

평가방법: 필답형 - 머신러닝_딥러닝

| | | | |
|------|------------------------|-----|--|
| 평가일시 | 2019년 10월 14일 | | |
| 과정명 | SBA 빅데이터 사이언스 실무자 양성과정 | | |
| 교과목 | 머신러닝_딥러닝I | 수강생 | |
| 단원명 | 머신러닝_딥러닝I | 확 인 | |

| 구분 (능력단위요소) | 문항 | 점수 |
|----------------|--|----------|
| | <p>(문항 1~ 문항 10) 8점, 문항 (11), (12) 10점</p> <p>다음이 설명하는 라이브러리는 무엇인가요?</p> <p>(1) 오픈 소스 이며, 매우 인기 높고 독보적인 파이썬 머신러닝 라이브러리입니다. ()</p> <p>(2) 파이썬으로 과학 계산을 하기 위한 꼭 필요한 패키지로서, 다차원 배열을 위한 기능을 가지고 있으며, 선형 대수 연산 기능을 가지고 있습니다. ()</p> <p>(보기) pandas, matplotlib, numpy, scikit-learn</p> <p>(정답) (A) _____ , (B) _____</p> <p>(3) 데이터를 훈련 데이터와 테스트 데이터 셋으로 나누기 위해 사용하는 함수는 무엇인가요? (정답) _____</p> <p>(4) O, X 맞으면 O, 틀리면 X 로 작성해 주세요. (A) 지도학습에는 두 가지 종류 분류와 회귀가 있다.() (B) k-mean 모델은 대표적인 지도학습 방법 중의 하나로 분류에 사용된다. () (C) 회귀는 정해져 있는 연속하지 않은 몇 가지 값을 예측하기 위한 방법이다. () (D) knn 모델은 범주형을 타깃 값을 예측할 때, 거리상 가장 가까운 k개의 데이터가 중에 많이 가지는 레이블의 값을 예측하게 된다.()</p> | 100/100점 |

(정답) (A) _____

(정답) (B) _____

(정답) (C) _____

(정답) (D) _____

(5) 아래 내용이 설명하는 것은 무엇인가요?

아주 복잡한 모델이다. 훈련 데이터 셋에 아주 잘 훈련된 현상을 말한다. 훈련 세트에만 정확하고 다른 일반 데이터 셋에는 정확하지 않을 가능성이 높다.

(A) 과대적합(Overfitting)

(B) 과소적합(underfittig)

(C) 일반화(generalization)

(D) 최적화(optimization)

(정답) _____

(6) 모델을 생성하고 이를 예측한 이후에 실제 값과 예측 값을 이용하여 평가하기 위한 대표적인 평가 지표로 회귀 모델에 적당하지 않은 것은 무엇인가요?

(A) MAE

(B) AUC

(C) MSE

(D) RMLSE

(정답) _____

(7) 회귀 모델은 특성이 많아질수록 선형모델의 성능이 높아져 과대적합이 될 가능성이 높아진다. 우리는 이를 위해 모델에 제한을 두는 방법을 사용한다. 두 가지 모델에 제한을 두는 방법을 적어보자.

_____ (A) _____ : 계수들을 0에 가깝게 제한

_____ (B) _____ : 계수를 어떤 것은 0으로 만들어 버린다.

(정답) _____ (A) _____

(정답) _____ (B) _____

(8) 아래 설명하는 것은 무엇인지 보기에서 골라주세요.

이 모델은 **분류**와 **회귀**문제에 널리 사용하는 모델이다.

스무고개 놀이의 질문과 비슷하다. 맨 위의 노드를 Root Node라고 하며, 맨 마지막 노드를 Leaf Node라고 한다. 이 모델은 발전하여 앙상블 모델에 사용되는 기본 모델로 사용된다.

- (A) 의사결정트리
- (B) 선형회귀
- (C) 랜덤포레스트
- (D) 그래디언트 부스팅

(정답) _____

(9) 아래 설명하는 것에 가장 가까운 모델은 무엇인지 보기에서 골라주세요.

이 모델은 **분류**와 **회귀**문제에 널리 사용하는 모델이다.

앙상블 기법중의 하나로서 여러 개의 트리를 만들어준다.

각각의 트리는 독립적으로 만들어진다.

- (A) 의사결정트리
- (B) 선형회귀
- (C) 랜덤포레스트
- (D) 그래디언트 부스팅

(정답) _____

(10) 빈칸을 채워 주세요.

- (A) cancer.target를 기준으로 동일한 기준으로 데이터 셋을 나눠주는 옵션
- (B) 나무의 깊이를 지정해 주는 옵션

```
def testTreeModel(TestSize=0.3, treedepth=3):  
    cancer = load_breast_cancer()  
    X_train, X_test, y_train, y_test = train_test_split(all_X, all_Y,  
        ____(A)____ = cancer.target,  
        test_size = TestSize,  
        random_state=77)  
    tree = DecisionTreeClassifier(____(B)____ =treedepth, random_state=0)  
    tree.fit(X_train, y_train)  
    print("훈련 세트 정확도 : {:.3f }".format(tree.score(X_train, y_train)))  
    print("테스트 세트 정확도 : {:.3f }".format(tree.score(X_test, y_test)))
```

(정답) (A)_____

(정답) (B)_____

(11) KNN모델을 생성하고 이를 예측하기 위한 소스 코드이다. 아래 코드를 완

성해 주세요. (10점)

```
model = KNeighborsClassifier(_____=3)
_____(B)_____ # 입력 X_train, 답(출력) y_train
pred = model.predict(X_test)
pred
```

(정답) (A)_____

(정답) (B)_____

(12) 랜덤 포레스트 모델은 앙상블 기법으로서 여러 머신러닝 모델을 연결하여 더 강력한 모델을 만드는 기법이다. 랜덤 포레스트는 각각의 트리 성격이 달라지도록 트리 생성 시에 2가지 방법을 이용하여 트리를 랜덤하게 만들어 주게 된다. (10점)

(A) _____

(B) _____

(정답) (A)_____

(정답) (B)_____

두 가지 방법을 적어 주세요.

총점