중간고사

2021년 4월 29일

1. 본인에게 배정된 두 국가 중 한 국가의 데이터**(global\_confirmed\_cases\_210420)**를 이용해 다음 분석을 시행하라.
2. 최근 3/20일까지 자료를 이용하여 예측한 Linear regression, Logistic, Gompertz model의 모형 결과로부터 4/20일까지의 **누적확진자수** 예측값을 구할 것 **(여기서 ~3/20일까지의 자료는 training dataset이고 3/21~4/20 까지는 testing dataset으로 처리한 분석임.)**
3. 각 모델의 coefficients 값을 정리

|  |  |  |  |
| --- | --- | --- | --- |
| Model |  |  | **R-square** |
| Linear regression |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | **a** | **b** | **c** | **R-square** |
| Logistic model |  |  |  |  |
| Gompertz model |  |  |  |  |

1. 모형의 예측값과 실제 관측값을 하나의 plot에 나타낼 것. X축은 첫 케이스가 발생한 날부터 흐른 날짜를 나타내고, Y축은 실제 관찰된 값을 점으로 표시하고 각 모형의 예측값은 선으로 나타낼 것.
2. 3/21 – 4/20 까지의 일별 관찰값과 예측값을 사용하여 test MSE 를 구할 것
3. MSE를 기반으로 세 모형 중에서 어떤 모형이 제일 자료를 잘 적합하는지 판단하시오.
4. **covid\_vaccine** 데이터의 최근 3/20 까지의 자료를 사용하여, predictor를 days만 이용하여 **일별확진자수**에 대한 segmented poisson regression model 모형을 적합하고 적합 결과를 그래프로 나타낼 것. **(논문 및 코드참조; break point 2개: , npsi=2, seg.Z = ~ Days\_after\_Start)**

(1) 모형의 예측값과 실제 관측값을 하나의 plot에 나타낼 것. X축은 첫 케이스가 발생한 날부터 흐른 날짜를 나타내고, Y축은 실제 관찰된 값을 점으로 표시하고 각 모형의 예측값은 선으로 나타낼 것.

(2) 3/21 – 4/20 까지의 일별 관찰값과 예측값을 사용하여 test MSE 를 구할 것

1. **covid\_vaccine** 데이터의 최근 3/20 까지의 자료를 사용하여, predictor를 days와 백신접종변수 people\_vaccinated를 이용하여 **일별확진자수**에 대한 segmented poisson regression model 모형을 적합하고 적합 결과를 그래프로 나타낼 것. **(논문 및 코드참조; break point 2개: npsi=2, seg.Z = ~ Days\_after\_Start,)**

(1) 모형의 예측값과 실제 관측값을 하나의 plot에 나타낼 것. X축은 첫 케이스가 발생한 날부터 흐른 날짜를 나타내고, Y축은 실제 관찰된 값을 점으로 표시하고 각 모형의 예측값은 선으로 나타낼 것.

(2) 3/21 – 4/20 까지의 일별 관찰값과 예측값을 사용하여 test MSE 를 구할 것

(3) 백신의 효과의 여부를 데이터에 기반하여 서술하시오