Berkeley Industrial Engineering & Operations Research COLLEGE OF ENGINEERING

LEC-2a Demand Forecast I

9/3/2024

INDENG 250 2024 Fall
Introduction to Production Planning and Logistics Models
University of California, Berkeley

Huiwen Jia
Assistant Professor
Industrial Engineering & Operations Research

Demand Forecast

Problem Model Algorithm Analysis

- Forecast the volume of demand of one or more future periods.
- Determine the forecast horizon.
- Collect data. Factors.
- …

Demand Forecast

Problem **Model Algorithm** Analysis

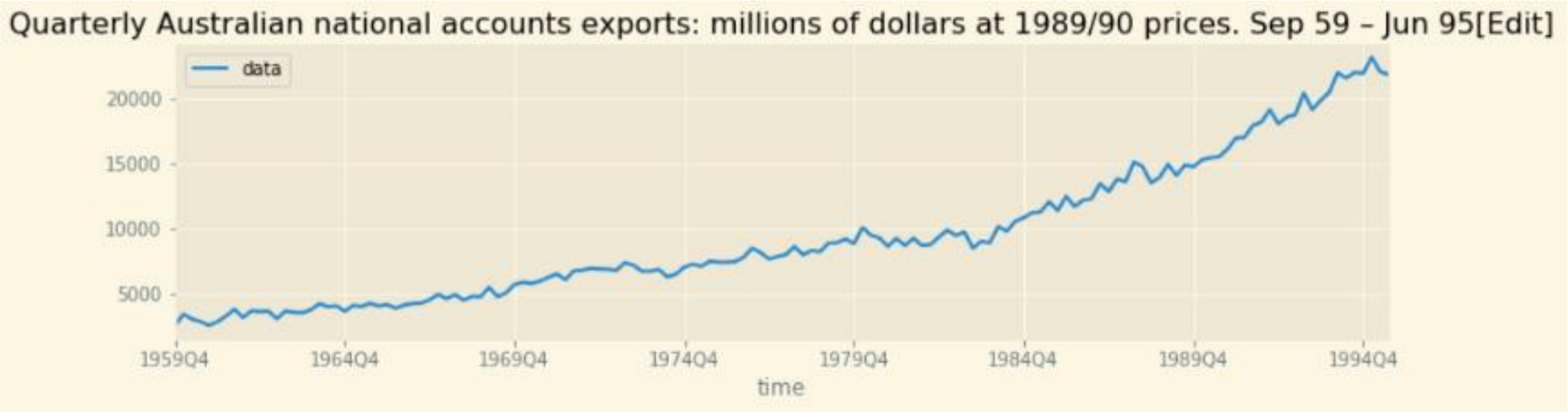
Time Series Models

- Simple average
- Moving average
- Weighted moving average
- Exponential smoothing
- Double exponential smoothing
- Triple exponential smoothing
- ARIMA
- …

General Supervised Learning Models

- Linear regression
- Gradient boosting
- …

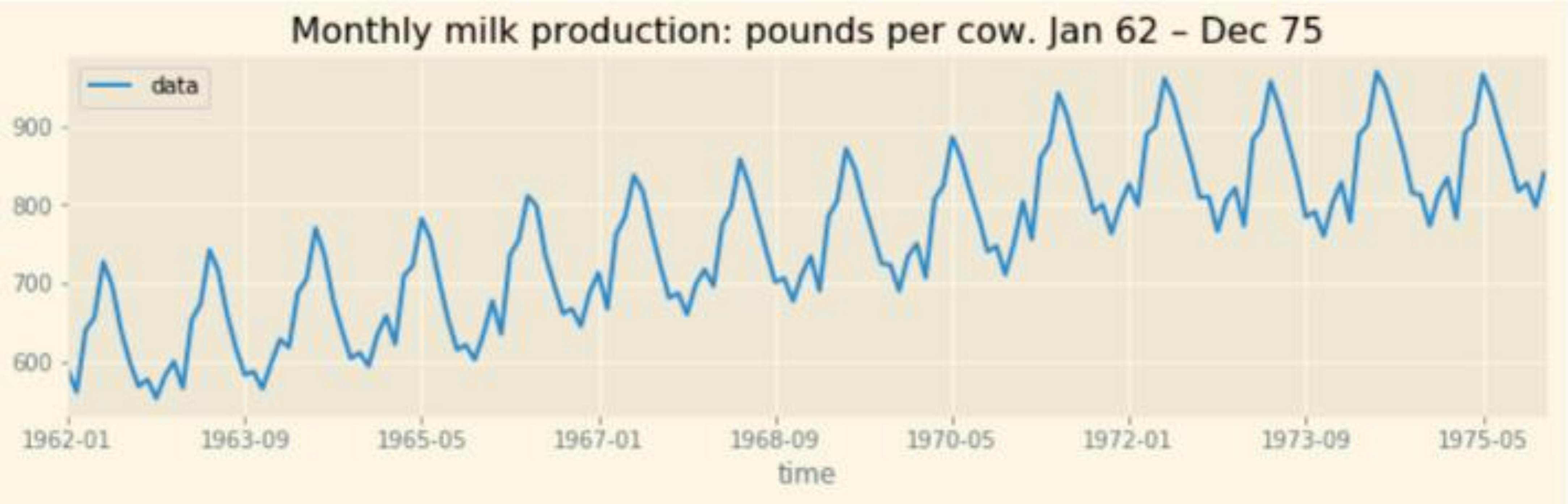
Examples:



