LEC002 Demand Forecasting

VG441 SS2021

Cong Shi
Industrial & Operations Engineering
University of Michigan

VG441 TA

Mr. Shunyi Zhu

- TA office hours: Wednesday Night 8:00-9:30pm
- Location: 326D
- Background: research on optimization algorithms in crowdsourcing and energy-efficient network node deployment strategies.
- Hobbies: Crayon Shin-chan

Properties

- ***Trend***: A long-term increase or decrease in the data. This can be seen as a slope (doesn't have to be linear) roughly going through the data.

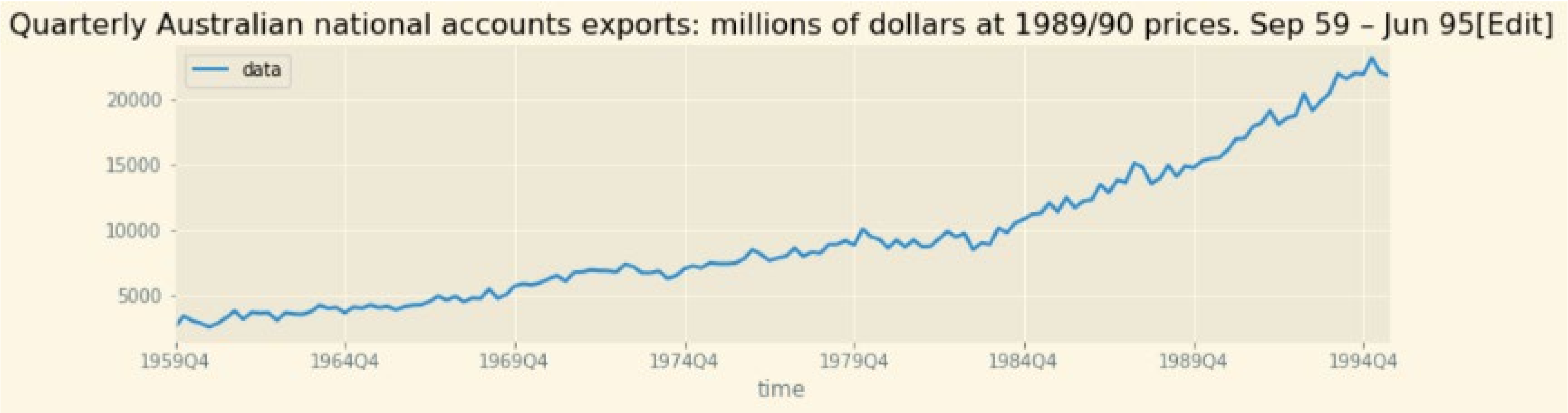
- ***Seasonality***: A time series is said to be seasonal when it is affected by seasonal factors (hour of day, week, month, year, etc.). Seasonality can be observed with nice cyclical patterns of fixed frequency.

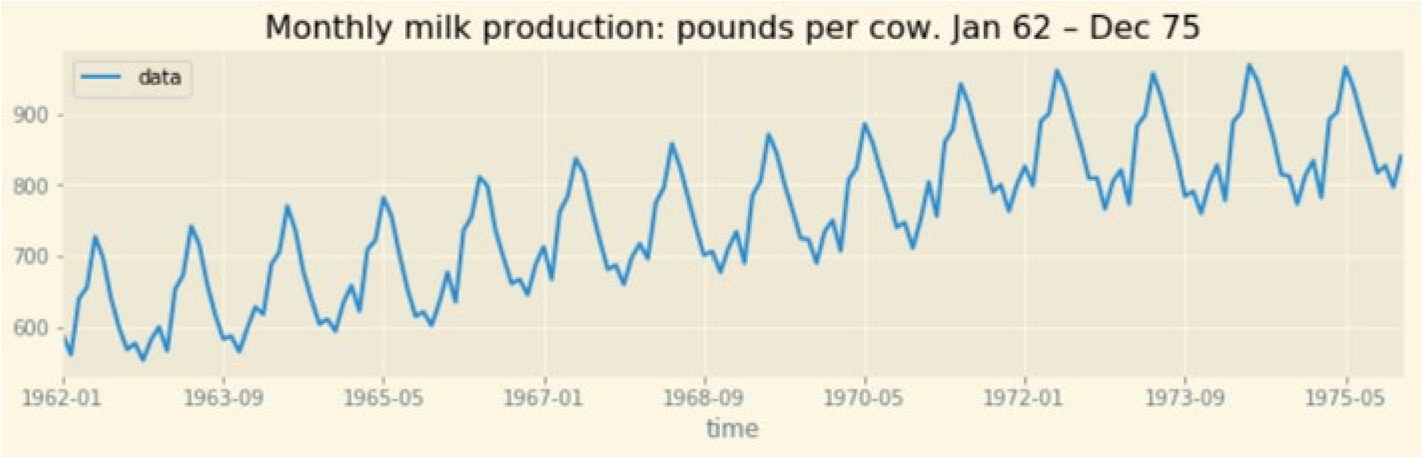
- ***Cyclicity***: A cycle occurs when the data exhibits rises and falls that are not of a fixed frequency. These fluctuations are usually due to economic conditions, and are often related to the "business cycle". The duration of these fluctuations is usually at least 2 years.

