Week 1

1-1 이동평균법

시계열 분석:

- 하나의 변수에 대한 시간에 따른 관측치의 특성을 요약하고 시간에 따른 패턴 분석과 시간에 따른 패턴을 바탕으로 모형화하고 미래값을 예측하는 것
- 자신의 변수의 과거 패턴이 미래에도 계속된다는 가정하에 변수의 과거값을 바탕으로 미래값 예측
- 시계열 패턴은 수평, 추세, 계절성이 복합된 것으로 간주

이동평균법

- 이동평균 (Moving average): 매 시점에서 직전 N개 데이터의 평균을 산출하여 평활치로 사용
    - 단순이동평균 (Simple Moving Average): 시계열데이터가 수평적 패턴인 경우

$$M_t = \frac{1}{N}(X_{t-N+1} + \cdots + X_t)$$

        - T에서의 이동평균을 사용해서 T+1의 값 예

$$f_{T,1} = M_T$$

        - N이 클수록 평활효과가 큼 (작을수록 최근 데이터 반영)
    - 이중이동평균 (Double moving average): 시계열데이터가 추세 패턴을 따르는 경우

$$X_t = c + bt + a_t$$

        - 단순이동평균 Mt는 추세를 늦게 따라가서 부적합 —> 이중이동평균 활용

$$M_t = \frac{1}{N}(X_{t-N+1} + \cdots + X_t)$$

$$M_t^{(2)} = \frac{1}{N}(M_{t-N+1} + \cdots + M_t)$$

예측

- 시점 T에서 다음 시점의 예측치 (한단계 이후 예측)

$$f_{T,1} = E[X_{T+1}|X_T, X_{T-1}, \ldots] = c + b(T+1)$$

$$\hat{f}_{T,1} = \hat{c} + \hat{b}(T+1) = 2M_T - M_T^{(2)} + \hat{b}$$

$$\hat{b} = \frac{2}{N-1}(M_T - M_T^{(2)})$$

- k-단계 이후 예측치

$$f_{T,k} = E[X_{T+k}|X_T, X_{T-1}, \ldots] = c + b(T+k),\ k = 1,2, \ldots$$

$$\hat{f}_{T,k} = \hat{c} + \hat{b}(T+k) = 2M_T - M_T^{(2)} + k\hat{b}$$

- 이중 이동평균법으로 예측하는 방법

예측성능 척도

1. 예측 오차
    a. 실제값과 예측값을 비교해서 예측 오찰를 산출
    b. 총 n개 시정에서 예측 오차를 산출하는 경우 :
        i. RMSE는 단위 일치

$$평균제곱오차: MSE = \frac{1}{n}\sum_{t=1}^{n} e_{t,1}^2$$

$$제곱근\ 평균제곱오차: RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_{t,1}^2}$$

$$평균절대오차: MAD = \frac{1}{n}\sum_{t=1}^{n} |e_{t,1}|$$

$$평균절대\ 퍼센트오차: MAPE = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{e_{t,1}}{X_{t+1}}\right|$$

지수평활법

- 전체 데이터를 사용하고 시간에 따른 가중치를 주면서 평활치 산출 (과거로 갈수록 지수적으로 감소하는 가중치)
- 단순 지수평활 - 시계열 데이터가 수평적 패턴인 경우 사용
- 이중 지수평활 - 시계열 데이터가 추세 패턴을 따르는 경우 사용
- 홀트 모형 - 시계열 데이터가 추세 패턴을 따르는 경우 사용

단순 지수평활

$$S_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \cdots$$

- 평활상수 α가 작을수록 평활효과가 큼 (stable)
- 시점 t에서의 시점 t+1의 값 예측은 t에서의 단순 지수평활

$$S_{t+1} = \alpha X_{t+1} + (1-\alpha)S_t$$

이중 지수평활법

$$X_t = c + bt + a_t$$

- 절편 (c)와 기울기 (b) 추정해야 하기 때문에 단순 지수평활치의 기대치와 시계열 기대치간의 격차가 존재 → 이중 지수평활을 활

$$S_t = \alpha X_t + (1-\alpha)S_{t-1}$$
$$S_t^{(2)} = \alpha S_t + (1-\alpha)S_{t-1}^{(2)}$$

예측

- 시점 T에서 다음 시점의 예측치 (한단계 이후 예측)

$$f_{T,1} = E[X_{T+1}|X_T, X_{T-1}, \ldots] = c + b(T+1)$$
$$\hat{f}_{T,1} = \hat{c} + \hat{b}(T+1) = 2S_T - S_T^{(2)} + \hat{b}$$
$$\hat{b} = \frac{\alpha}{1-\alpha}(S_T - S_T^{(2)})$$

- k-단계 이후 예측치

$$f_{T,k} = E[X_{T+k}|X_T, X_{T-1}, \ldots] = c + b(T+k), \ k = 1,2,\ldots$$
$$\hat{f}_{T,k} = \hat{c} + \hat{b}(T+k) = 2S_T - S_T^{(2)} + k\hat{b}$$

홀트 모형

- 수평 수준과 추세를 각각 갱신하는 모형
  - 수평수준: $L_t = \alpha X_t + (1-\alpha)(L_{t-1} + b_{t-1})$, $(0<\alpha<1)$
  - 추세: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$, $(0<\beta<1)$

T+k 예측:

$$f_{T,k} = L_T + kb_T, \qquad k = 1,2,\ldots$$

계절성 고려 모형:

- 윈터스 모형 - 홀트 모형에 계절성을 추가반영
  - 가법모형과 승법모형
- 분해법 - 추세와 계절성을 분해한 후 예측시 다시 결합
  - 가법모형과 승법모형

윈터스 모형:

- st = 계절성 지수 - t = 1,2,3,…m
- m: 계절 주기, 분기별 데이터의 경우 m = 4

- 수평수준: $L_t = \alpha \frac{X_t}{s_{t-m}} + (1-\alpha)(L_{t-1} + b_{t-1}), \quad (0<\alpha<1)$
- 추세: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}, \quad (0<\beta<1)$
- 계절성: $s_t = \gamma \frac{X_t}{L_t} + (1-\gamma)s_{t-m}, \quad (0<\gamma<1)$

초기치들이 필요하고, 계절성지수는 평균 1이 되도록 조정이 필요하다

예측:

$$f_{T,k} = (L_T + kb_T))s_{T-m+k}, \qquad k = 1,2,\ldots$$

분해법:

분해법에 의한 예측 절차

1. 중심 이동평균으로 평활치 산출

2. 추세제거 시계열 산출
3. 계절성 지수 산출
4. 계절성 제거 시계열 산출
5. 회귀모형으로 추세 추정
6. 추세 및 계절성지수를 결합하여 예측치 산출

- 승법적 모형

m=12
q=6

- 중심이동평균 (m개의 데이터 이용)

$$CM_t = \begin{cases} \frac{1}{m}\left(0.5X_{t-q} + X_{t-q+1} + \cdots + X_{t+q+1} + 0.5X_{t+q}\right) & m = 2q \text{ (주기 짝수)} \\ \frac{1}{m}\left(X_{t-q} + X_{t-q+1} + \cdots + X_{t+q}\right) & m = 2q+1 \text{ (주기 홀수)} \end{cases}$$

- 추세제거 시계열

$$DX_t^{(T)} = \frac{X_t}{CM_t}$$

- 계절성지수

계절별 추세제거 시계열값의 평균으로 계절성 지수 $s_i, i = 1, \ldots, m$ 산출,
이때 계절성지수 평균이 1이 되도록

- 계절성제거 시계열

$$DX_t^{(S)} = \frac{X_t}{s_t}$$