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.

Examples:



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.

Examples:

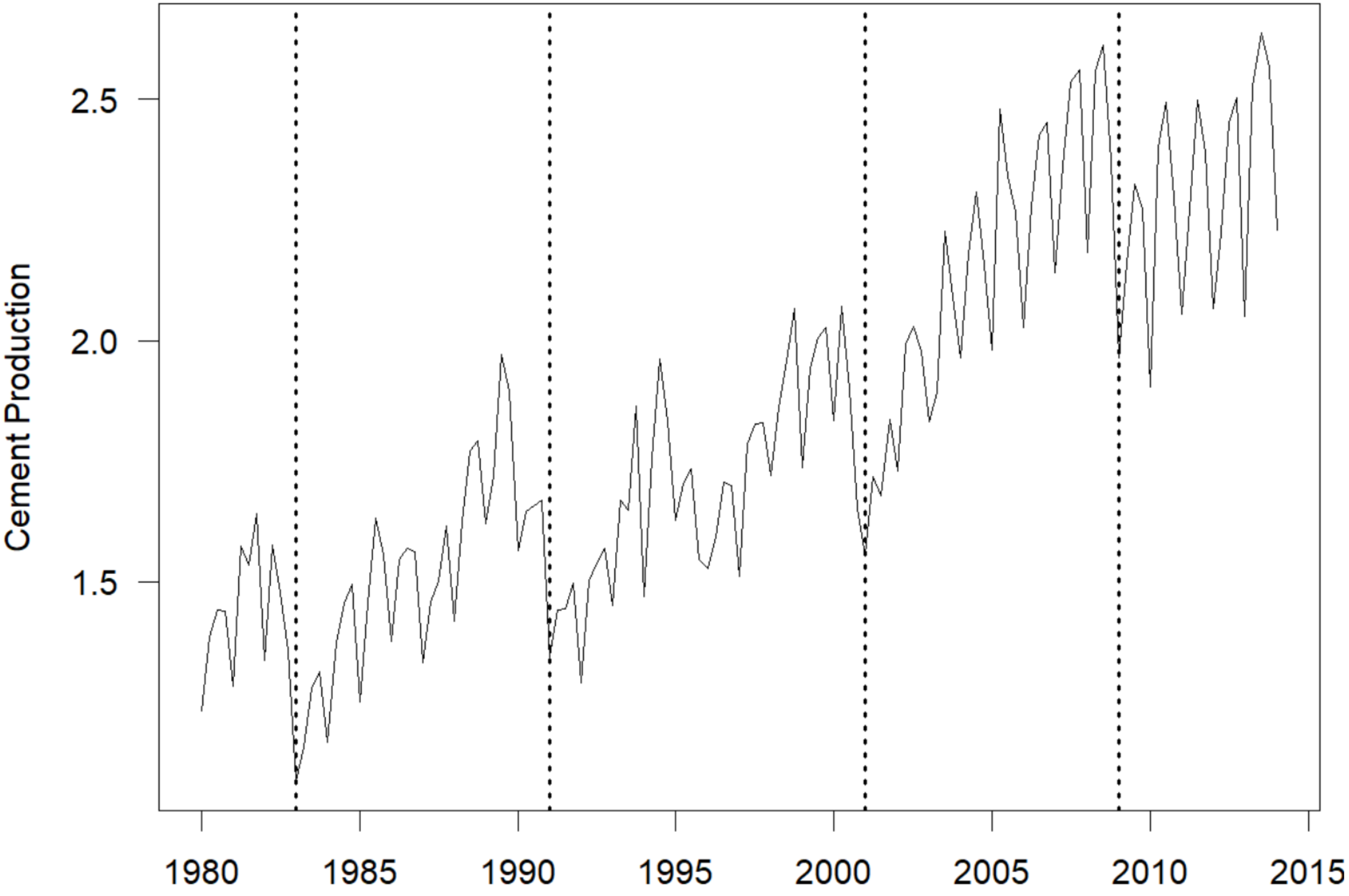
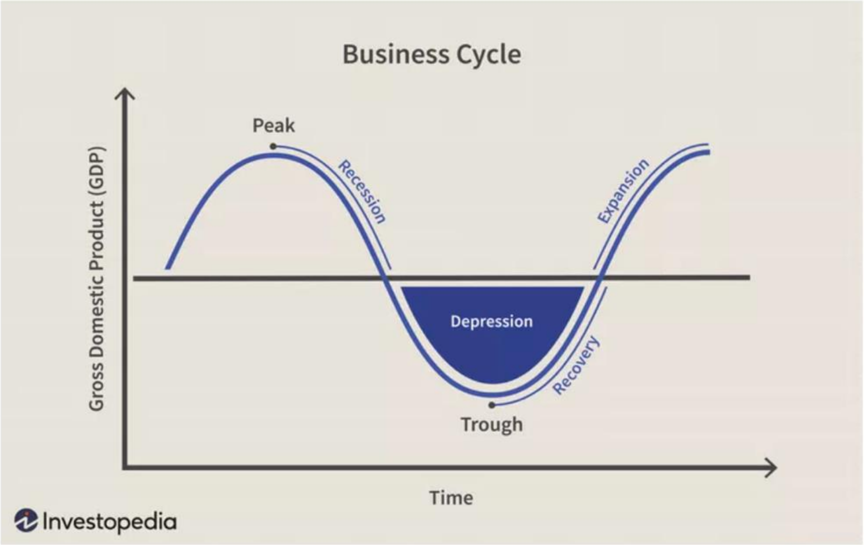