- ***Residuals***: Each time series can be decomposed in two parts:
  - A forecast, made up of one or several *forecasted* values
  - Residuals. They are the difference between an observation and its predicted value at each time step. Remember that

Value of series at time t = Predicted value at time t + Residual at time t

Examples


Quarterly Australian national accounts exports: millions of dollars at 1989/90 prices. Sep 59 – Jun 95[Edit]
data
20000
15000
10000
5000
1959Q4
1964Q4
1969Q4
1974Q4
1979Q4
1984Q4
1989Q4
1994Q4
time


Monthly milk production: pounds per cow. Jan 62 – Dec 75
data
900
800
700
600
1962-01
1963-09
1965-05
1967-01
1968-09
1970-05
1972-01
1973-09
1975-05
time

Examples

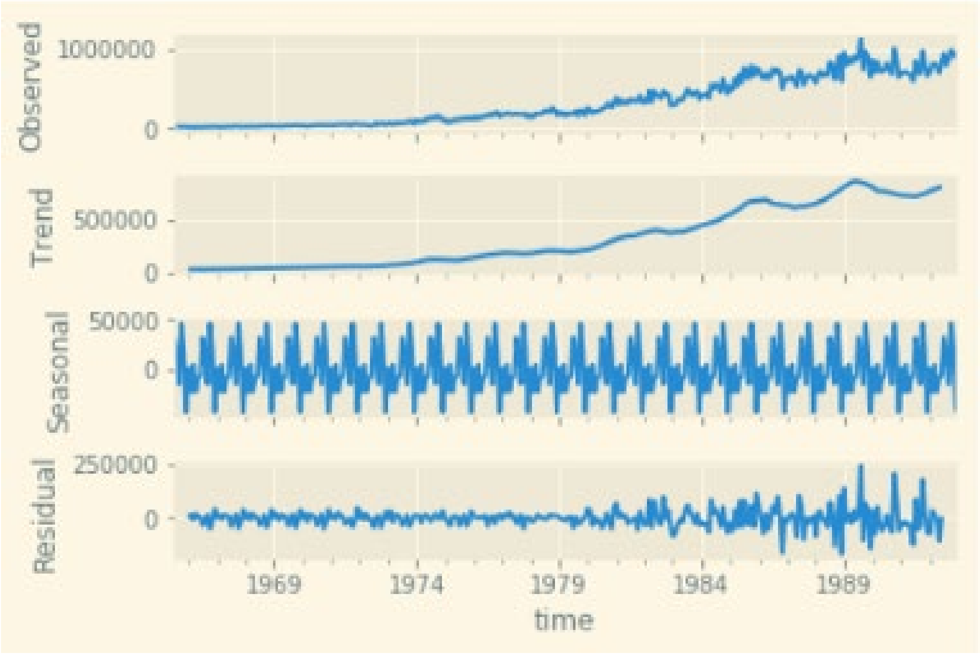

Industrial production, Spain: monthly (78-82)
data
180
160
140
120
100
1977-12
1978-10
1979-08
1980-06
1981-04
1982-02
1982-12
time


IBM common stock closing prices
data
600
550
500
450
400
350
300
1962-01-02
1962-10-16
1963-08-02
1964-05-20
1965-03-08
1965-12-21
time

Decomposition of Time Series

Each time series can be thought as a mix between several parts :

