

1. 인공지능 (Artificial Intelligence, AI)의 기초

| 용어 | 정의 및 핵심 내용 |
|--------------------------|--|
| 인공지능 (AI) | 인간의 지능적인 행동을 모방하는 컴퓨터 시스템 기술 사고나 학습 등 인간의 지적능력을 컴퓨터를 통해 구현하는 기술 |
| 약한 인공지능 (Weak AI) | 특정 문제의 해결에 국한되며, 인간 두뇌의 일부 기능 모사하여 제한된 지능을 가짐. |
| 강한 인공지능 (Strong AI) | 사람처럼 사고하며, 인간 두뇌의 대체 가능한 수준으로 다목적 과제 수행이 가능한 범용적 지능 |
| 튜링 테스트 (Turing Test) | 질의자가 응답자가 사람인지 컴퓨터인지 구분하지 못할 경우 인공지능으로 판단하는 시험 |
| 자연어 처리 (NLP) | 인간이 사용하는 언어를 기계가 이해하고 처리하여 정보를 추출하거나 생성하는 기술 |
| 컴퓨터 비전 (Computer Vision) | 이미지, 비디오 데이터를 처리하여 객체 검출, 얼굴 인식, 이미지 분류 등 환경 인식을 수행하는 기술 |
| 로봇 공학 (Robotics) | 로봇 시스템을 개발하여 환경에서 작업을 수행하고 인간과 상호 작용할 수 있도록 하는 분야 (HW + SW) |

2. 머신러닝 (Machine Learning, ML)의 기초

| 용어 | 정의 및 핵심 내용 |
|--|--|
| 머신러닝 (ML) | 명시적으로 프로그래밍하지 않아도 컴퓨터가 데이터를 통해 학습하여 패턴을 인식하고 이를 기반으로 예측/결정을 내리는 기술. 실수 또는 복소 행렬을 세 개의 다른 행렬로 분해하는 강력한 고 널리 사용되는 행렬 분해 기법이다. |
| 지도 학습 (Supervised Learning) | 문제와 정답(레이블)을 모두 알려주는 학습 방법. 분류(Classification)와 회귀(Regression) 문제 사용. |
| 비지도 학습 (Unsupervised Learning) | 정답(레이블)을 가르쳐주지 않고 공구하는 방법. 군집화(Clustering), 차원 축소(Dimension Reduction)에 사용. |
| 강화 학습 (Reinforcement Learning) | 보상을 최대화하고 벌을 최소화하는 방향을 행위를 강화하는 학습 |
| 분류 (Classification) | 주어진 입력 데이터를 사전에 정의된 여러 범주(클래스)로 구분하는 지도 학습 문제. (예: 스팸, 암/정상) |
| 회귀 (Regression) | 두 변수 사이의 관계를 분석하고 연속적인 실수 값을 예측하는 지도 학습 문제. (예: 주택 가격, 판매량) |
| 오버피팅 (Overfitting) | 모델이 특정 데이터셋에 과적합되어 새로운 데이터에 대해 일반화 능력이 떨어지는 현상 |
| 특이값 분해 (Singular Value Decomposition, SVD) | 실수 또는 복소 행렬을 세 개의 다른 행렬로 분해하는 강력하고 널리 사용되는 행렬 분해 기법이다. |

3. 데이터 준비 및 전처리 (Data Preparation & Preprocessing)

| 용어 | 정의 및 핵심 내용 |
|--|---|
| 데이터 정제 | 데이터의 품질을 보장하기 위해 결측값, 이상값, 중복 데이터 등을 처리하는 과정. |
| 결측값 (Missing Value) | 데이터셋에서 누락된 값. 모델 정확성 저하 및 문제 유발. 삭제, 대체(평균/최빈값 등), 예측 등으로 처리. |
| 이상값 (Outlier) | 데이터의 일반적인 패턴에서 크게 벗어난 값. 모델이 잘못된 학습을 진행하게 하므로 제거, 대체, 변환 등으로 처리. |
| 특성 공학 (Feature Engineering) | 머신러닝 모델의 성능을 향상시키기 위해 데이터의 특성을 만들거나 변형하는 과정. |
| 표준화 (Standardization) | 데이터의 평균을 0, 표준편차를 1로 조정하는 특성 스케일링. (정규분포에 유용) |
| 정규화 (Normalization) | 데이터 값을 주로 0과 1 사이의 일정한 범위로 조정하는 스케일링. (최소-최대 정규화) |
| 범주형 데이터 | 값이 명확한 범주 또는 그룹으로 나누어지는 데이터. (예: 색상, 성별, 교육 수준) |
| 원-핫 인코딩 (One-Hot Encoding) | 범주형 데이터를 이진 벡터 형태로 변환하는 기법. (예: [1, 0, 0, 0, 0, 0, 0]) |
| 레이블 인코딩 (Label Encoding) | 범주형 데이터의 각 범주를 고유한 숫자로 변환하는 기법. (예: Apple → 1, Chicken → 2) |
| 토큰화 (Tokenization) | 텍스트를 분석의 최소 단위인 단어나 문장으로 나누는 작업. |
| SIFT (Scale-Invariant Feature Transform) | 크기 및 회전 불변 특징 추출 및 기술 |
| SURF (Speeded Up Robust Features) | SIFT와 유사한 강건성을 유지하며 속도 향상 |

