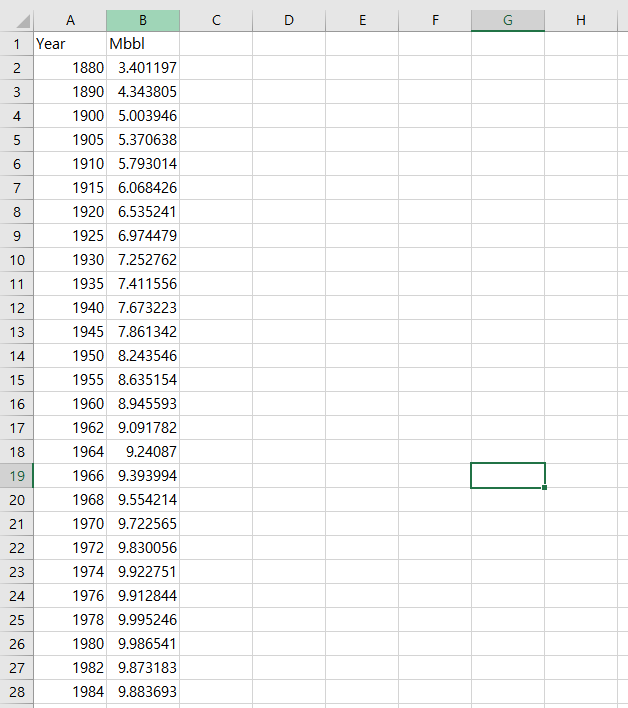
컴퓨터 응용확률 과제1

12171651 컴퓨터공학과 오윤석

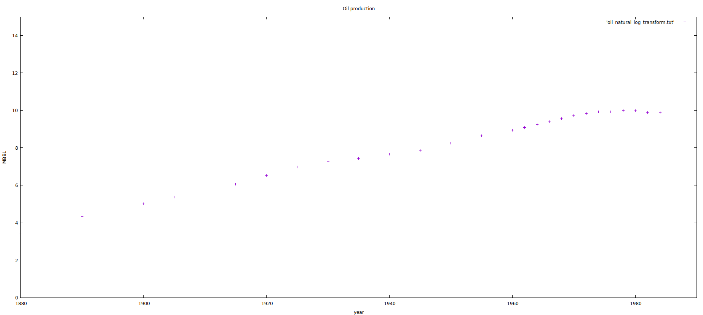
**1. Draw graphs using GNU plot**

우선 이 데이터를 분석하려면 그냥 데이터를 분석하기 보다는 데이터 변환이 필요하다. Background를 보면 자연로그를 취함으로써 데이터를 변환하라고 적혀 있었다. 물론 프로그래밍으로 자연로그 변환을 하면 좋지만 좀 더 가시적으로 데이터를 보고싶어 엑셀 함수를 활용하여 데이터를 변환하였다.

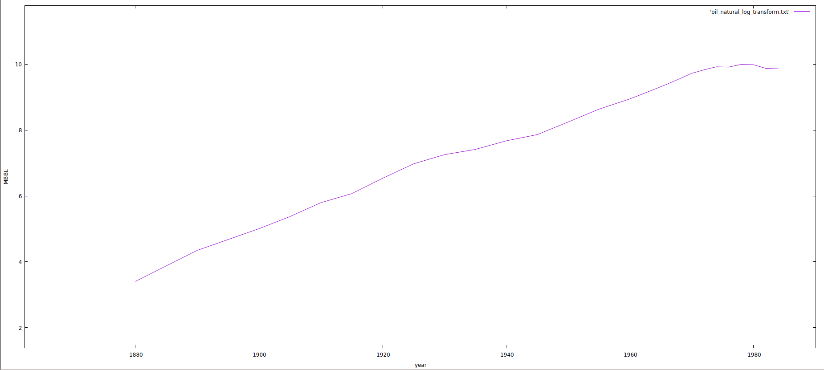


<엑셀 함수 활용한 자연로그 데이터 변환>

그 다음 GNU plot을 활용하여 그래프를 그렸다. 윈도우에서 다운을 받았지만 프로그램이 실행이 되지 않아 우분투에서 실행을 하였다.



<GNU plot을 사용해 점으로 데이터를 표현>



<line을 사용하여 조금 더 가시적으로 표현>

변수가 2개뿐이라 따로 삭제하거나 추가해야 되는 부분은 필요하지 않았다.

