

Pattern Recognition Assignment: Regression Analysis Using the California Housing Dataset폰트, 그래픽, 로고, 그래픽 디자인이(가) 표시된 사진

자동 생성된 설명

제출일 2024.11.1

과목명 패턴인식

전공 모바일시스템공학과

학번 32204012

이름 전혜림

목차

1. Introduction

1. 과제의 목적 및 중요성
2. 캘리포니아 주택 데이터셋 및 그 특성

2. Dataset Overviews

1) California housing dataset의 구조와 feature

3. Experiment and Results

1) Task 1: Regression using Least Squares

2) Task 2: Regression using Maximum Likelihood Estimation (MLE)

3) Task 3: Regression Using Scikit-Learn

4. Conclusion.

1. Introduction (Background)

이번 과제의 목적은 캘리포니아 주택 데이터셋을 기반으로 다양한 회귀 분석 기법을 적용하여 주택 가격을 예측하고, 각 기법의 성능을 비교하여 회귀 분석에 대해 이해하는 것입니다. 회귀 분석은 통계 모델링과 머신러닝에서 핵심적인 기법으로, 주어진 데이터의 패턴을 분석하여 미래의 값을 예측하는 데 중요한 역할을 합니다. 특히, 본 과제는 수동으로 구현한 회귀 모델과 Scikit-learn의 라이브러리 기반 모델을 비교하는 과정에서 각 모델의 장단점을 파악하고 회귀 모델링의 원리에 대한 직관을 기르는 것을 목표로 합니다.

과제는 크게 세 가지 regression 기법을 포함하며, 각각의 기법을 사용하여 주택 가격을 예측하고 성능을 비교합니다. 첫 번째는 최소제곱법(Least Squares) regression 로, 잔차의 제곱합을 최소화하는 방식으로 모델의 파라미터를 수동으로 계산합니다. 두 번째는 Maximum Likelihood Estimation(MLE)을 활용한 regression으로, Negative Log-Likelihood, (NLL)를 최적화하여 파라미터를 추정합니다. 마지막으로, Scikit-Learn의 LinearRegression 클래스를 사용하여 라이브러리 기반의 선형 회귀 모델을 간편하게 구현하고 앞선 두 task의 구현방식과 성능을 비교합니다.

이러한 접근을 통해 회귀 분석의 이론과 실제 구현을 모두 학습하며, 데이터의 패턴을 효율적으로 학습하고 예측하는 모델링 기법의 핵심을 이해할 수 있습니다.

2.Dataset overviews

1)California housing dataset

본 과제에서 사용하는 캘리포니아 주택 데이터셋은 1990년 캘리포니아 주택 조사 데이터를 기반으로, 지역별 주택 가격을 예측하기 위해 설계된 대표적인 머신러닝 학습용 데이터셋입니다. 이 데이터셋은 각 지역의 주택 및 인구 통계에 대한 다양한 특성들을 포함하고 있어, 주택 가격에 영향을 미치는 여러 인구통계학적 변수를 통해 다양한 요인을 분석하고 예측하는 데 유용하게 활용됩니다. 중간 소득(Median Income), 주택 연령, 인구 수 등 다양한 변수로 구성된 이 데이터셋에서 본 과제는 중간 소득(`MedInc’) 변수 하나를 사용해 단순 회귀 분석을 수행합니다. 이를 통해 단일 특성만으로 구축된 회귀 모델이 주택 가격을 얼마나 효과적으로 예측할 수 있는지 평가합니다.

Scikit-Learn 라이브러리의 fetch\_california\_housing() 함수를 통해 불러올 수 있습니다.

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

Scikit-learn library와 dataset load

3. Experiment and Results

3.1) Task 1: Regression using Least Squares

Task1에서는 캘리포니아 주택 데이터셋을 사용하여 최소제곱법(Least Squares) 기반의 단순 회귀 모델을 구현하고, 이를 통해 주택 가격을 예측하였습니다.

최소제곱법은 회귀 분석에서 많이 사용하는 기법으로, 잔차 제곱합(Sum of Squared Residuals) 을 최소화하는 파라미터를 구하는 방식입니다. 이 방법은 데이터를 가장 잘 설명하는 직선을 찾는 데 사용되며, 잔차(예측 값과 실제 값의 차이)가 작을수록 예측의 정확성이 높아집니다. 회귀 모델의 식은 다음과 같습니다.

