인공지능 설계 및 실습 Homework #2

19017029 김주엽

Problem 1. data-03-diabetes.cav 파일을 사용하여 비만 여부를 판단하는 모델을 설명 및 구현하고 해당 모델이 충분히 학습된 모델임을 보이시오. (50점)

* 비만 여부를 판별하는 것은 비만인 경우(0)와 정상인 경우(1)로 나눌 수 있으므로 분류에 해당하는 문제로 논리회귀분석을 사용하여 모델을 구현할 수 있습니다.
* 논리회귀분석을 사용하고 최대 훈련 횟수를 50000, learning rate를 0.01로 했을 때 모델 정확도 및 cost함수의 그래프는 다음과 같습니다.

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

1. 정확도 그래프 2. cost 함수 그래프

* 모델을 0~2만 번 훈련했을 때 정확도 그래프는 처음에는 증가하지만 중간에 감소와 증가를 반복하는 것을 확인할 수 있었고 cost함수의 값은 계속 감소하는 것을 볼 수 있습니다. 모델을 2~5만 번 훈련했을 때 정확도와 cost함수의 변화량이 미미한 것을 보아 더 이상 훈련을 진행해도 값의 변화가 적기에 모델이 학습을 마쳤다고 볼 수 있습니다. 이때 정확도는 약 0.77%이고 cost함수의 값은 약 0.47에 가까운 값을 얻을 수 있었습니다. 따라서 이 모델은 데이터 불균형으로 인해서 정확도가 낮아 모델을 활용하는데 어려움이 있다고 생각합니다.

Problem 2. data-04-zoo.csv 파일을 사용하여 동물을 분류하는 모델을 설명 및 구현하고 해당 모델이 충분히 학습된 모델임을 보이시오. (50점)

* 위의 파일은 동물에 대한 여러 정보와 그에 맞는 동물 타입을 담고 있습니다. 동물의 타입이 총 0~6까지 7개로 다중논리회귀분석에 적합하며 동물 타입을 분류하기 위해서 우선적으로 타입값을 원-핫 인코딩 해준 후에 다중논리회귀분석 알고리즘을 이용하여 구현하였습니다.

텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명텍스트, 라인, 도표, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. 정확도 그래프 2. cost 함수 그래프

* 모델을 0~7500번까지 훈련했을 때 정확도는 계속해서 증가하고 cost함수 값은 감소하는 것을 확인할 수 있습니다. 또한 약 7000~8000번 훈련했을 때 이미 모델의 정확도는 100%로 예측값과 실제값을 비교했을 때 일치하는 것을 볼 수 있었습니다. 이후에 정확도가 100%임에도 불구하고 cost함수 값은 계속해서 감소하는데 그 값의 변화량은 크지 않는 것을 확인할 수 있었습니다. 따라서 위의 모델은 동물을 분류하는 데 사용하기에 적합한 모델이라고 생각합니다.