**2. Data transform**

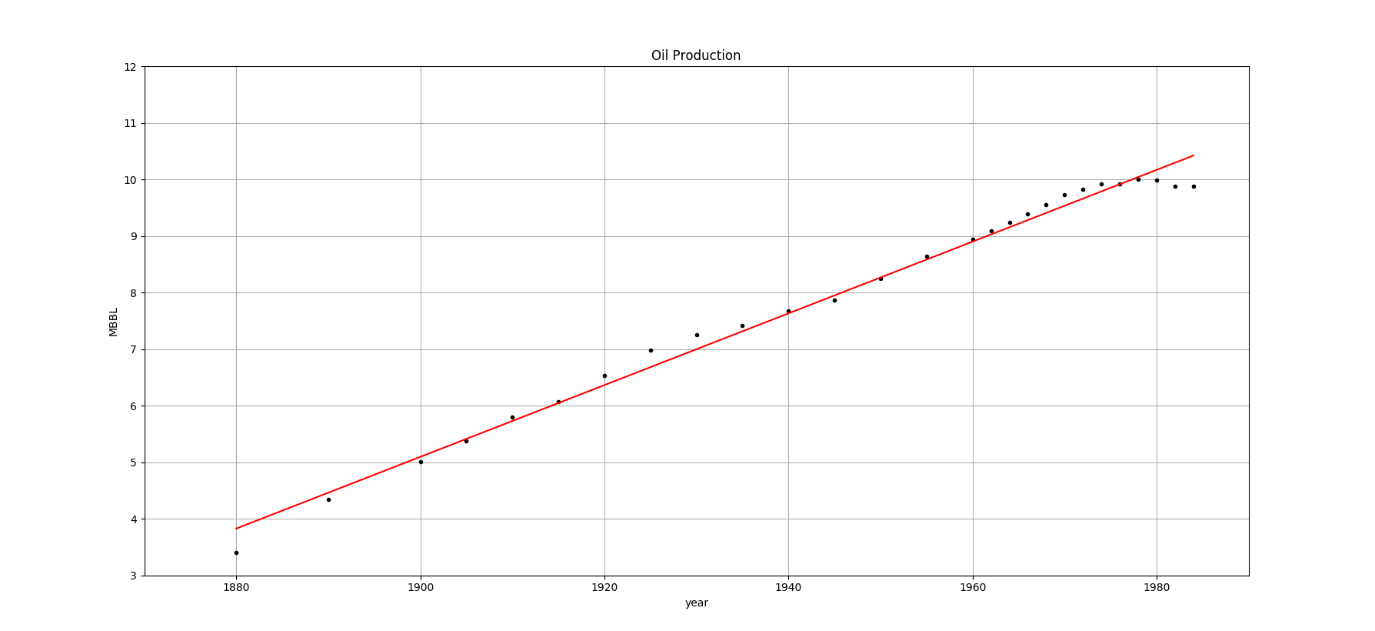
사용할 수 있는 언어는 많지만 본인은 python이 편리하여 python을 사용하여 수행하였다.

python에는 matplotlib이라는 그래프를 편리하게 그려주는 라이브러리와 sklearn이라는 회기 분석에 필요한 도구가 되는 라이브러리를 사용하였다.



<소스 코드>

이 코드를 보면 LinearRegression()이라는 함수가 보이는데 이 함수에서 MSE를 통하여 예측 그래프를 자동으로 그려준다. 그 후 원래 데이터와 이 예측선을 그래프에 그려보았다. 여기서 MSE란 Mean Squared error인데 한국어로 하면 평균 제곱 오차이다. 이를 통해 다음 값을 추측을 할 수 있다.



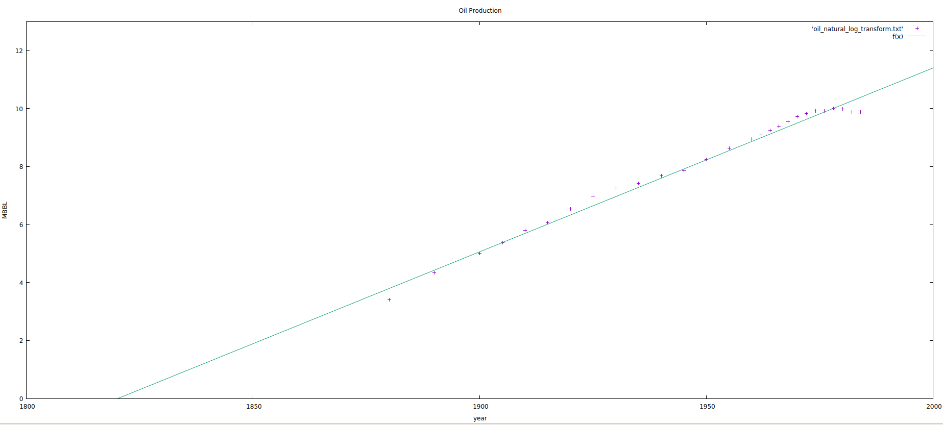
<matplotlib으로 그린 그래프>

우리 과제에서는 GNU plot을 활용한 그래프가 필요해 우선 이 예측선의 함수를 구하였다. (코드 맨 아래에 기울기와 y절편을 구할 수 있는 코드가 있다.)



이를 통해 그래프 방정식을 구해보면 y=0.06342-115.4003으로 구할 수 있다. (소수점4자리 반올림)

이 그래프와 변환된 데이터를 GNU plot에 넣어보면 결과는 아래와 같다.



<GNU plot을 활용한 예측 그래프>

이를 통해 Pattern of exponential growth의 problem을 증명할 수 있다.

**3. Provide your own analysis about the oil production**

2.c도 포함하여 서술을 하겠다. 1973년부터 원유 생산량은 감소하기 시작하였다. 그 이유는 사회적 영향이 컸다. 바로 1973년 중동지역 정치적 혼란으로 인해 원유 생산이 정체되었기 때문이다.

만약 이러한 사건들과 마찬가지로 원유 생산의 영향이 가는 사건들이 없다면 본인이 제작한 예측 그래프에 따라 증가할 것이다.

하지만 원유는 한정된 자원이기 때문에 무한하게는 증가하지 못할 것이다.

**4. 개발환경**

Ubuntu 18.04, python3, GNUplot, matplotlib, sklearn