Figure 6.6: Cement production in Australia (millions of tonnes). Business cycles in the early 1980s, 1990s, 2000s, and around 2008 are demarcated by vertical dotted lines



Business Cycle Graph. Image by Julie Bang © Investopedia 2019

Example from: https://dfep.netlify.app/sec-tsfeatures

Examples:

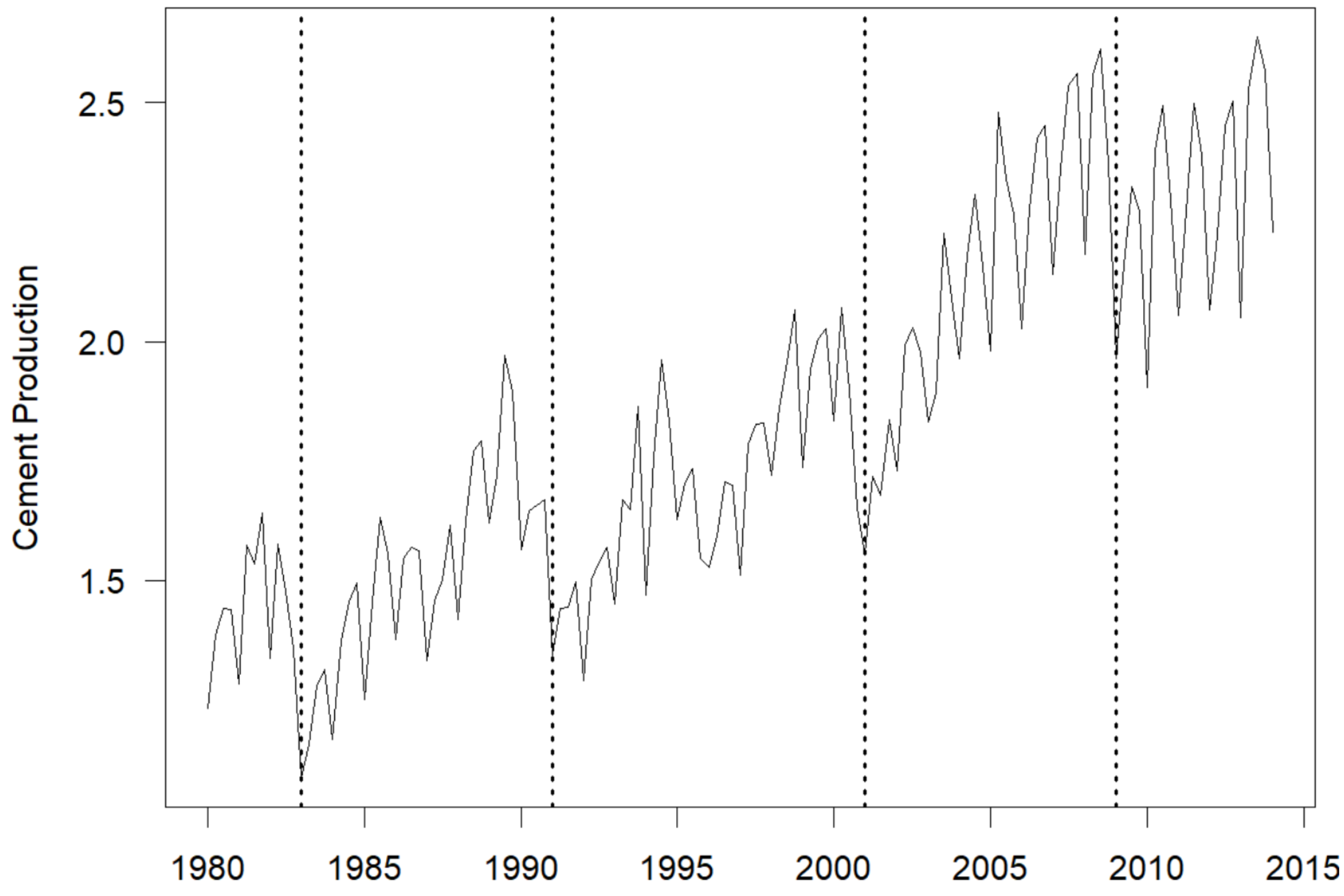


Figure 6.6: Cement production in Australia (millions of tonnes). Business cycles in the early 1980s, 1990s, 2000s, and around 2008 are demarcated by vertical dotted lines

Example from: https://dfep.netlify.app/sec-tsfeatures

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.

Compotents

- *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.

