

Deep Learning 중간고사 (20 점)

1. O, X 문제. 각 1점식, 단 틀리면 -1 감점

1-1. 인공지능은 과거의 패턴을 학습하여 새로운 문제를 해결할 수 있는 능력을 포함한다. (O)

1-2. 지도학습은 레이블이 있는 데이터를 기반으로 학습하지만, 훈련 과정에서 사용된 데이터에만 잘 맞도록 설계되므로 테스트 데이터에 대한 일반화 성능은 고려하지 않는다. (X)

1-3. 비지도학습에서는 입력에 대한 피드백이 존재하지 않기 때문에, 모델이 찾는 군집은 반드시 사람이 기대하는 분류 결과와 일치한다고 보장할 수 없다. (O)

1-4. 강아지와 고양이를 구분하는 문제에서, 정답을 알려주지 않고 학습 알고리즘이 자동으로 유사한 그룹을 나누도록 한다면 이는 지도학습이다. (X)

1-5. 강화학습과 비지도학습은 둘 다 외부로부터 정답을 직접 제공받지 않으며 데이터 또는 환경으로부터 패턴이나 최적의 정책을 스스로 탐색한다는 공통점이 있다. (O)

2. 다음 중 머신러닝 모델 개발 과정의 순서와 역할을 가장 부정확하게 설명한 것은? (2점)

A. 데이터 전처리는 결측치 보완, 이상치 제거, 형식 통일 등을 통해 데이터 품질을 높이는 과정이며, 모델 학습과 평가 이전에 수행된다.

B. 모델 학습에서는 정제된 데이터를 사용하여 패턴을 학습하며 이때 테스트 데이터를 활용해 과적합을 방지하고 하이퍼파라미터를 조정한다.

C. 예측은 학습된 모델을 사용해 새로운 데이터에 대한 결과를 생성하는 단계로, 학습에 사용되지 않은 데이터를 입력으로 사용한다.

D. 모델 평가는 테스트셋으로 수행하며, 정확도, 정밀도, 재현율, F1-score 등 다양한 지표로 성능을 분석한다.

3. 10명의 유전자 데이터를 활용하여 머신러닝 기반의 유전자 예측 시스템을 개발하고자 한다. 이때, leave-one-out 교차 검증을 3번 반복한다고 가정할 때, 한 명의 샘플이 테스트에 포함될 수 있는 최대 횟수는 몇 번인가?(2점)

3번

4. 다음 중 분류 알고리즘보다 회귀 알고리즘을 적용하는 것이 더 적절한 상황은?(1점)

A. 환자의 유전자 패턴을 바탕으로 암의 존재 여부를 예측하는 모델을 구축하려고 한다.

B. 치매 환자의 뇌 영상 데이터를 바탕으로 병의 중증도를 "경증, 중등도, 중증" 3단계로 예측하려고 한다.

C. 심전도 신호를 바탕으로 환자의 심박수를 bpm 단위의 실수값으로 예측하려고 한다.

D. 환자의 과거 병력과 나이를 기반으로 미래 1년 내 재입원 여부를 예측하고자 한다.

학과 : 학번 : 이름

4. 한 소상공인이 유튜브 광고비, 오프라인 행사비 등을 반영한 다양한 변수들을 이용해 매달 판매량을 예측하고자 한다. 다음 중 가장 적절한 모델은? (1점)

A. 선형회귀: 출력이 수치형이지만, sigmoid 출력층을 통해 판매량 예측을 안정화시킬 수 있다.

B. K-Nearest Neighbors: 비선형성과 지역적 데이터 밀도까지 고려할 수 있어 작은 데이터셋에 유리하다.

C. 선형회귀: 입력값과 출력값 간의 선형적 관계가 가정되며, 해석 가능성이 높고 과적합 방지를 위한 정규화가 가능하다.

D. K-Nearest Neighbors: 계절성과 비용 간의 상호작용을 비선형적으로 반영할 수 있으며 과적합에도 강하다.

6. 선형 회귀는 입력 데이터를 가장 잘 설명하는 (기울기)과 (절편)을 찾는 문제이다.(1점)

