

1. 인공지능의 소개 (중간고사 대비 단답형 문제)

다음 빈칸에 알맞은 답을 쓰거나, 질문에 대해 간결하게 답하시오.

1. 인공지능은 인간의 _____ 행동을 모방하는 컴퓨터 시스템입니다.
2. 특정 문제 해결에 유용한 제한된 지능을 가지며, 지능이 있는 것처럼 프로그래밍되는 인공지능을 _____ 인공지능이라고 합니다.
3. 사람처럼 사고하고 인간 두뇌의 대체 가능한 수준으로 다목적 과제 수행이 가능한 범용적 지능을 _____ 인공지능이라고 합니다.
4. 앨런 튜링이 제안한 것으로, 질의자가 대화하는 상대가 사람인지 컴퓨터인지 구분하지 못할 경우 인공지능이라고 판단하는 기준이 되는 시험을 _____라고 합니다.
5. 1956 년에 개최되어 'AI'라는 용어가 공식화되고 인공지능의 아버지들(The Founding Fathers of AI)이 모였던 역사적인 회의의 이름은 무엇입니까? _____.
6. 1980 년대에 등장하여 상업적으로 성공한 최초의 _____ 중 하나로 미니컴퓨터 시스템 구성을 위해 사용되었던 시스템은 R1/XCON 입니다.
7. 2016 년 이세돌 9 단과의 바둑 대결에서 승리하여 딥러닝과 AI 의 대중화를 상징했던 Google DeepMind 의 AI 프로그램 이름은 무엇입니까? _____.
8. 이미지, 비디오 데이터를 처리하여 객체 검출, 얼굴 인식, 자율주행차의 환경 인식 등에 사용되는 인공지능의 주요 분야는 무엇입니까? _____.
9. 이미지에서 특정 객체를 식별하고 그 위치를 찾는 작업을 _____이라고 하며, 이는 자율주행차의 환경 인식 등에 사용됩니다.
10. 금융 분야에서 고객의 거래 패턴을 모니터링하고 이상 행동을 감지하여 신용 카드 사기 및 금융 범죄를 예방하는 인공지능의 활용 기술은 무엇입니까? _____.

1. 인공지능의 소개 (중간고사 대비 단답형 정답)

1. 지능적인
2. 약한 (Weak)
3. 강한 (Strong)
4. 튜링 테스트 (Turing Test)
5. 다트머스 회의 (Dartmouth Conference)
6. 전문가 시스템
7. 알파고 (AlphaGo)
8. 컴퓨터 비전 (Computer Vision)
9. 객체 검출 (Object Detection)
10. 금융 사기 탐지

2. 머신 러닝의 기초 (중간고사 대비 단답형 문제)

다음 빈칸에 알맞은 답을 쓰거나, 질문에 대해 간결하게 답하시오.

1. 머신 러닝은 명시적으로 _____ 하지 않아도 컴퓨터가 데이터를 통해 학습하여 패턴을 인식하고 예측이나 결정을 내리는 기술입니다.
2. 머신 러닝의 학습 단계는 데이터 수집, 피쳐 정의, 가설 정의, _____ 정의, 학습의 순서로 진행됩니다.
3. 머신 러닝 모델이 학습하고 예측을 수행하는 데 필요한 입력이며, 좋은 품질이 모델을 정확하게 학습시키는 데 중요한 것은 무엇입니까? (정형 데이터와 비정형 데이터가 있음) _____
4. 머신 러닝에서 예측된 출력과 실제 출력 간의 차이를 최소화하는 방향으로 파라미터를 조정하여 모델이 데이터의 패턴을 학습하고 예측 능력을 갖추도록 하는 과정을 무엇이라고 합니까? _____
5. 문제와 _____을(를) 모두 알려주고 공부시키는 방법이며, 주로 예측(Regression)과 분류(Classification) 작업에 사용되는 머신 러닝의 종류는 무엇입니까? _____ 학습
6. 답을 가르쳐주지 않고 데이터의 패턴이나 구조를 스스로 파악하도록 공부시키는 방법이며, 연관 규칙, _____ (Clustering), 차원 축소 등에 사용되는 머신 러닝의 종류는 무엇입니까? _____ 학습
7. 머신 러닝에서 _____을(를) 통해 상은 최대화하고 벌은 최소화하는 방향으로 행위를 강화하는 학습 방식은 무엇입니까? _____ 학습
8. 지도 학습에서 결과가 연속적인 숫자 값(예: 주택 가격, 택시비)을 예측하는 문제를 _____ 이라고 합니다.
9. 지도 학습에서 입력 데이터를 두 개 이상의 클래스(범주) 중 하나로 나누는 문제(예: 스팸/비 스팸, 강아지/고양이)를 _____ 이라고 합니다.
10. 모델이 훈련 데이터에 너무 완벽하게 맞춰져서(too good to be true) 새로운 데이터나 테스트 데이터에 대한 예측 성능이 떨어지는 현상을 무엇이라고 합니까?