- A trend (upward or downwards movement)
- A seasonal component
- Residuals



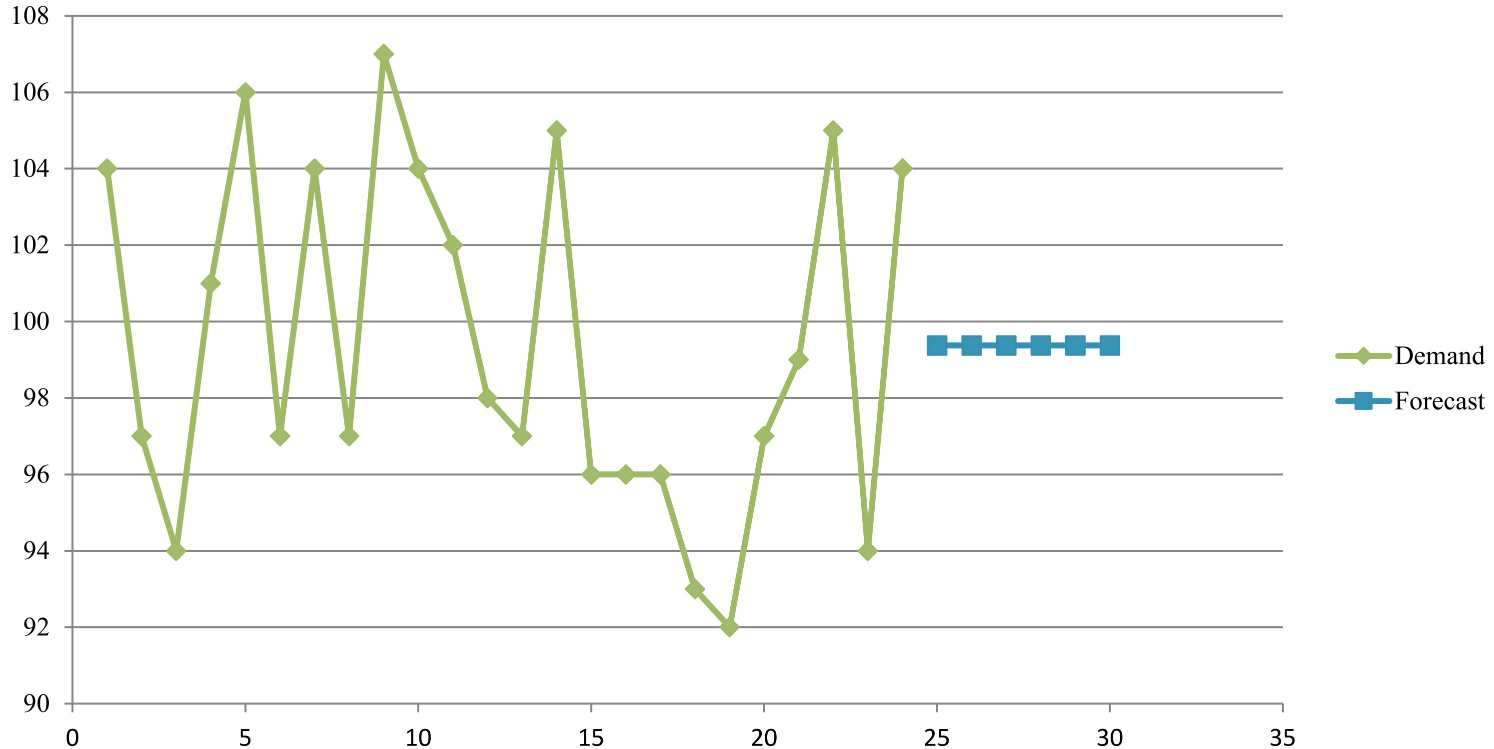
Simple Average

- Stationary model $D_t = I + \epsilon_t$
- Static forecast $\hat{y} = \frac{\sum_{t=1}^{N} D_t}{N}$
- Derived based on minimizing MSE

$$e_t = D_t - \hat{y}$$

$$\frac{d\left(\sum_{t=1}^{N} e_t^2\right)}{d\hat{y}} = \frac{d\left[\sum_{t=1}^{N} (d_t - \hat{y})^2\right]}{d\hat{y}} = -2\sum_{t=1}^{N} (d_t - \hat{y}) = 0$$

