

문제 1)

	문제 1-a)	문제 1-b)		문제 1-a)	문제 1-b)
최대전력수요	61726	62036	11~12h	61288	61634
최소전력수요	50260	49100	12~13h	57080	57441
0~1h	54680	53381	13~14h	59317	59770
1~2h	51830	50551	14~15h	60076	60588
2~3h	50503	49223	15~16h	59787	60300
3~4h	50260	49103	16~17h	60038	60430
4~5h	51012	50032	17~18h	59996	60247
5~6h	52416	51563	18~19h	60477	60813
6~7h	54277	53635	19~20h	60155	60665
7~8h	55829	55519	20~21h	58331	58748
8~9h	58795	58810	21~22h	57130	57356
9~10h	61019	61191	22~23h	58482	58768
10~11h	61726	62036	23~0h	59109	59451

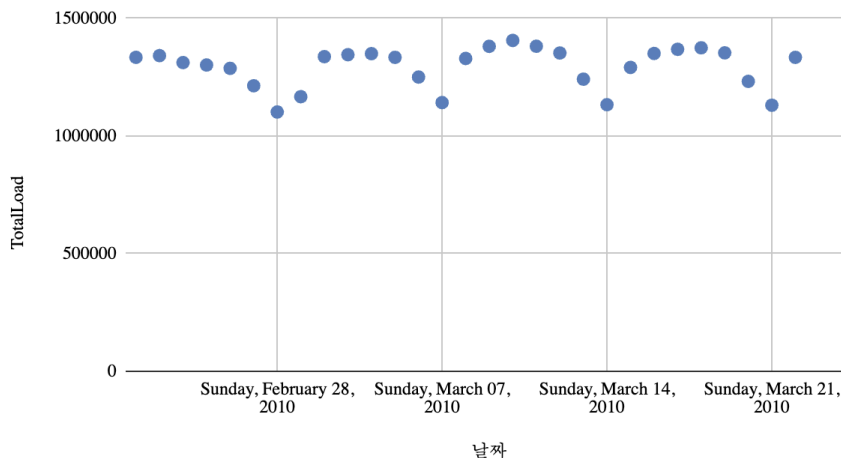
	최대전력수요 예측오차	최소전력수요 예측오차	전력수요 일평균 예측오차
문제 1-a)	2.47%	0.38%	1.69%
문제 1-b)	2.98%	2.68%	2.59%

3월 12일 금요일의 전력수요를 예측하기 위해 a)에서는 3월 10일 수요일, 3월 9일 화요일, 3월 5일 금요일의 전력수요 데이터를 사용하여 이동평균법으로 예측하였으며, b)에서는 동일하게 이동평균법을 사용하였으나 3월 5일 금요일 대신 3월 8일 월요일의 전력수요 데이터를 사용하였다. 방법 a)가 전체적으로 예측오차가 낮은 이유는, 3월 5일 금요일의 전력수요가 3월 8일 월요일의 데이터보다 3월 12일 금요일의 전력수요와 더 유사한 패턴을 보이기 때문이다. 전력수요는 산업용 수요의 영향을 크게 받는데, 월요일의 경우 휴무일인 일요일 다음날이라, 주말에 설비를 멈추는 경우 설비를 재가동해야 하며, 근로자들을 새벽에 출근시키기 어려운

측면이 있다. 따라서 다른 평일보다 전력수요가 낮은 특성을 보이며, 금요일의 전력수요를 예측할 때 해당 주 월요일의 데이터보다 전 주 금요일의 데이터를 사용하는 것이 더 적절하다.
문제 2)

