 것은 무엇인가요?

17) 분류(Classification)란 무엇인가요?

18) 분류에 기반한 머신러닝 알고리즘 3개를 말해주세요.

19) 정형 데이터의 분석에서 높은 예측 성능을 발휘하는 것은 무엇인가요?

20) 결정 트리에서 과적합이 되는 경우는 어떤 경우인가요?

21) DecisionTreeClassifier는 무엇을 기본으로 데이터 세트를 분할하나요?

22) 지니 계수를 간략히 설명해주세요.

23) 앙상블 학습의 유형에는 무엇이 있나요?

24) 보팅과 배깅의 다른 점은 무엇인가요?

25) 스택킹이란 무엇인가요?

26) 랜덤 포레스트의 예측 방식을 간단하게 설명해주세요.

27) 부트스트래핑(bootstrapping) 분할 방식이란 무엇인가요?

28) 부스팅 알고리즘이란 무엇인가요?

29) 경사 하강법(Gradient Descent)에 대해 간략히 설명해주세요.

30) GBM(Gradient Boost Machine)의 장단점을 간략하게 설명해주세요.

31) XGBoost의 장점 3가지를 말해주세요.

32) 파라미터 `early_stopping_rounds`의 역할은 무엇인가요?

33) 트리 분할에 있어 LightGBM이 가지는 일반 GBM 계열과의 차이는 무엇인가요?

34) 오버샘플링(oversampling)에 대해 간략하게 설명해주세요.

35) 오버샘플링을 사용하는 이유는 무엇인가요?

36) 사분위에서 IQR은 어떻게 정의되나요?

37) IQR을 통해 이상치 데이터를 검출하는 방식에 대해 간략히 설명해주세요.

38) 배깅 및 부스팅에 비해 스택킹이 가지는 가장 큰 차이는 무엇인가요?

39) 책의 279쪽 오른쪽 상단에 있는 그림에 대해 간략히 설명해주세요.