2. 머신 러닝의 기초 (중간고사 대비 단답형 정답)

1. 프로그래밍
2. 비용 코스트 함수
3. 데이터
4. 학습 (Training)
5. 정답, 지도 (Supervised)
6. 군집, 비지도 (Unsupervised)
7. 보상 (Reward), 강화 (Reinforcement)
8. 회귀 (Regression)
9. 분류 (Classification)
10. 과적합 (Overfitting)

3. 데이터 준비와 전처리의 기초 (중간고사 대비 단답형 문제)

다음 빈칸에 알맞은 답을 쓰거나, 질문에 대해 간결하게 답하십시오.

1. 모델의 정확도 향상, 일반화 능력 향상, 과적합 방지 등을 위해 반드시 필요한 데이터 준비 과정을 통틀어 _____라고 합니다.
2. 특정 웹 사이트에서 필요한 데이터를 수집하는 것으로, 웹 크롤링보다 좁은 범위의 데이터 수집 방식을 _____이라고 합니다.
3. 데이터 수집을 자동화하고 구조화된 방식으로 데이터를 획득하기 위해 사용하는, 일반적으로 HTTP 요청을 통해 데이터를 제공받는 인터페이스를 _____라고 합니다.
4. 데이터 분석과 모델 훈련 과정에서 문제(정확성 저하 등)를 유발하며, 누락된 값을 의미하는 용어는 _____입니다.
5. 결측값 처리 방법 중, 가장 많이 나오는 값이나 평균값 등 통계로 계산한 대푯값으로 결측치를 대신하는 방법을 _____라고 합니다.
6. 머신 러닝 모델의 성능을 향상시키기 위해 데이터의 특성을 만들거나 변형하는 과정이며, 중요한 특성을 선택하고 변환하여 모델이 더 효율적으로 학습하게 하는 과정을 _____이라고 합니다.
7. 데이터 값을 0 과 1 사이의 일정한 범위로 조정하는 스케일링 방법으로, 스케일링 시 최대, 최소값이 사용되는 것은 _____입니다.
8. 데이터의 평균을 0, 표준편차를 1로 조정하며 특정 범위로 제한되지 않는 스케일링 방법은 _____입니다.
9. 색상(빨강, 파랑)이나 도시(뉴욕, 런던)처럼 범주 간에 순서나 계급이 없는 범주형 데이터는 _____ 데이터라고 합니다.
10. 범주형 데이터를 처리하는 방법 중, 각 범주를 독립적인 이진(0 또는 1) 특성으로 표현하여 벡터로 변환하는 인코딩 방식은 _____ 인코딩입니다.

3. 데이터 준비와 전처리의 기초 (중간고사 대비 단답형 정답)

1. 전처리 (또는 데이터 전처리)
2. 웹 스크래핑 (Web Scraping)
3. API (Application Programming Interface)
4. 결측값 (Missing Value)
5. 대체
6. 특성 공학 (Feature Engineering)
7. 정규화 (Normalization)
8. 표준화 (Standardization)
9. 명목형 (Nominal)
10. 원-핫 (One-Hot)

4. 선형 회귀와 로지스틱 회귀 (중간고사 대비 단답형 문제)

다음 빈칸에 알맞은 답을 쓰거나, 질문에 대해 간결하게 답하시오.