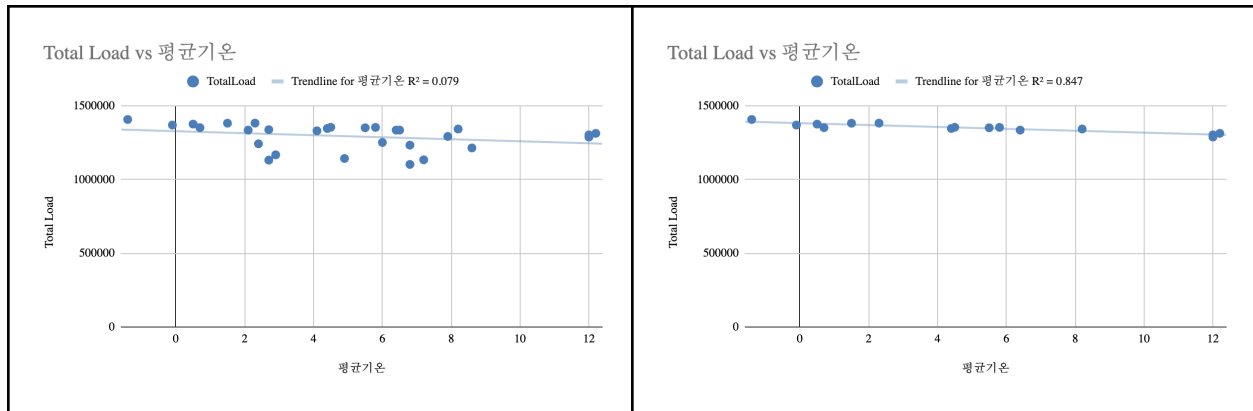
주어진 데이터는 2월 22일부터 3월 22일까지의 시간대별 전력수요, 최대전력수요, 최소전력수요, 총 전력 수요, 최고기온, 최저기온, 평균기온, 운량, 일조시간이며, 이 데이터를 기반으로 3월 24일 수요일의 전력수요 데이터를 예측하고자 한다. 3/1 이 삼일절로 공휴일이라는 특이사항이 존재한다.

TotalLoad vs. 날짜



우선 날짜별 총 전력수요를 그래프로 그려보면, 사인 함수의 형태를 띠는 것을 확인할 수 있다. 대체로 수요일/목요일에 전력수요가 최대로 나타나고, 주말에 줄어들어 일요일에 최소가 되며, 월요일부터 회복되어 화요일, 수요일까지 소폭 상승하여 수요일/목요일에 최고점이 나타나는 추세를 확인할 수 있다. 전력수요에 상업/공업이 큰 영향을 미치기 때문에, 휴무일인 토요일/일요일에 전력수요가 줄어들며, 평일에 높은 추세가 나타나는 것이다. 이러한 특성으로, 전력수요를 모델링하는 데 요일을 고려하는 것이 매우 중요하다. 따라서, 3월 24일 수요일의 전력수요를 예측하기 위해 활용할 적절한 날짜들을 선별해야 한다. 대체로 월요일은 휴무일 다음날이기 때문에 주말동안 쉬었던 설비를 재가동해야 하는 경우가 존재하며, 새벽에 직원들을 출근시키기 상대적으로 어렵기 때문에 화요일, 수요일, 목요일, 금요일과는 다른 패턴을 보인다. 따라서, 24일의 전력수요를 예측하기 위해 주어진 데이터 중 화요일, 수요일, 목요일, 금요일의 데이터만 사용한다. 또한, 3월 1일 월요일이 3.1절 공휴일을 낀 연휴였기 때문에, 3월 2일 화요일도 통상 월요일과 비슷한 전력수요 패턴을 보일 개연성을 고려하면 3월 2일은 제외하였다.

최종적으로 2/23, 2/24, 2/25, 2/26, 3/3, 3/4, 3/5, 3/9, 3/10, 3/11, 3/12, 3/16, 3/17, 3/18, 3/19 총 15일의 데이터를 활용하여 3월 24일의 전력수요를 예측하기로 결정하였다.



기온의 경우 전력수요의 큰 비중을 차지하는 난방수요와 직결되기 때문에, 매우 중요한 변수이다. 요일을 구분하지 않고 모든 데이터를 사용하여 일별 총 전력수요와 평균기온을 모델링했을 때, R^2 값이 0.079로 낮게 나오지만, 토요일, 일요일, 월요일, 공휴일인 3/1일, 공휴일 다음날인 3/2일을 제외하고 전력수요와 평균기온을 모델링하면 R^2 값이 0.847로 매우 상관관계가 높게 나타난다. 또한, 평균기온의 R^2 값이 최고기온, 최저기온의 R^2 값 (각각 0.753, 0.843) 보다 높게 나타났다. 따라서, 평균기온이 전력수요를 예측하는 데 있어서 유효한 변수임을 확인할 수 있다.

또한, 운량과 일조시간 역시 체감온도와 관련되어 냉난방 수요에 영향을 미치고, 태양광 소규모 자가발전기의 발전량에 영향을 미쳐 수요가 줄어드는 것 처럼 보이게 하는 효과가 있기 때문에 유의미한 변수라고 할 수 있다. 다만, 예측을 수행하려는 2010년 당시에는 태양광 소규모 자가발전이 현재보다 덜 보급되어, 일조량이 현재만큼 큰 영향을 미치지 는 못할 것이다.