Simple Average Model



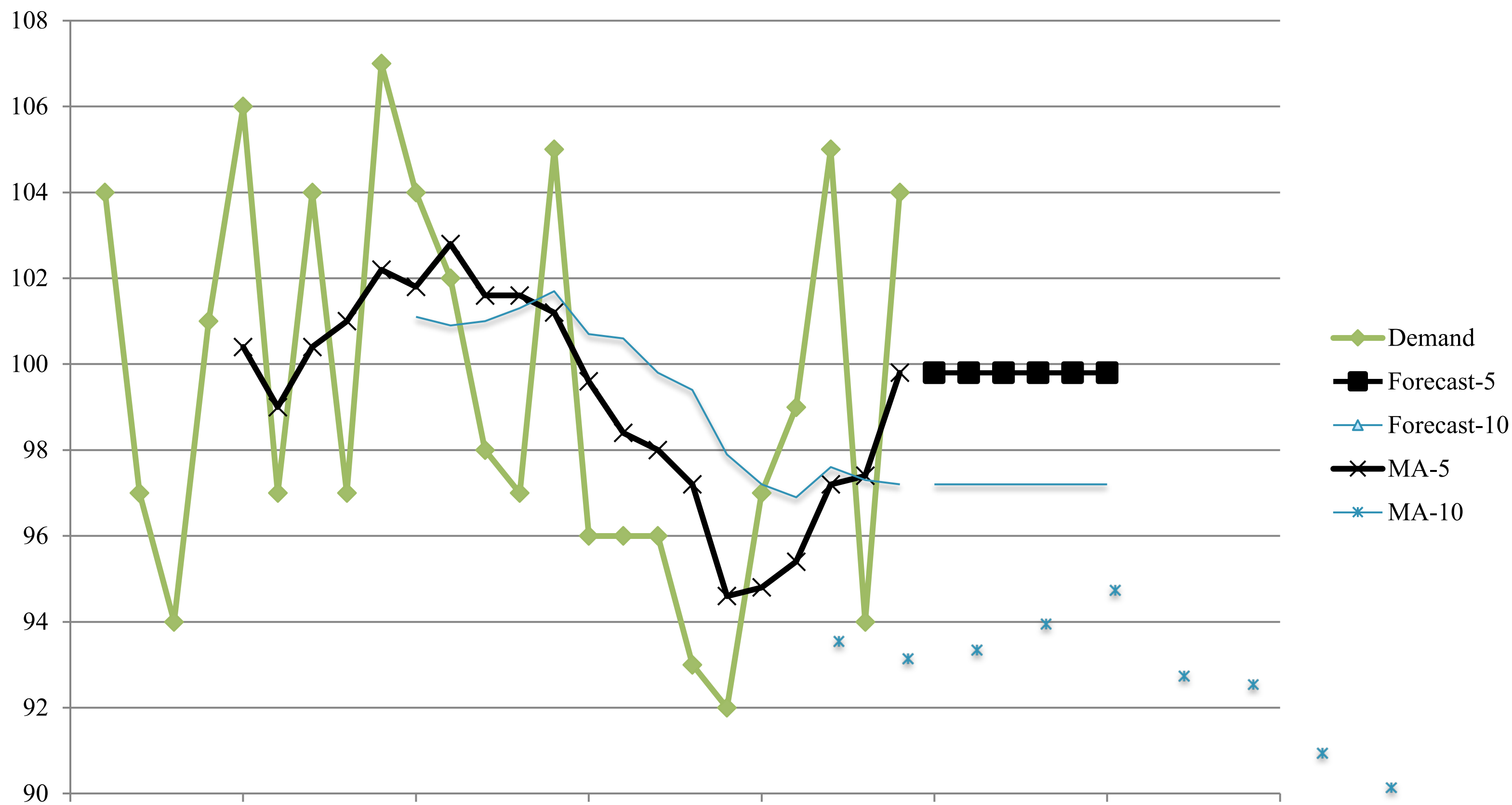
Moving Average (MA)

- Average only the most recent data points

$$y_t = \frac{1}{N} \sum_{i=t-N}^{t-1} D_i$$

- Smooth out noise
- Can respond to change in process

Moving Average (MA)