4. 주요 분류 및 군집화 알고리즘

| 용어 | 정의 및 핵심 내용 |
|---|---|
| 로지스틱 회귀 (Logistic Regression) | 종속 변수가 범주형(이진)일 때 사용하며, 선형 회귀 결과를 시그모이드 함수를 사용하여 0과 1 사이의 확률로 변환하여 예측. |
| 시그모이드 함수 (Sigmoid Function) | 로지스틱 회귀에서 사용되는 함수로, 입력값을 0과 1 사이의 확률로 변환하는 S자 형태의 곡선. |
| K-최근접 이웃 (K-Nearest Neighbors, KNN) | 주변의 가장 가까운 K개의 데이터를 보고 새로운 데이터가 속할 그룹을 판단하는 분류 알고리즘. (유클리드 거리 사용) |
| 서포트 벡터 머신 (Support Vector Machine, SVM) | 데이터를 구분하는 최적의 결정 경계(Decision Boundary)를 정의하는 모델. 결정 경계와 서포트 벡터 사이의 마진을 최대화함. |
| 결정 경계 (Decision Boundary) | 분류 모델에서 서로 다른 클래스에 속하는 데이터 포인트를 구분하는 경계선이나 초평면 |
| K-means 알고리즘 | 비지도 학습의 대표적인 군집화 알고리즘으로, 데이터 포인트를 K개의 그룹으로 나누고, 각 그룹의 평균(중심점, centroid)을 기준으로 반복하여 군집으로 형성. |
| 엘보우 기법 (Elbow Method) | K-means에서 최적의 군집 수(K)를 찾기 위해, K값을 증가시키면서 군집 내 거리의 합이 급격히 줄어드는 구간을 찾는 방법. |
| 계층적 군집화 (Hierarchical Clustering) | 데이터 포인트들을 계층적인 트리 구조(덴드로그램)로 군집화하는 알고리즘. (병합적/분할적 방법) |
| DBSCAN | 밀도에 기반하여 클러스터를 형성하는 군집화 알고리즘. 이상치(outlier)를 효율적으로 처리하며, 비구형 클러스터를 잘 찾음. |

5. 모델 평가 지표

| 용어 | 정의 및 핵심 내용 | |
|--------------------------|---|---|
| 혼동 행렬 (Confusion Matrix) | 특정 분류 모델의 성능을 평가하는 지표로, 실제값과 예측값을 한눈에 확인할 수 있도록 나타낸 행렬. | |
| 정확도 (Accuracy) | 모델이 전체 데이터 중에서 정확하게 예측한 비율. | $TP + TN / TP + FP + FN + TN$ |
| 정밀도 (Precision) | 모델이 Positive로 예측한 것 중에서 실제 True Positive인 비율. (False Positive를 낮추는 데 중요) | $TP / TP + FP$ |
| 재현율 (Recall) | 실제 Positive인 것 중에서 모델이 True Positive로 정확하게 예측한 비율. (False Negative를 낮추는 데 중요) | $TP / TP + FN$ |
| F1 스코어 (F1-score) | 정밀도와 재현율의 조화 평균으로, 두 지표의 균형 잡힌 성능을 나타냄. | $2 * (Precision * Recall / (Precision + Recall))$ |
| 교차 검증 (Cross Validation) | 모델의 일반화 능력을 평가하기 위해 전체 데이터를 k개로 나누어 일부는 학습하고 나머지는 테스트하는 과정을 반복하는 방법. (예: K-fold validation) | |

| | | 실제 정답 | |
|-------|-------|----------------|----------------|
| | | True | False |
| 분류 결과 | True | True Positive | False Positive |
| | False | False Negative | True Negative |