따라서, 수요예측을 위해 해당일의 평균기온, 운량, 일조시간을 X로, 각 시간대의 전력수요를 y로 하여 선형회귀 모델을 설계하였다. 사용할 수 있는 데이터가 15일치로 제한적이기 때문에, 머신러닝 기반 모델을 사용하기에는 무리가 있어, 비교적 간단하고 설명가능한 선형회귀 모델을 선택하였다. 모델의 정확도를 검증하기 위해, 각 날짜마다 나머지 14일의 데이터를 사용해 선형회귀 모델을 피팅하고, 그 모델을 사용해 해당 날짜의 오차를 계산하였다 (Leave-one-out Method). 편의를 위해 파이썬을 사용하였다.

```

0~1h : 2.18 %
1~2h : 2.51 %
2~3h : 2.57 %
3~4h : 2.47 %
4~5h : 2.22 %
5~6h : 1.90 %
6~7h : 1.81 %
7~8h : 1.80 %
8~9h : 1.72 %
9~10h : 1.47 %
10~11h : 1.20 %
11~12h : 1.02 %
12~13h : 1.16 %
13~14h : 1.05 %
14~15h : 1.03 %
15~16h : 0.98 %
16~17h : 0.58 %
17~18h : 0.85 %
18~19h : 1.09 %
19~20h : 0.84 %
20~21h : 0.73 %
21~22h : 0.62 %
22~23h : 0.51 %
23~0h : 0.65 %

Overall mean MAPE across 24 models: 1.37 %

```

선형회귀 모델을 피팅하여, 시간대별로 위와 같은 평균 오차를 얻을 수 있었다. 여기서, 새벽과 아침시간에 해당하는 0시 ~ 10시의 오차가 눈에 띄게 높은 것을 관찰할 수 있다. 특히, 0시~5시의 새벽시간대는 일관되게 2% 이상의 오차가 나타났다. 새벽시간의 경우 대부분의 사람들이 취침하는 시간대로, 기온의 변화에 난방 사용량이 영향을 받기 어렵다. 종속변수로 해당일의 평균기온을 사용하였는데, 사람들의 생활 패턴상, 새벽시간대의 전력사용량은 해당일의 기온보다 전날의 기온에 영향을 받는다고 보는 것이 타당하다. 따라서, 각 시간대에 대해 1. 이전에 피팅한 해당일의 평균기온을 사용한 모델 2. 전날의 평균기온을 사용한 모델 3. 해당일의 평균기온, 전날의 평균기온 모두 사용한 모델 3가지의 결과를 확인해 보겠다.

```

0~1h : 2.18 %
1~2h : 2.51 %
2~3h : 2.57 %
3~4h : 2.47 %
4~5h : 2.22 %
5~6h : 1.90 %
6~7h : 1.81 %
7~8h : 1.80 %
8~9h : 1.72 %
9~10h : 1.47 %
10~11h : 1.20 %
11~12h : 1.02 %
12~13h : 1.16 %
13~14h : 1.05 %
14~15h : 1.03 %
15~16h : 0.98 %
16~17h : 0.58 %
17~18h : 0.85 %
18~19h : 1.09 %
19~20h : 0.84 %
20~21h : 0.73 %
21~22h : 0.62 %
22~23h : 0.51 %
23~0h : 0.65 %

Overall mean MAPE across 24 models: 1.37 %

```

해당일 평균기온 사용한 모델

```

--- Average MAPE for each hour ---
0~1h : 1.05 %
1~2h : 1.22 %
2~3h : 1.32 %
3~4h : 1.28 %
4~5h : 1.23 %
5~6h : 1.18 %
6~7h : 1.21 %
7~8h : 1.30 %
8~9h : 1.13 %
9~10h : 1.24 %
10~11h : 1.26 %
11~12h : 1.31 %
12~13h : 1.79 %
13~14h : 2.05 %
14~15h : 2.08 %
15~16h : 2.14 %
16~17h : 2.03 %
17~18h : 1.96 %
18~19h : 1.72 %
19~20h : 1.87 %
20~21h : 1.73 %
21~22h : 1.51 %
22~23h : 1.50 %
23~0h : 1.67 %

Overall mean MAPE across 24 models: 1.53 %