1. 두 변수 사이의 관계를 분석하고 이를 _____으로 표현하는 통계 기법으로, 종속 변수(y)가 연속적인 숫자 값일 때 예측에 사용되는 모델은 무엇입니까?
2. 선형 회귀 방정식 $y=mx+b$ 에서, 독립 변수 x가 1단위 증가할 때 종속 변수 y의 증가량을 나타내는 파라미터 m을 무엇이라고 부릅니까?
3. 선형 회귀의 주된 목적은 주어진 데이터를 가장 잘 설명하는 _____을(를) 찾는 것입니다.
4. 독립 변수와 종속 변수 간의 관계를 모델링하여 독립 변수의 값이 주어졌을 때 종속 변수를 _____하는 것이 선형 회귀의 목적입니다.
5. 로지스틱 회귀는 종속 변수가 _____ (예: 예/아니오, 0/1)일 때 사용되며, 출력값이 0과 1 사이의 _____로 변환됩니다.
6. 로지스틱 회귀에서 선형 회귀 결과를 0과 1 사이의 확률로 제한하기 위해 사용하는 'S'자 모양의 곡선 함수를 무엇이라고 합니까?
7. 주어진 입력 변수들이 특정 사건 발생에 영향을 미치는지를 모델링하고, _____이(가) 발생할 확률을 예측하는 것이 로지스틱 회귀의 주된 목적입니다.
8. 로지스틱 회귀의 종류 중, 결과 변수가 두 개의 범주일 때 사용되는 회귀는 _____ 로지스틱 회귀입니다.
9. 로지스틱 회귀의 종류 중, 결과 변수가 세 개 이상의 범주일 때 사용되는 회귀는 _____ 로지스틱 회귀입니다.
10. 로지스틱 회귀 모델의 평가 지표로 사용되는 행렬로, 실제 결과와 예측 결과를 비교하여 모델의 정확도, 정밀도 등을 계산하는 데 사용되는 것은 무엇입니까?
_____ 행렬

4. 선형 회귀와 로지스틱 회귀 (중간고사 대비 단답형 정답)

1. 직선 (선형 방정식), 선형 회귀 (Linear Regression)
2. 기울기 (슬로프)
3. 직선
4. 예측
5. 범주형, 확률
6. 로지스틱 함수 (또는 시그모이드 함수)
7. 해당 사건 (또는 특정 사건)
8. 이진 (Binary)
9. 다항 (Multinomial)
10. 혼동 (Confusion)

5. 분류 알고리즘 (중간고사 대비 단답형 문제)

다음 빈칸에 알맞은 답을 쓰거나, 질문에 대해 간결하게 답하시오.

1. 분류 문제는 주어진 입력 데이터를 사전에 정의된 여러 범주 또는 클래스로 구분하는 _____ 학습 문제 유형 중 하나입니다.
2. 분류 문제에서 입력 데이터의 특성을 바탕으로 해당 데이터가 어느 클래스에 속할지 예측하는 대상인 출력 범주를 _____라고 합니다.
3. 결과를 '암 진단: 긍정 vs 부정'과 같이 두 개의 클래스로 구분하는 분류 문제를 _____ 분류라고 합니다.
4. 가장 간단한 분류 알고리즘으로, 새로운 데이터가 주변의 가장 가까운 K 개의 데이터 중 다수결의 원칙을 따라 속할 그룹을 판단하는 알고리즘은 무엇입니까? _____
5. K -최근점 이웃(KNN) 알고리즘에서 데이터 간의 거리를 측정할 때 주로 사용되는 거리 측정 방식은 무엇입니까? _____ 거리
6. 서포트 벡터 머신(SVM)에서 두 클래스를 분리하는 기준으로 정의되며, 2 차원에서는 직선, 고차원에서는 초평면이 되는 경계를 _____라고 합니다.
7. SVM 모델이 최적의 결정 경계를 찾기 위해, 결정 경계와 가장 가까이 있는 데이터 포인트(서포트 벡터) 사이의 거리를 _____하고 최대화하는 원리를 사용합니다.
8. 변수들로 기준을 만들고 이를 통해 샘플을 분류하는 나무 모양의 구조를 가지며, 해석력이 높다는 장점이 있는 분류 모형은 무엇입니까? _____
9. 여러 개의 의사 결정 나무를 학습시켜 그 결과를 모아 다수결 투표(Majority Voting)로 최종 결과를 도출하는 앙상블(Ensemble) 기법은 무엇입니까? _____
10. 분류 모델 평가 지표 중, 실제값과 예측값을 비교하여 TP, FP, FN, TN 을 한눈에 확인할 수 있게 정리한 표를 _____ 행렬이라고 합니다.