Weighted Moving Average

- A generalization of MA with weights

$$y_t = \frac{\sum_{i=t-N}^{t-1} w_i D_i}{\sum_{i=t-N}^{t-1} w_i}$$

e.g.,

$$w_{t-1} = N, w_{t-2} = N - 1, \ldots, w_{t-N} = 1$$

Exponential Smoothing

- Adjust forecast based on the recent data point

$$y_t = \alpha D_{t-1} + (1-\alpha) y_{t-1}$$

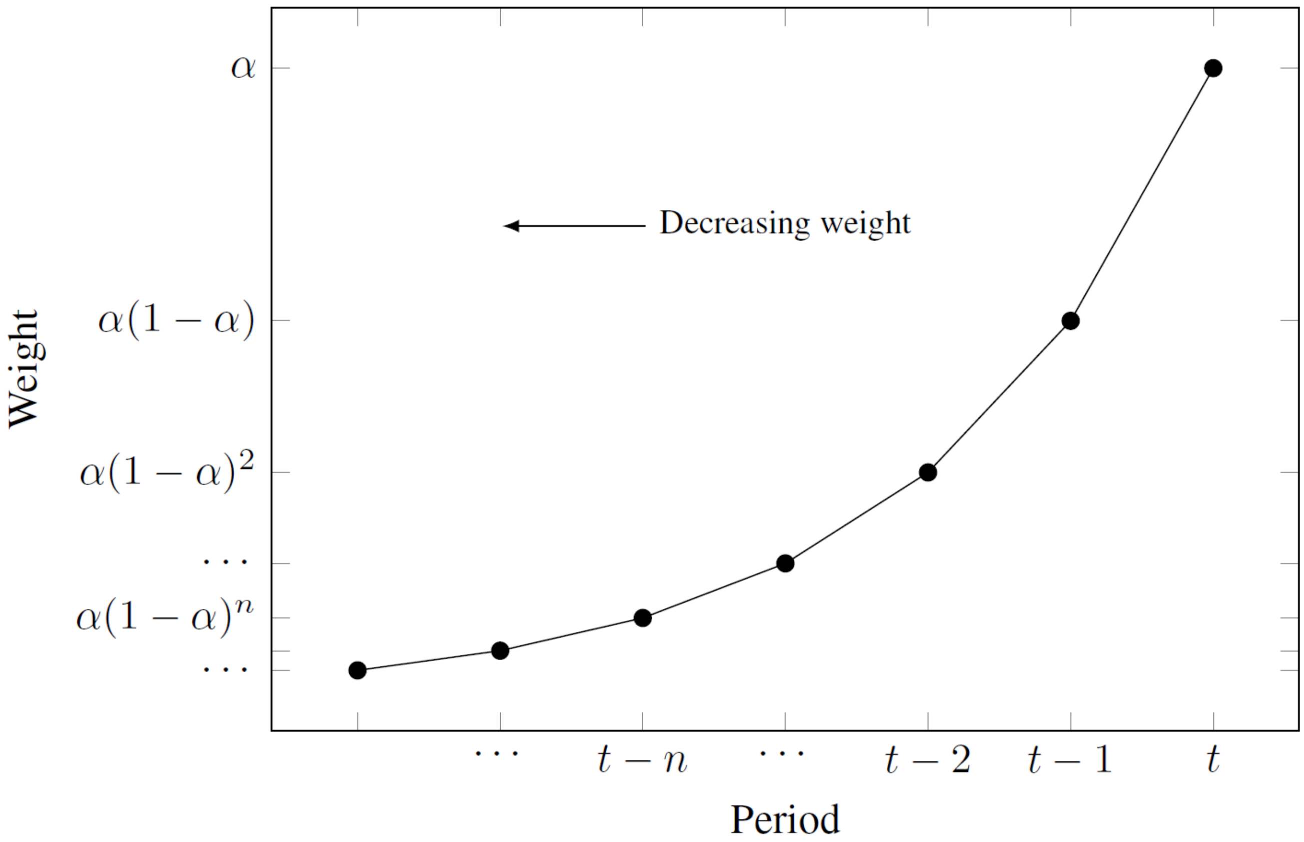
- It is a weighted average of all historical data points, with the weight decreasing exponentially with age

$$y_{t-1} = \alpha D_{t-2} + (1-\alpha) y_{t-2}$$

$$y_t = \alpha D_{t-1} + \alpha(1-\alpha) D_{t-2} + (1-\alpha)^2 y_{t-2}$$

$$y_t = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i D_{t-i-1} = \sum_{i=0}^{\infty} \alpha_i D_{t-i-1}$$

Exponential Smoothing



Double Exponential Smoothing (Holt)

Double exponential smoothing can be used to forecast demands with a linear trend

$$D_t = I + tS + \epsilon_t$$

The predictor consists of base and slope:

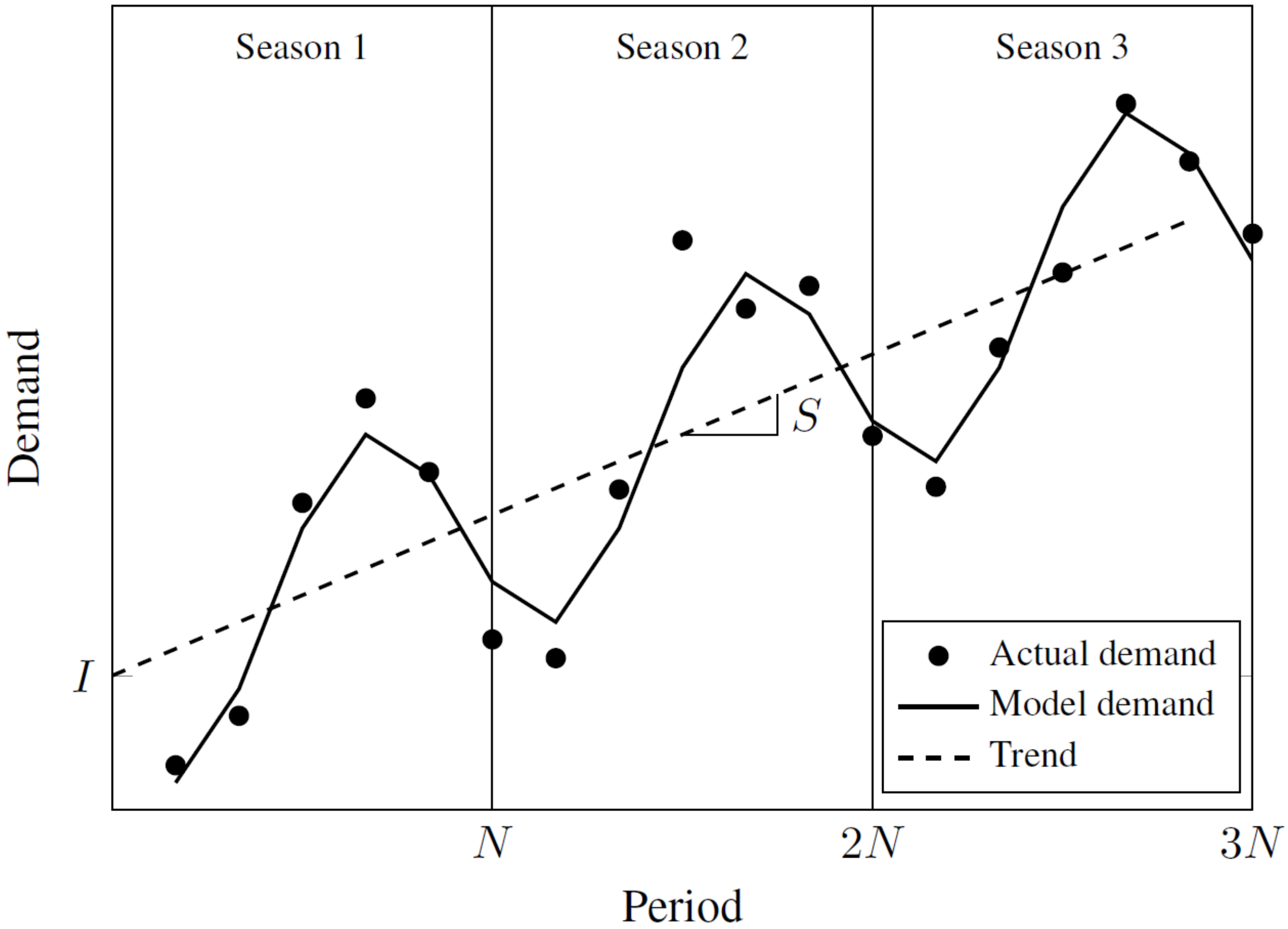
$$y_{t+1} = I_t + S_t$$

$$\begin{aligned} I_t &= \alpha D_t + (1-\alpha)\left(I_{t-1} + S_{t-1}\right) \\ S_t &= \beta\left(I_t - I_{t-1}\right) + (1-\beta)S_{t-1} \end{aligned}$$

Alpha is the smoothing constant and Beta is the trend constant

Triple Exponential Smoothing (Holt-Winters)

- Random demands with trend and seasonality



Triple Exponential Smoothing (Holt-Winters)

Demand model

$$D_t = (I + tS)c_t + \epsilon_t \qquad \sum c_t = N$$

The predictor

$$y_{t+1} = (I_t + S_t)\, c_{t+1-N}$$

Basic idea is to “de-trend” and “de-seasonalize”

