

18 -1 중간 데이터사이언스 개론 이기용 교수님

총 24문제였습니다 ! 그리고 22문제 기입하였습니다.

어느 한 군데에서 많이 나온 것 없이 모든 파트에서 세세하게 나왔습니다!  $\pi$

1. Data scientist, Data engineer 가 하는 일 구분

(예시 주어지고 데이터 사이언티스트 인지 데이터 엔지니어 인지 씀)

2. 모델 구분

(예시 주어지고 각각 예시들이 clustering, classifier 할지 등등 씀)

3. 데이터 사이언티스트의 자질2가지 쓰기

4. supervised, unsupervised 모델 구분하기

(예시 주어지고 각각 모델이 supervised, unsupervised 모델인지 씀)

5. Information Gain 구하기 문제 (달린 문제까지 총 3개)

6. 트리 주어지고 Space partition 그래프 그리기

7. 각각 정확도 다르게 주어진 두개 모델 주워지고 가장 쓰기 알맞은 모델 고르고 이유쓰기

8. 1)트레이닝 데이터로만 검사한 모델 2) 테스트 데이터로 한번더 검사한 모델

### 3) Cross- Validation 한 모델

위의 세가지 모델을 믿을 만한 모델 순서대로 쓰기

9. Data-preparation 단계에서 하는 일 2가지 쓰기

10. 두 개의 variable 주어지고 Logistic regression 이용해서 두 variable이 같은

클래스에 속하는지 여부와 그 이유 쓰기

11. variable 주어지고 리니어 클래사파이어 그래프 그리기

12. 표 주어지고  $k=2$  인 2-NN 으로 거리계산하여서 Targer value 값이 무엇인지 구하기

13. 리니어 와 트리 에서 오버피팅 regulation 하는 방법 2개씩 쓰기

14. 피팅 그래프에서 오버피팅 어떻게 찾는지 방법 쓰기

15. 러닝그래프 그리기

16. (nested cross validation 주어지고) 트리가 총 몇개 생성되는지 개수 쓰기

17. 리니어 리그래션에서 최소로 해야하는 조건 계산식 쓰기

18. SVM 그래프 그리고 Margin 표시하기

19. (SVM그래프 2개 주어지고) SVM 에서 hinge 기울기가 더 큰 그래프 원지 고르기

20. k NN 모델을 실제로 적용하면 안좋은 이유 쓰기

21. (예시 주어지고) K-NN 에서  $k = 1$  일때와  $k = n$  ( $n =$  전체) 일때 확률 계산하기

22. 거리 구하기문제 (총 3개 맨하탄, jaccard, cosine 거리)

23.

24.