5. 분류 알고리즘 (중간고사 대비 단답형 정답)

1. 지도
2. 라벨 (Label)
3. 이진 (Binary)
4. K-최근점 이웃 (K-Nearest Neighbors, KNN)
5. 유클리드 (Euclidean)
6. 결정 경계 (Decision Boundary)
7. 마진 (Margin)
8. 의사 결정 나무 (Decision Tree)
9. 랜덤 포레스트 (Random Forest)
10. 혼동 (Confusion)

6. 군집화 알고리즘 (중간고사 대비 단답형 문제)

다음 빈칸에 알맞은 답을 쓰거나, 질문에 대해 간결하게 답하시오.

1. 군집화(Clustering)는 입력 데이터를 사전에 정의된 라벨 없이 스스로 그룹화하는 _____ 학습의 한 종류입니다.
2. K-means 알고리즘에서 각 클러스터를 대표하며, 반복 계산을 통해 위치를 업데이트하는 클러스터의 중심점을 무엇이라고 부릅니까? _____
3. K-means 알고리즘에서 데이터 포인트를 가장 가까운 중심점에 할당하기 위해 거리를 측정할 때 사용하는 방식은 _____ 거리입니다.
4. K-means 알고리즘에서 군집의 개수(K 값)를 변화시키면서 각 군집 내 거리의 합(WSS)이 급격히 줄어드는 지점을 K로 선택하는 기법을 _____ 기법이라고 합니다.
5. 데이터 포인트들을 계층적으로 군집화하며, 결과를 트리 구조(Dendrogram)로 시각화하여 데이터 간의 계층적 관계를 보여주는 알고리즘은 무엇입니까? _____.
6. 계층적 군집화에서, 각 데이터 포인트를 개별 클러스터로 지정한 후 가장 가까운 두 클러스터를 반복적으로 병합하여 하나의 클러스터로 묶어 나가는 상향식(Bottom-Up) 방법을 _____ 방법이라고 합니다.
7. 두 클러스터에서 가장 가까운 두 데이터 포인트 사이의 거리를 클러스터 간 거리 기준으로 삼는 계층적 군집화의 거리 계산 방법은 무엇입니까? _____.
8. 군집의 모양과 데이터의 밀도에 기반하여 클러스터를 형성하며, 미리 지정된 밀도 기준을 만족하는 데이터 포인트들을 하나의 군집으로 묶는 알고리즘은 무엇입니까? _____.
9. DBSCAN 알고리즘의 장점 중 하나로, K-means가 잘 찾지 못하는 _____ 클러스터(예: 원형이 아닌 복잡한 모양)를 잘 찾고 노이즈 및 _____에 강하다는 특징이 있습니다.
10. K-means, 계층적 군집화, DBSCAN을 포함한 군집화 알고리즘들이 비슷한 성향을 가진 고객 그룹을 식별하여 타겟 마케팅 등에 활용되는 가장 대표적인 활용 사례는 무엇입니까? _____.

6. 군집화 알고리즘 (중간고사 대비 단답형 정답)

1. 비지도 (Unsupervised)
2. 중심점 (Centroid)
3. 유클리드 (Euclidean)
4. 엘보우 (Elbow)
5. 계층적 군집화 (Hierarchical Clustering)
6. 병합적 (Agglomerative)
7. 최단 거리 (Single Linkage)
8. DBSCAN
9. 비구형, 이상치 (Outlier)
10. 고객 세분화 (Customer Segmentation)