는 예측된 주택가격, 는 중간소득(Median Income), 는 절편(intercept), 는 기울기(coefficient)를 나타냅니다.

최소제곱법은 다음과 같은 식을 통해 최적의 파라미터 를 구합니다.

는 feature matrix, 는 target variable입니다. 위 수식을 통해 를 계산하면 잔차 제곱합이 최소화되는 linear regression 모델을 얻을 수 있습니다.

3.1.1) code 분석

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

1. Load the California Housing Dataset

Scikit-Learn의 fetch\_california\_housing() 함수를 사용하여 데이터셋을 load합니다. 데이터셋의 다양한 특성 중 중간 소득(Median Income) 특성만을 선택하여 단순 회귀 분석을 수행하였습니다. X는 중간 소득(MedInc) 특성으로 구성되고, y는 목표 변수인 주택 가격으로 구성됩니다.

2. Add a Bias (Intercept) Term

회귀 모델에서 절편 항을 고려하기 위해 X에 1로 구성된 열을 추가하여 X\_b를 생성하였습니다. 이를 통해, 회귀 모델의 식에 절편이 포함될 수 있습니다.

3. Calculate Parameters using Least Squares Formula

최소제곱법 수식인 를 이용하여 파라미터 를 계산하였습니다.

np.linalg.inv와 행렬 곱셈을 사용하여 수동으로 회귀 계수를 구하였습니다.

4. Make Predictions

구한 파라미터 를 이용하여 y\_pred로 예측 값을 계산하였습니다. 예측은 np.dot을 사용하여 행렬 곱으로 수행되었습니다.

5. Evaluate the Model using Mean Squared Error (MSE)

예측의 정확성을 평가하기 위해, 평균제곱오차(Mean Squared Error, MSE)를 계산하였습니다. MSE는 잔차 제곱합의 평균을 나타내며, 값이 작을수록 모델이 데이터에 잘 맞는다는 것을 의미합니다.

6. Visualization of Results

실제 주택 가격과 예측된 주택 가격을 시각화하여 비교하였습니다. 산점도는 실제 데이터를, 선 그래프는 예측된 데이터를 나타내며, 두 값의 일치도를 시각적으로 확인할 수 있습니다.

3.1.2) 실행결과와 분석

스크린샷, 텍스트, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

* Intercept (): 절편은 약 0.4509로 계산되었습니다. 이 값은 중간 소득이 0일 때 예측되는 기본 주택 가격을 나타냅니다. 물론 실제 중간 소득이 0이 될 수는 없지만, 절편은 회귀 모델이 전체 데이터를 잘 설명하기 위해 보정하는 값으로 이해할 수 있습니다.
* Slope (): 기울기 값인 0.4179는 중간 소득이 단위 1만큼 증가할 때 예상되는 주택 가격의 증가 정도를 나타냅니다. 즉, 중간 소득이 증가할수록 주택 가격도 함께 증가하는 양의 상관관계를 가지며, 중간 소득이 주택 가격에 상당히 중요한 요인임을 보여줍니다.
* Mean Squared Error (MSE): 평균 제곱 오차가 약 0.7011로, 이는 예측된 주택 가격과 실제 가격 간의 평균 오차의 크기를 나타냅니다. 이 값이 0에 가까울수록 모델이 실제 데이터를 잘 설명하고 있다는 의미입니다. MSE 값이 다소 큰 이유는 본 모델이 단일 특성인 중간 소득(Median Income)만을 사용하여 주택 가격을 예측하고 있기 때문입니다. 주택 가격에 영향을 미치는 변수는 중간 소득 외에도 여러 가지가 있을 수 있으며, 이를 반영하지 않은 단순 모델이기 때문에 오차가 발생할 수 있습니다.
* 시각화 결과를 보면, 파란색 산점도는 실제 주택 가격을 나타내며, 빨간색 선은 예측된 주택 가격을 나타냅니다. 선형 모델이 데이터의 전체적인 경향을 설명하는데 적절한 예측을 하고 있음을 확인할 수 있습니다. 중간 소득이 증가함에 따라 주택 가격이 증가하는 추세를 잘 반영하고 있으며, 선형 관계를 통해 중간 소득과 주택 가격 간의 양의 상관관계를 확인할 수 있습니다.