```

전날 평균기온을 사용한 모델

```

--- Average MAPE for each hour ---
0~1h : 1.14 %
1~2h : 1.31 %
2~3h : 1.44 %
3~4h : 1.39 %
4~5h : 1.34 %
5~6h : 1.26 %
6~7h : 1.27 %
7~8h : 1.29 %
8~9h : 1.16 %
9~10h : 1.10 %
10~11h : 1.05 %
11~12h : 1.00 %
12~13h : 1.30 %
13~14h : 1.14 %
14~15h : 1.05 %
15~16h : 0.95 %
16~17h : 0.64 %
17~18h : 0.98 %
18~19h : 1.28 %
19~20h : 0.94 %
20~21h : 0.80 %
21~22h : 0.70 %
22~23h : 0.59 %
23~0h : 0.75 %

Overall mean MAPE across 24 models: 1.08 %

```

해당일 평균기온, 전날
평균기온 모두 사용한 모델

전날 평균기온을 사용하면 새벽시간대의 오차가 줄어들지만, 이후 시간대의 오차가 늘어나며, 전날, 해당일의 평균기온을 모두 사용하면 적절하게 **trade off**가 발생하여 평균 오차가 1.08%로 최소화된다.

일반적으로 최고기온은 낮 시간대에, 최저기온은 새벽이나 밤 시간대에 발생한다. 이런 특성을 고려하면, 단순히 평균기온만을 사용하는 것이 아니라 최저기온과 최고기온 역시 모델의

입력으로 사용한다면 오차를 더 줄일 수 있을 것이다. 최종적으로 선형회귀 모델의 입력으로 고려되는 변수들은 최고기온, 최저기온, 평균기온, 전일 평균기온, 일조량, 운량 총 6가지이다. 각 시간대 별로, 이 6가지 변수로 만들 수 있는 모든 부분집합에 대해 선형회귀 모델을 훈련하고, 각 모델에 대해 leave one out 방법으로 15일에 대한 평균 오차를 최소화하는 모델을 찾는다. Python 코딩을 사용하였으며, 결과는 다음과 같다.

```

----- Summary -----
0~1h : 0.84 % (평균기온, 최저기온, temp_yesterday)
1~2h : 0.95 % (평균기온, 최저기온, temp_yesterday)
2~3h : 1.08 % (평균기온, 최저기온, temp_yesterday)
3~4h : 1.03 % (평균기온, 최저기온, temp_yesterday)
4~5h : 1.01 % (평균기온, 운량, 일조시간, 최저기온, 최고기온, temp_yesterday)
5~6h : 1.01 % (평균기온, 최저기온, temp_yesterday)
6~7h : 0.88 % (평균기온, 최저기온)
7~8h : 0.71 % (최저기온)
8~9h : 0.63 % (최저기온, temp_yesterday)
9~10h : 0.67 % (최저기온, temp_yesterday)
10~11h : 0.90 % (최저기온, 최고기온)
11~12h : 0.97 % (평균기온)
12~13h : 1.13 % (운량, 일조시간, 최고기온)
13~14h : 0.94 % (운량, 일조시간, 최고기온)
14~15h : 0.82 % (운량, 일조시간, 최고기온)
15~16h : 0.73 % (운량, 일조시간, 최고기온)
16~17h : 0.54 % (평균기온, 일조시간)
17~18h : 0.78 % (평균기온, 일조시간)
18~19h : 0.94 % (평균기온, 일조시간)
19~20h : 0.82 % (평균기온)
20~21h : 0.73 % (평균기온, 운량, 일조시간)
21~22h : 0.58 % (최저기온, 최고기온)
22~23h : 0.51 % (평균기온, 운량, 일조시간)
23~0h : 0.58 % (평균기온, 운량, 일조시간, 최고기온)

최종 평균 오차: 0.82 %

```

예상했던 대로, 새벽 시간대에는 전일의 평균 기온과 최저기온이 반영된 모델이 좋은 성능을 보였으며, 낮 시간에는 최고기온과 운량, 일조시간을 고려한 모델이 좋은 성능을 보였다. 또한, 새벽, 아침 시간대에는 최저기온이, 낮 시간에는 일조시간, 운량, 최고기온이 반영되는 추세가 확인되었다. 하루 중 이른 아침 시간이 가장 춥고, 낮 시간이 가장 덥다는 점을 생각하면, 각 시간대별 온도에 맞게, 평균기온과 최고기온, 최저기온, 전일기온이 적절하게 조합되었다고 할 수 있다. 해당일의 평균, 최고, 최고기온은 해당일 전체의 기온 정보인데, 전일 기온 정보와 함께 해당 값들에 적절한 가중치를 부여하면 해당 시간대의 구체적인 기온을 예측할 수 있어, 더 정확한 결과를 얻을수 있는 것으로 보인다. 오버피팅을 막기 위해 이러한 정성적 지식에 기반해서 각 시간대별로 변수를 제거하고, 되도록 인접한 시간대와 통일하겠다. 일례로 4~5시에서 운량과 일조시간이 반영되는데, 아직 해가 뜨지 않은 시간대에 일조시간이 반영되는 것은 적절치 않다.

