

## < CUI 4기 BASIC 트랙 야금야금 머신러닝 1회차 >

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 1회차 Basic\_홍길동)으로 제출해주세요!

1) 머신러닝은 어떤 알고리즘 기법을 통칭하나요?

2) 일반적으로 머신러닝은 어떻게 분류되나요?

3) 지도학습의 머신러닝에는 어떤 것들이 있나요?

4) 머신러닝의 가장 큰 단점은 무엇일까요?

5) 파이썬 패키지 중 NumPy는 무엇인가요?

6) 파이썬 패키지 중 시각화 패키지는 어떤 것이 있나요?

7) reshape()에 -1값을 인자로 적용하면 결과가 어떻게 되나요?

8) 넘파이 ndarray에서 인덱싱하는 방법에는 무엇이 있나요?

9) Pandas의 핵심 객체는 무엇인가요?

10) DataFrame에서 칼럼을 드롭할 때 어떤 메서드를 사용하나요?

11) DataFrame 메서드에서 inplace=True의 의미는 무엇인가요?

12) `iloc[]`과 `loc[]`의 차이는 무엇인가요?

13) 결손 데이터를 대체해야 하는 이유는 무엇인가요?

14) `apply` `lambda` 식으로 데이터를 가공할 때 장점은 무엇인가요?

15) Scikit-Learn의 특징 두 가지를 적어주세요.

16) 지도학습의 수행과정을 간략하게 적어주세요.

17) 학습용 데이터와 테스트용 데이터를 분리하는 이유는 무엇인가요?

18) 학습용 데이터와 테스트용 데이터는 어떤 API로 구분되나요?

19) 책에서 언급한 붓꽃 데이터 세트의 프로세스를 4단계로 적어주세요.

20) 모듈 `sklearn.model_selection`에 대해 간략히 설명해주세요.

21) 모듈 `sklearn.preprocessing`에 대해 간략히 설명해주세요.

22) `train_test_split`의 파라미터 `test_size`에 대해 간략히 설명해주세요.

23) 교차 검증을 수행하는 이유는 무엇인가요?

24) K폴드 교차 검증이란 무엇인가요?

25) Stratified K폴드란 무엇인가요?

26) 최적 하이퍼 파라미터 튜닝에는 사이킷런의 어떤 API가 사용되나요?

27) 피쳐 값 중 Null 값이 얼마 되지 않거나 대부분이라면 각각 어떻게 해야 하나요?

28) 레이블 인코딩이란 무엇인가요?

29) 원-핫 인코딩과 레이블 인코딩의 차이는 무엇인가요?

30) 피쳐 스케일링(feature scaling)이란 무엇인가요?

31) fit(), transform(), fit\_transform() 이용 시 유의할 점은 뭔가요?

32) 책의 타이타닉 생존자 예측 중 전처리 과정에서 무엇을 수행했나요?