이번 Task 1에서는 최소제곱법을 사용하여 단순 회귀 모델을 성공적으로 구현하였습니다. 중간 소득이라는 단일 특성을 통해 주택 가격을 예측하면서, 중간 소득과 주택 가격 사이에 양의 상관관계가 있음을 확인할 수 있었습니다. 다만, 모델의 평균 제곱 오차(MSE)가 다소 높게 나온 것은 단일 변수만을 사용한 한계로 볼 수 있으며, 더 정확한 예측을 위해서는 여러 특성을 추가하여 다중 회귀 모델을 구성하는 것도 고려할 수 있습니다. 따라서 이번 Task 1의 결과는 중간 소득이 주택 가격에 중요한 영향을 미치는 요인임을 보여주며, 이를 통해 회귀 분석의 기본적인 원리와 성능 평가 방법을 학습하는 데 의의가 있습니다.

3.2) Task 2 - Regression using Maximum Likelihood Estimation (MLE)

이번 Task 2에서는 Maximum Likelihood Estimation(MLE)을 사용하여 단순 회귀 모델을 구현하고, 이를 통해 주택 가격을 예측하였습니다. MLE는 주어진 데이터를 가장 잘 설명하는 모델 파라미터를 추정하는 기법으로, 본 과제에서는 Negative Log Likelihood(NLL)를 최소화하여 회귀 모델을 최적화합니다.

MLE는 관측된 데이터를 가장 잘 설명하는 확률 분포의 파라미터를 추정하는 방법입니다. 회귀 분석에서 MLE는 목표 변수 와 예측 변수 간의 관계를 선형 모델로 가정하고, 이를 통해 관측된 데이터를 설명하는 최적의 회귀 계수와 분산을 찾습니다.

회귀 모델의 식은 다음과 같습니다.

여기서 는 예측된 주택 가격, 는 중간 소득(Median Income), 는 절편(intercept), 는 기울기(coefficient)를 나타냅니다. MLE에서는 데이터의 잔차를 정규 분포로 가정하여 음의 로그 가능도를 최소화하는 파라미터를 찾습니다.

텍스트, 폰트, 화이트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

3.2.1) code 분석

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. Load the California Housing Dataset

Scikit-Learn의 fetch\_california\_housing() 함수를 사용하여 캘리포니아 주택 데이터셋을 로드하였습니다. Task 1과 마찬가지로, 단일 특성인 MedInc(중간 소득)를 선택하여 단순 회귀 분석을 수행하였습니다.

2. Add a Bias (Intercept) Term

절편 항을 반영하기 위해 np.c\_를 사용하여 X에 1로 구성된 열을 추가하였습니다. 이를 통해 회귀 모델에 절편 항이 포함되도록 설정합니다.

3. Define the Negative Log Likelihood (NLL) Function

negative\_log\_likelihood() 함수를 정의하여 MLE를 위한 NLL 수식을 구현하였습니다. 이 함수는 절편, 기울기, 잔차의 분산을 매개변수로 받아 음의 로그 가능도를 계산합니다. 예측 값과 실제 값의 차이를 바탕으로 NLL을 계산하며, 이 값이 최소화될 때 가장 적합한 파라미터를 찾을 수 있습니다.

4. Set Initial Parameters

파라미터 추정을 시작할 초기값으로 절편과 기울기를 0, 분산을 1로 설정하였습니다.

5. Minimize NLL to Find Optimal Parameters

scipy.optimize.minimize() 함수와 L-BFGS-B 알고리즘을 사용하여 NLL을 최소화하고, 최적의 절편과 기울기, 분산을 찾았습니다. bounds 옵션을 사용하여 기울기와 절편의 값 범위를 제한하고, 분산의 최솟값을 설정하여 안정적인 최적화를 수행하였습니다.

6. Calculate Predictions using Optimal Parameters

구한 최적의 파라미터를 사용하여 예측 값을 계산하였고, 이를 통해 모델이 얼마나 잘 학습되었는지를 확인하였습니다.

7. Calculate Mean Squared Error (MSE)

mean\_squared\_error() 함수를 사용하여 평균 제곱 오차(MSE)를 계산하고 모델의 성능을 평가하였습니다.