7. 실제 데이터와 예측 값 사이의 차이를 나타내는 것을 (손실함수)이라 하며, 선형회귀에서 가장 많이 사용되는 방법은 (MSE)이다. (1점)

8. 과적합(overfitting)이란 (훈련데이터)에서는 뛰어난 성능을 보이지만, (테스트데이터)에서는 일반화 성능이 낮은 현상을 말한다. (1점)

9. 다음 중 경사 하강법(gradient descent)에 대한 설명 중 가장 틀린 것은? (1점)

A. 손실 함수 f(W)의 도함수인 기울기를 기준으로, 기울기가 양수이면 가중치를 감소시키는 방향으로 이동한다.

B. 경사 하강법은 손실 함수가 미분 가능한 경우에만 적용 가능하며, 이때 기울기는 하나의 스칼라 값으로 계산된다.

C. 학습률이 지나치게 크면 최솟값을 지나쳐 발산할 수 있으므로 적절한 값의 선택이 중요하다.

D. 기울기가 0이 되는 지점에서는 이동이 멈추므로 해당 지점이 손실 함수의 전역 최솟값이라는 보장이 없다.

10. 퍼셉트론 학습 규칙의 수학적 또는 개념적 설명으로 틀린 것은? (1점)

A. 퍼셉트론은 출력이 정답과 같을 경우 가중치를 변경하지 않고, 오답일 경우에만 입력과 학습률을 고려하여 가중치를 조정한다.

B. 퍼셉트론의 업데이트 식은 $w_i \leftarrow w_i + \eta(y - y^{\wedge}) \cdot x_i$ 로 주어지며, 이때 y는 실제 정답, y^{\wedge} 는 모델의 출력이다.

C. 퍼셉트론은 주어진 입력 데이터를 분류하기 미분한 값을 통해 가중치를 업데이트한다.

D. 퍼셉트론은 선형 분리 가능한 데이터에 대해서는 수렴을 보장하지만, 선형 분리 불가능한 경우에는 업데이트가 무한 반복될 수 있다.

11. 퍼셉트론의 한계를 보완하기 위해 등장한 (MLP)는 (은닉층)을 포함하며, 각 뉴런에는 (비선형) 활성화 함수가 적용되어 복잡한 비선형 함수도 근사할 수 있다. (1점)

12. 다음 중 역전파 알고리즘의 원리에 대한 설명 중 가장 틀린 것은? (1점)
- A. 역전파는 출력층에서 계산된 손실 함수를 각 층을 거슬러 올라가며 편미분함으로써, 각 가중치가 손실에 끼친 영향도를 계산하고 이에 따라 가중치를 갱신하는 알고리즘이다.
 - B. 역전파 알고리즘은 각 층에서 활성화 함수의 도함수 값이 0이 아닌 한, 손실 함수의 기울기가 중간 층까지 정상적으로 전달되어 학습이 이루어진다.
 - C. 역전파는 손실 함수 E 를 각 가중치에 대해 편미분한 $\frac{\partial E}{\partial w}$ 를 계산하여, $w(t+1)=w(t)-\eta\frac{\partial E}{\partial w}$ 형태로 업데이트한다.
 - D. 역전파 알고리즘은 미분 가능한 손실 함수와 연속적인 활성화 함수가 사용되면, 경사 하강법을 통해 항상 최적의 파라미터를 찾을 수 있다.

13. 자율주행 자동차의 주행 중 수집되는 센서 데이터를 바탕으로, 현재 상황이 긴급 제동이 필요한 상황인지 아닌지를 판단하는 모델을 설계 하시오.(2점)

- 단, 아래 조건들을 만족할 것.
- 조건 1. 센서 데이터는 연속형 특성 80개로 구성되어 있으며, 샘플은 40명이다.
 - 조건 2. 인공지능 모델은 수업시간에 다룬 알고리즘 중 선택할 할 것
 - 조건 3. 성능 검증 시 과적합을 줄이고 일반화 가능성을 높일 수 있는 방법을 사용 것
 - 조건 4. 모델의 성능을 파악할 수 있는 방법을 1건 이상 제공할 것