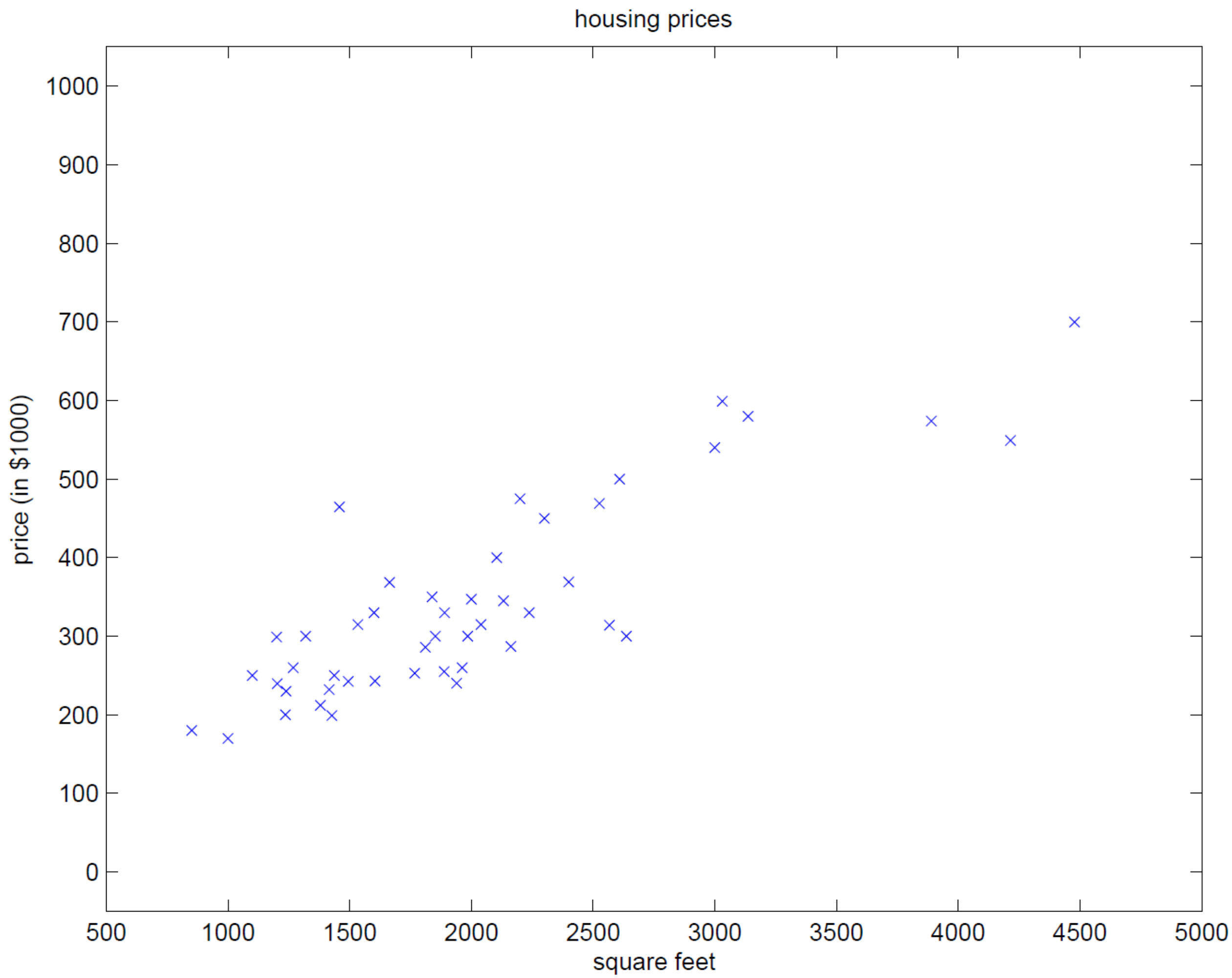
$$\begin{aligned} I_t &= \alpha \frac{D_t}{c_{t-N}} + (1-\alpha)\left(I_{t-1} + S_{t-1}\right) \\ S_t &= \beta\left(I_t - I_{t-1}\right) + (1-\beta)S_{t-1} \\ c_t &= \gamma \frac{D_t}{I_t} + (1-\gamma)c_{t-N} \end{aligned}$$

Python Time!

- statsmodels.tsa.seasonal
- statsmodels.tsa.holtwinters

Linear Regression

- "Best fitting line"



Linear Regression

- (Linear) Hypothesis Function:

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

- Minimizing the least-squares cost:

$$J\left(\theta_{0\ldots n}\right) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Matrix Derivation

- Turn everything into matrix notation

$$h_\theta(x) = \theta^T x \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_n \end{pmatrix} \in \mathbb{R}^{n+1} \qquad x = \begin{pmatrix} x_0 \triangleq 1 \\ x_1 \\ \dots \\ x_n \end{pmatrix} \in \mathbb{R}^{n+1}$$

- Design matrix

$$X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m-1)} \\ x^{(m)} \end{bmatrix} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & \dots & \dots & x_n^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(m)} & \dots & \dots & \dots & x_n^{(m)} \end{bmatrix}$$

