**REPORT**

(CHAPTER 6)

로고, 상징, 등록 상표, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

|  |  |
| --- | --- |
| 이름 | 김민서 |
| 학과 | 컴퓨터인공지능공학부 |
| 학번 | 202213083 |
| 과목명 | 인공지능입문 |
| 교수님 | 진예지 교수님 |
| 분반 | 105분반 |
| 제출일 | 2025.05.14 |

**2. 다음은 예측 모델의 평가 지표다. 지표 1) ~ 4)의 계산 방법과 해석 방법을 설명하시오.**

|  |  |  |  |
| --- | --- | --- | --- |
| 지표 | 우리말 이름 | 계산 방법 | 해석 방법 |
| MSE | 평균 제곱 오차 | 각 데이터에서 발생한 오차를 제곱한 후 계산된 평균 | 예측 결과가 정확할수록 0에 가까운 값을 가짐 |
| RMSE | 평균 제곱근 오차 | MSE의 제곱근 |
| MAE | 평균 절대 오차 | 각 데이터에서 발생한 오차 절댓값의 평균 |
| R2 | 결정 계수 | 분산을 사용해 정규화된 MSE를 1에서 뺀 값 | 예측 결과가 정확할수록 1에 가까운 값을 가짐 |

**3. 다음 중 특정 변수의 수치를 직접 예측하는 데에 사용하는 모델은 무엇인가?**

1) 로지스틱 회귀분석

* → 범주형을 예측하는 분류 모델
* 정답 아님

2) **선형 회귀분석**

* → 연속적인 수치형 변수를 예측
* 정답

3) 랜덤 포레스트

* 주된 목적이 분류이므로 정답 아님

4) 의사 결정 트리

* 주로 분류에서 사용
* 선형 회귀분석만큼 명확하지 않음

**8. 지도 학습의 성능 평가 지표에는 분류 정확도(CA), 정밀도(Precision), 재현율(Recall), F1 score, AUC등이 있다. 혼동행렬(Confusion Matrix)을 이용하여 각 평가 지표를 계산하는 방법에 대해 설명하시오.**

- TN / FP / FN / TP에 의한 계산식

|  |  |
| --- | --- |
| 지표 | 계산 |
| CA | (TP / TN) / Total |
| Precision | TP / (TP + FP) |
| Recall | TP / /(TP + FN) |
| F1 score | 2 / ((1 / Precision) + (1 / Recall)) |
| AUC | TPR – FPR 커브의 아래쪽 면적 |

* TPR = TP / (FN / TP)
* FPR = FP / (TN + FP)

**10. 6.5장에서 SVM과 Neural Network 위젯을 이용하여 심장병이 있는지 학습하고 Test and Score 위젯으로 모델의 성능을 검증하였다. 이때, Neural Network 위젯의 Neurons in hidden layers 항목에서 은닉층의 수를 변화시키면 성능이 어떻게 달라지는지 분석하시오.**

- Neural Network의 복잡도는 은닉층에 있는 노드 수에 따라 결정된다.

- 복잡한 데이터를 인식해야 하는 경우는 노드의 숫자를 늘려야 하고, 특징의 수가 작고 이해하기 쉽다면 노드를 적게 사용하는 것이 좋음.

|  |  |  |
| --- | --- | --- |
| 노드 수 변화에 따른 결과 | | |
|  |  |  |
| Layers = 200, 400, 600 | Layers = 400 | Layers = 800 |

- Layers = 200은 예제에서 다뤘기에 생략함

1) 은닉층의 노드 수 증가

- 모델의 표현력이 향샹됨

- 학습 속도가 저하함

2) 은닉층의 노드 수 감소

- 모델의 표현력이 부족해짐 -> 데이터의 패턴을 충분히 학습하지 못함

- 노드 수가 너무 적으면 오히려 성능이 줄어듬

3) 은닉층의 수 증가

- 복잡한 패턴을 점진적으로 추상화하며 학습 가능

- 충분한 데이터와 적절한 정규화에 따라 성능이 향상될 수 있음

- 학습 시간이 증가

- 데이터가 부족하면 성능이 오히려 저하될 수 있음