8. Visualization of Results

실제 주택 가격과 예측된 주택 가격을 산점도와 선 그래프로 시각화하여, MLE 모델이 데이터의 경향을 얼마나 잘 설명하는지 시각적으로 확인할 수 있도록 하였습니다.

3.2.2) 실행결과와 분석

텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

* Intercept (): 절편 값은 약 0.4509로, 중간 소득이 0일 때 예측되는 주택 가격을 의미합니다. 현실적으로 중간 소득이 0일 가능성은 없지만, 회귀 모델이 전체 데이터를 설명하기 위해 최적화된 값을 보정하여 제공하고 있습니다.
* Slope (): 기울기 값은 약 0.4179로, 중간 소득이 단위 1만큼 증가할 때 예상되는 주택 가격의 증가 폭을 의미합니다. 이 값은 중간 소득과 주택 가격 사이에 양의 상관관계가 있음을 보여주며, 중간 소득이 주택 가격 예측에서 중요한 변수임을 확인할 수 있습니다.
* Variance (σ²): 분산 값은 0.7011로, 모델의 잔차(residual)의 분산을 나타냅니다. 이는 데이터가 모델의 예측에서 벗어나는 정도, 즉 잡음의 크기를 반영합니다. 잔차의 분산이 낮을수록 모델이 데이터의 변동성을 더 잘 설명하고 있음을 의미합니다.
* Mean Squared Error (MSE): 평균 제곱 오차는 약 0.7011로, Task 1에서의 최소제곱법(Least Squares) 결과와 매우 유사합니다. MSE는 예측된 주택 가격과 실제 가격 사이의 평균 오차 크기를 나타내며, 값이 작을수록 모델이 데이터를 더 잘 설명합니다. 이 결과는 MLE와 최소제곱법이 동일한 데이터와 단일 특성을 사용한 경우 비슷한 예측 성능을 나타낼 수 있음을 보여줍니다. 이는 두 방법이 모두 회귀 분석에서 강력한 예측 기법임을 시사합니다.
* Task1과 마찬가지로 시각화 결과의 파란색 산점도는 실제 주택 가격을, 빨간색 선은 MLE 모델이 예측한 주택 가격을 나타냅니다. 예측된 선형 모델이 중간 소득과 주택 가격의 관계를 잘 설명하고 있으며, MLE 모델이 데이터의 전반적인 경향을 반영하여 중간 소득이 증가할수록 주택 가격도 증가하는 양상을 효과적으로 나타내고 있음을 확인할 수 있습니다.

Task 2에서 Maximum Likelihood Estimation(MLE)을 사용한 단순 회귀 모델을 성공적으로 구현하고, 주택 가격 예측을 통해 모델의 성능을 평가하였습니다. 최적의 파라미터로 구한 절편과 기울기는 중간 소득이 주택 가격에 중요한 양의 상관관계를 갖는다는 점을 다시 한번 확인할 수 있었습니다. 또한, MLE는 분산을 추정할 수 있기 때문에 데이터의 불확실성까지 반영하여 통계적 신뢰도를 높일 수 있었습니다.

평균 제곱 오차(MSE) 측면에서 Task 1의 Least squeares과 유사한 성능을 보여줬으며, 이를 통해 두 가지 기법이 회귀 분석에서 모두 효과적인 예측을 수행할 수 있음을 확인하였습니다.

3.3) Task 3 - Regression Using Scikit-Learn

Task 3에서는 Scikit-learn의 LinearRegression 클래스를 사용하여 단일 변수 선형 회귀 모델을 학습하였습니다. LinearRegression은 Python의 Scikit-learn 라이브러리에 포함된 선형 회귀 모델로, 데이터의 입력 변수와 출력 변수 간의 관계를 설명하는 가장 기본적인 회귀 기법 중 하나입니다. 이 클래스는 선형 회귀의 기본 수식을 기반으로 하여 데이터를 모델링하며, 최소 제곱법을 통해 최적의 회귀 계수를 찾습니다. Scikit-learn의 LinearRegression 클래스는 데이터 전처리, 모델 학습, 예측 및 성능 평가 등 다양한 기능을 제공하여 사용자에게 직관적인 API를 통해 선형 회귀 분석을 손쉽게 수행할 수 있도록 합니다.