최종적으로 각 시간대별 예측을 위해 사용되는 변수들을 다음과 같이 정했다.

시간대	평균기온	최저기온	최고기온	전일평균기온	운량	일조시간
0~7h	O	O		O		
8~10h		O		O		
10~11h	O			O		
11~12h	O					

12~16h			O		O	O
16~19h	O					O
19~21h	O					
21~22h		O	O			
22~24h	O				O	O

최종적으로 시간대별 평균 오차 **0.87%**를 얻을 수 있었으며, 이는 문제1의 이동평균법으로 예측했을 때의 오차 **1.69%**보다 유의미하게 개선되었다. 선형회귀식 들은 다음과 같다.

```

0~1h ≈ +54681.0406 +146.4288·평균기온 -238.5109·최저기온 -228.1035·temp_yesterday
1~2h ≈ +51793.3606 +181.8380·평균기온 -268.7609·최저기온 -244.7581·temp_yesterday
2~3h ≈ +50461.8549 +190.9594·평균기온 -292.1150·최저기온 -236.1655·temp_yesterday
3~4h ≈ +50193.4937 +173.2565·평균기온 -301.6061·최저기온 -212.3254·temp_yesterday
4~5h ≈ +51063.3754 +154.6380·평균기온 -282.0995·최저기온 -192.1523·temp_yesterday
5~6h ≈ +52279.4824 +151.9804·평균기온 -300.3749·최저기온 -130.6253·temp_yesterday
6~7h ≈ +53939.5949 +134.8353·평균기온 -336.5942·최저기온 -91.6213·temp_yesterday
7~8h ≈ +57742.8483 -295.5758·평균기온 -29.7513·일조시간 -146.1714·운량
8~9h ≈ +58953.2350 -250.9304·최저기온 -91.3122·temp_yesterday
9~10h ≈ +60768.6717 -243.2319·최저기온 -95.5125·temp_yesterday
10~11h ≈ +62019.2473 -200.7519·평균기온 -105.0999·temp_yesterday
11~12h ≈ +61332.1503 -254.0303·평균기온
12~13h ≈ +61107.0169 -198.0514·최고기온 -317.6155·운량 -332.6834·일조시간
13~14h ≈ +63416.9614 -236.9238·최고기온 -289.2547·운량 -351.4917·일조시간
14~15h ≈ +64183.0378 -248.2176·최고기온 -282.9689·운량 -373.9680·일조시간
15~16h ≈ +63619.5716 -255.3063·최고기온 -255.6956·운량 -380.4848·일조시간
16~17h ≈ +60877.5638 -303.5474·평균기온 -267.5409·일조시간
17~18h ≈ +60701.5276 -316.5233·평균기온 -249.9710·일조시간
18~19h ≈ +60912.1466 -292.4824·평균기온 -134.5351·일조시간
19~20h ≈ +60678.2670 -334.0269·평균기온
20~21h ≈ +58917.5225 -312.4945·평균기온
21~22h ≈ +57424.5286 -152.5257·최저기온 -114.6682·최고기온
22~23h ≈ +61509.2569 -278.1424·평균기온 -225.8740·일조시간 -265.7587·운량
23~0h ≈ +62736.4799 -313.8395·평균기온 -293.4449·일조시간 -320.9390·운량

```

3월 24일의 경우 최고기온 13.8도, 최저기온 4.3도, 평균기온 8.7도, 운량 6.5도, 일조시간 7.4도, 전일 평균기온 5.9도이다. 앞서 구한 선형회귀 식들에 이 값들을 대입하여 다음과 같이 3월 24일의 전력수요를 예측하였다.

	문제 2)
최대전력수요	59,653
최소전력수요	49,151
0~1h	53,584
1~2h	50,776
2~3h	49,474
3~4h	49,151
4~5h	50,062
5~6h	51,539
6~7h	53,125
7~8h	54,001
8~9h	57,335
9~10h	59,159
10~11h	59,653
11~12h	59,122
12~13h	53,848
13~14h	55,666
14~15h	56,151
15~16h	55,619
16~17h	56,257

17~18h	56,098
18~19h	57,372
19~20h	57,772
20~21h	56,199
21~22h	55,186
22~23h	55,691
23~0h	55,748