Normal Equation

- Cost function

$$J(\theta) = \frac{1}{2m}(X\theta - y)^T(X\theta - y)$$

- Throwing out the constant

$$J(\theta) = \theta^T X^T X\theta - 2(X\theta)^T y + y^T y$$

- The "famous" normal equation

$$\frac{\partial J}{\partial \theta} = 2X^T X\theta - 2X^T y = 0$$

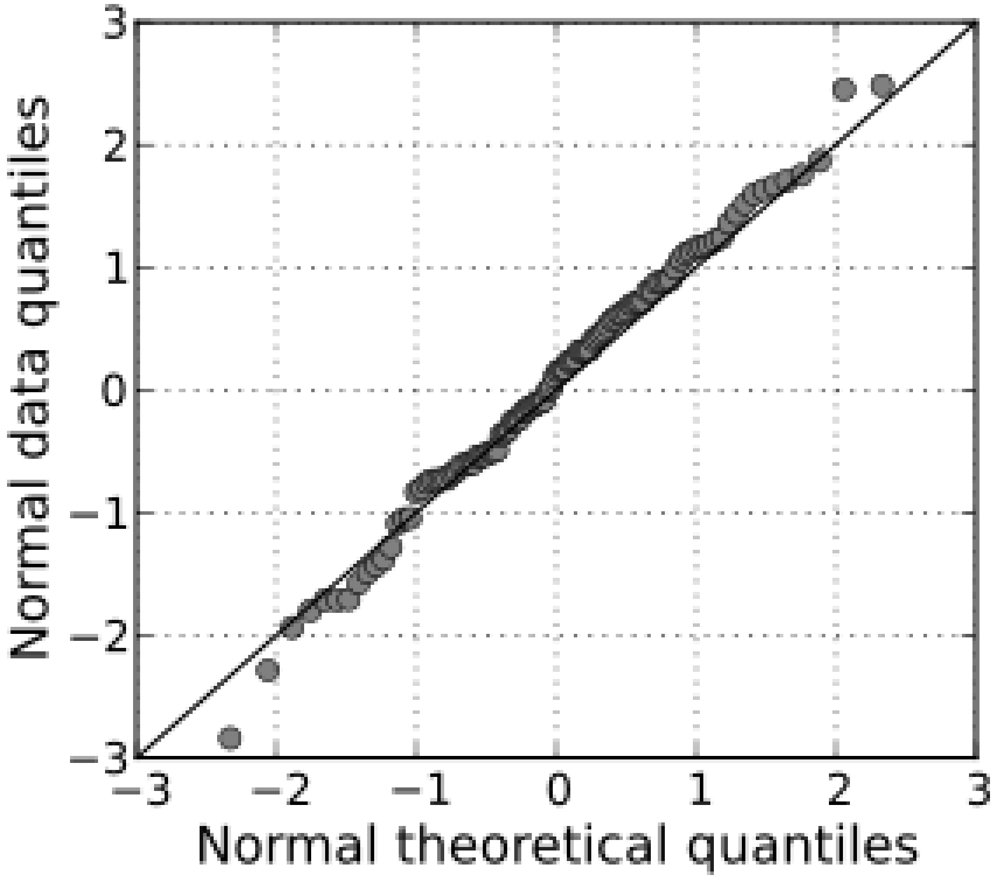
$$X^T X\theta = X^T y$$

$$\theta = \left(X^T X\right)^{-1} X^T y$$

Residuals

- The residual should be normally distributed

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$



Probabilistic Interpretation

- Deriving the log-likilihood function

$$\epsilon^{(i)} = y^{(i)} - \theta^T x^{(i)} \sim \mathcal{N}\left(0, \sigma^2\right)$$

$$p\left(\epsilon^{(i)}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(\epsilon^{(i)}\right)^2}{2\sigma^2}\right)$$

$$p\left(y^{(i)} | x^{(i)}; \theta\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

Probabilistic Interpretation

- Deriving the log-likilihood function

$$L(\theta) = L(\theta; X, \vec{y}) = p(\vec{y}|X; \theta)$$

$$L(\theta) = \prod_{i=1}^{m} p\left(y^{(i)}|x^{(i)}; \theta\right)$$
$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right)$$

Probabilistic Interpretation

- Maximum Likelihood Estimator (MLE)

$$
\begin{aligned}
\ell(\theta) &= \log L(\theta) \\
&= \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right) \\
&= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y^{(i)} - \theta^T x^{(i)}\right)^2}{2\sigma^2}\right) \\
&= m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^{m} \left(y^{(i)} - \theta^T x^{(i)}\right)^2
\end{aligned}
$$

Python Time!

- from sklearn import linear_model
- from statsmodels.api import OLS