3.3.1) code 분석

텍스트, 스크린샷, 폰트, 문서이(가) 표시된 사진

자동 생성된 설명텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

1. 데이터 로딩

fetch\_california\_housing 함수를 사용하여 캘리포니아 주택 데이터를 로드했습니다. 이 데이터는 주택 가격을 예측하기 위해 다양한 특징을 포함하고 있으며, 여기서는 단일 변수 회귀를 위해 'MedInc' (지역 중간 소득)를 유일한 독립 변수로 사용했습니다.

2. 모델 학습

LinearRegression 클래스를 초기화하고 fit 메서드를 통해 모델을 학습했습니다. fit 메서드는 데이터의 입력 값과 타겟 값(주택 가격)을 받아 선형 회귀 모델을 훈련시킵니다.

3. 파라미터 추출

학습된 모델의 절편(intercept\_)과 기울기(coef\_)를 추출하여 최적의 회귀 계수를 확인했습니다. 이 값들은 모델이 데이터를 기반으로 학습한 결과로, 타겟 값을 예측하는 데 사용됩니다.

4. 예측 및 성능 평가

학습된 모델을 사용해 predict 메서드로 주택 가격을 예측했습니다. 예측된 값과 실제 값을 비교하여 평균 제곱 오차(MSE)를 계산하고 모델의 성능을 평가했습니다.

3.3.2) 실행결과와 분석

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

이 값들은 중간 소득(MedInc) 값이 주택 가격에 미치는 영향을 설명합니다.

* Intercept (): 절편 값은 약 0.45085로, 중간 소득이 0일 때 예측되는 주택 가격을 의미합니다. 이는 지역의 중간 소득이 없는 상태에서도 기본적으로 존재하는 주택 가격을 나타내며, 주택 가격의 하한선으로 해석될 수 있습니다.
* Slope (): 기울기 값은 약 0.41794로, 중간 소득이 1 단위 증가할 때 예상되는 주택 가격의 증가량을 나타냅니다. 이는 중간 소득이 증가함에 따라 주택 가격이 증가한다는 것을 의미하며, 주택 시장에서 소득과 가격 간의 양의 상관관계를 잘 설명합니다.
* Mean Squared Error(MSE): 모델의 예측값과 실제 주택 가격 간의 오차를 제곱하여 평균낸 값인 평균 제곱 오차(MSE)는 0.7011로 나타났습니다. 이 값이 낮을수록 모델의 예측 정확도가 높다는 의미입니다. Task 1과 Task 2에서의 MSE와 거의 동일하게 나타나 Task 3의 예측 성능이 일관적임을 확인할 수 있었습니다. 이는 서로 다른 회귀 접근 방식이 유사한 결과를 도출할 수 있음을 시사합니다.

3.4) Task1,2,3 비교분석

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

* 정확도: 세 가지 방법에서 계산된 평균 제곱 오차(MSE)는 모두 0.70113…으로 동일하게 나타나, 세 모델이 예측 성능에서 유사한 수준의 정확도를 보여줍니다. 이러한 결과는 각 회귀 방법이 데이터의 패턴을 효과적으로 학습했음을 나타내며, 각 접근 방식이 주어진 데이터에 대해 일관된 성능을 발휘하고 있음을 시사합니다. 이는 모델의 신뢰성을 높이며, 실제 응용에서 다양한 기법 중 어느 하나를 선택해도 비슷한 예측 결과를 기대할 수 있음을 의미합니다.
* 계산 효율성: Task 3에서 사용한 Scikit-learn의 LinearRegression 클래스는 가장 간단하고 직관적인 방식으로 모델을 구축할 수 있게 해줍니다. 최적화가 자동으로 이루어져 코드의 복잡성을 줄여주며, 데이터 전처리 및 모델 학습 단계에서의 효율성을 극대화할 수 있습니다. 따라서 실제 응용에서는 Scikit-learn을 사용하는 것이 매우 효율적입니다.

반면, Task 1의 최소제곱법은 수식으로 계산이 간단하여 작은 데이터셋에서는 적합하지만, 데이터의 양이 많아질수록 계산 속도가 느려질 수 있습니다. 따라서 대규모 데이터셋을 다루는 경우에는 비효율적일 수 있습니다. Task 2의 최대 우도 추정(MLE) 방식은 최적화 알고리즘을 사용하여 강력한 추정을 제공하지만, 그 과정에서 시간이 더 소요될 수 있습니다. 이러한 점은 실제 데이터 분석에 있어 속도와 효율성 간의 균형을 고려해야 함을 보여줍니다.

