|  |  |  |
| --- | --- | --- |
| **[EEC3600-001] 수치해석** | | |
| 소속: 전기전자공학부 | 학번: 12191529 | 이름: 장준영 |
| **Term Project** | | **Prob #3** |

1. **Problem**
2. 문제

|  |
| --- |
|  |

본 문제에서는 2016년 5월 한 달간 LAX의 시간당 온도 데이터를 기반으로, Auto-Regressive (AR) 모델을 구성하고, Least-Squares 방법을 통해 계수 벡터 를 추정한다.

1. **Solution (a), (b) : M = 8**

|  |
| --- |
| ■ **데이터 정의**    • LAX의 5월 온도 시계열 데이터를 벡터 형태로 반환하는 함수이다.  • 전체 길이는 개로 구성되어 있다.  ■ **데이터 로딩 및 파라미터 설정**    • 시계열데이터를 불러오고, AR 모델의 메모리 길이를 로 설정한다.  • 이에 따라 학습 가능한 입력-출력 쌍은 736개로 제한된다.  ■ **디자인 행렬 및 목표 벡터 구성**    • 각 입력 는 8개의 과거 온도값으로 구성되며, 출력 는 그 다음 시점의 온도이다.  • 이렇게 구성된 는 Linear-regression 문제로 정리 가능하다.  ■ **최소제곱 해 구하기**    • 최소제곱 해법 를 통해 AR 계수를 추정한다.  • 이는 NumPy의 “lstsq()” 함수를 사용하여 수치적으로 안정적으로 계산된다.  ■ **예측 및 RMS 오차 계산**    • 예측값은 로 계산된다.  • 실제값과의 차이에 대한 Root Mean Square (RMS) 오차는 모델의 예측 정확도를 나타낸다.  ■ **예시 실행 및 결과 확인**       * 실제 온도와 예측 온도를 함께 시각화하여 예측 성능을 직관적으로 확인할 수 있다. |

**[결과 분석]**

1. **AR(8) 계수**

**텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.**

계수는 최근 값 (1-step lag)에 큰 비중을 두고, 나머지는 비교적 낮거나 음수 계수로 반영되어 있다. 이는 최근 온도 정보가 미래 예측에 가장 중요함을 시사한다.

1. **RMS 오차**

**텍스트, 라인, 그래프, 도표이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.**

****

평균적으로 약 오차 내에서 예측이 이루어지고 있다. 시계열의 추세를 잘 따라가며, 예측 정확도가 준수한 수준임을 보여준다.

1. **Solution (c): M=4, 8, 12, 24**
2. **시뮬레이션 목적 및 방법**

문제 3-(a), (b)에서 수행한 Auto-Regressive (AR) 모델 학습과 RMS 오차 계산 과정을, 서로 다른 메모리 크기 에 대해 반복 수행한다. 이 과정을 통해 모델의 메모리 길이에 따른 예측 정확도 (RMS Error) 변화를 관찰하고자 한다.

각 메모리 길이 M에 대해 다음 과정을 반복한다:

1. 총 744개의 온도 시계열 데이터에서, 개의 과거 값으로 구성된 입력 벡터 와 다음 시점의 실제값 를 구성한다:
2. 선형 최소제곱법을 이용하여 AR 계수 을 추정한다.
3. 예측값 를 계산하고, 실제 와 비교하여 RMS 오차를 구한다.
4. 이렇게 얻어진 RMS 오차를 M에 따라 정리하고, 그래프 (Memory size vs. RMS Error)를 시각화한다.
5. **Python Code 설명**

**텍스트, 스크린샷, 소프트웨어, 운영 체제이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.**

* 각 에 대해 design matrix 와 target 를 구성하고, 최소제곱해 를 구하여 예측을 수행한다.
* 예측 오차를 RMS 기준으로 평가하여 rms\_errors 리스트를 작성한다.

앞서 (a)에서 다음과 같은 선형 시스템 전개식을 유도하였다:

1. **시뮬레이션 결과 및 그래프 해석**

**라인, 텍스트, 그래프, 도표이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.**

이번 문제에서는 AR 모델의 메모리 크기 (M)에 따라 예측 성능이 어떻게 달라지는지를 분석하였다. 이를 위해 M=4, 8, 12, 24의 네 가지 값을 설정하고, 각각의 경우에 대해 AR 모델을 최소제곱법으로 학습한 후, 예측값과 실제값의 차이를 Rood Mean Square (RMS) 오차로 평가하였다. 그 결과, M=4일 때는 RMS 오차가 약 1.05로 가장 높았으며, 이는 과거 데이터의 양이 부족해 시계열의 패턴을 충분히 반영하지 못한 것으로 해석된다. 반면, M=8, M=12에서는 각각 RMS Error가 약 1.02와 1.00으로 다소 감소하였고, 이는 메모리 길이가 증가함에 따라 모델이 더 많은 시계열 정보를 활용하여 예측력을 개선하였음을 의미한다.

가장 흥미로운 결과는 M=24에서 나타났는데, 이 경우 RMS 오차가 약 0.83까지 감소하였다. 이는 모델이 하루치 온도(24시간)를 하나의 패턴으로 인식하여 이를 효과적으로 학습했다는 점을 시사하며, 실제 온도 시계열의 일변 주기성과도 잘 부합한다. 즉, 충분한 과거 정보를 제공하면 AR 모델은 계절성 또는 주기성을 포착하여 정확한 예측을 수행할 수 있다.

이러한 결과는 AR 모델의 성능이 메모리 길이에 따라 큰 영향을 받는다는 사실을 보여준다. 작은 M은 정보 부족으로 인해 정확도가 떨어지는 반면, 큰 M은 더 많은 시계열 정보를 포착할 수 있어 성능 향상에 기여한다. 다만 일반적으로 메모리 크기를 무한히 늘리는 것이 항상 좋은 결과를 보장하지는 않으며, 데이터의 주기성과 잡음 수준, 그리고 계산 복잡도 등을 고려하여 최적의 M값을 선택하는 것이 중요하다.