Scikit-learn의 큰 장점은 코드가 간결하고 구현이 쉬워 개발 시간이 단축된다는 점입니다. 그러나 자동화된 과정 덕분에 회귀 분석의 수학적 원리를 깊이 이해하기 어렵다는 단점도 있습니다. 라이브러리의 작동 원리를 알지 못하면 모델의 한계나 가정에 대한 이해가 부족할 수 있습니다.

반면, Task 1과 Task 2는 수동으로 회귀 모델을 구현하는 과정에서 회귀 분석의 기초적인 수학적 원리를 직접 체험할 수 있는 기회를 제공합니다. 이는 데이터 과학의 이론적 기초를 확고히 하고, 모델이 작동하는 방식을 명확히 이해하는 데 도움이 됩니다.

4. Conclusion.

이번 과제를 통해 캘리포니아 주택 데이터셋을 기반으로 세 가지 회귀 기법인 Least Squares, Maximum Likelihood Estimation(MLE), 그리고 Scikit-learn의 LinearRegression을 적용하여 데이터 예측 성능을 비교하고 분석하였습니다.

과제를 수행하며 Least Squares, Maximum Likelihood Estimation(MLE)와 같은 수동 구현과 Scikit-learn라이브러리 사용 간의 차이를 깊이 이해할 수 있었습니다. Task 1과 Task 2에서는 직접 파라미터를 계산하면서 회귀 분석의 수학적 원리와 최적화 방법을 배웠습니다. 반면, Task 3에서는 Scikit-learn을 활용함으로써 코드의 간결함과 시간 절약, 그리고 자동화된 최적화로 모델 구현의 효율성을 체감할 수 있었습니다. 이를 통해 실제 응용에서는 Scikit-learn과 같은 라이브러리를 사용하는 것이 적합하나, 기본 개념을 이해하기 위해 직접 구현해보는 과정이 중요하다는 것을 깨달었습니다.

Least Squares와 Maximum Likelihood Estimation(MLE)은 모두 주어진 데이터를 가장 잘 설명하는 최적의 선형 회귀 계수를 찾기 위한 통계적 기법이지만, 수학적 접근 방식에서 차이가 있습니다. Least Squares는 잔차 제곱합을 최소화하는 고전적인 회귀 접근으로, 간단한 수식을 통해 해를 구할 수 있어 이해와 구현이 용이합니다. 반면, MLE는 음의 로그 가능도를 최소화하여 데이터를 가장 잘 설명하는 회귀 계수를 찾으며, 회귀 모델에 대한 확률적 관점에서 데이터의 불확실성을 고려한 강력한 추정을 제공합니다.

Scikit-learn의 LinearRegression은 코드가 간결하고 효율적인 라이브러리 기반 회귀 방법으로, 실제 응용에서 빠르고 정확하게 모델을 구성할 수 있었습니다. 특히, 자동화된 최적화와 검증 기능은 대규모 데이터셋에서도 신속하게 처리할 수 있는 이점을 제공합니다. 실습 결과, 세 가지 방법 모두 일관된 MSE(Mean Squared Error)를 나타내어 유사한 예측 성능을 보였으며, 이는 다양한 회귀 접근 방식이 동일한 결과를 도출할 수 있음을 보여줍니다.

과제를 마치며, 향후에는 다중 회귀 분석 및 교차 검증을 포함한 다양한 평가 방법을 추가하여 모델 성능을 보다 심층적으로 분석하고 싶다는 생각이 들었습니다. 특히, 추가 특성을 고려한 다변수 회귀 분석과 비선형 모델링 기법을 도입하여 실제 더 복잡한 데이터를 정확하게 예측할 수 있는 모델을 개발할 수 있을 것이라 생각합니다.

과제를 수행하면서 수학적 원리를 직접 적용하고, 다양한 회귀 기법의 장단점을 비교하는 과정을 통해 패턴인식 과목에 대한 이해도가 한층 높아졌다고 느낍니다.

앞으로도 지속적인 학습과 실험을 통해 이론적 지식을 실제 문제에 적용할 수 있는 능력을 키워나가고 싶습니다.