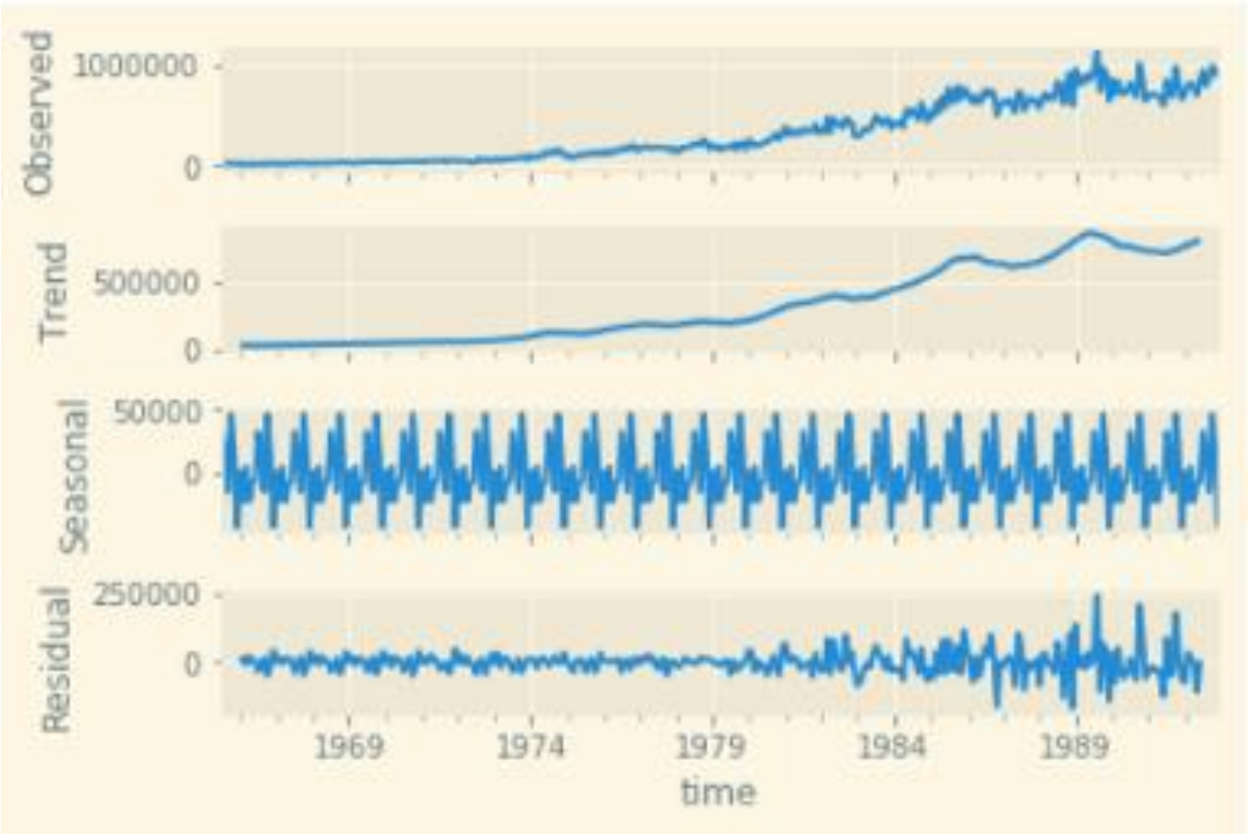
- *Residuals*: Each time series can be decomposed in two parts:
  - A forecast, made up of one or several *forecasted* values
  - Residuals. They are the difference between an observation and its predicted value at each time step.
  - Remember that

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

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



Time Series Models

Time Series Models

- Simple average
- Moving average
- Weighted moving average
- Exponential smoothing
- Double exponential smoothing
- Triple exponential smoothing

Parameters & Variables:

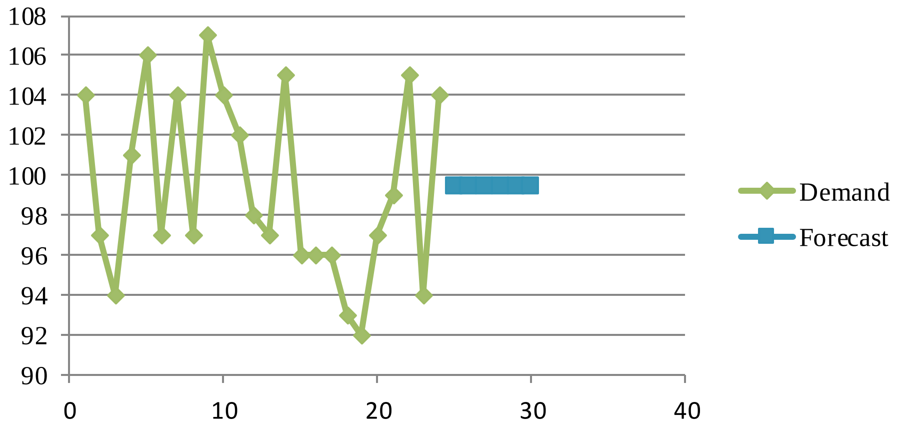
- Time periods: $t = 1, 2, \ldots, N, N+1, \ldots$
- Historical actual demand in period t: D_t
- The forecast of demand in period t: y_t
- Forecast horizon h: h-step-ahead forecast, i.e., forecast period $t+h$ at the end of period t.
- For this lecture, we mainly focus on 1-step-ahead forecast, i.e., forecast period t at the end of period $t-1$.

One basic rule for scientific writing: define notations before referring to them.

- Historical demand: D_t
- Forecast: y_t

Simple Average

- Stationary model (for all t) $D_t = I + \epsilon_t$
- Static forecast $y_t = \dfrac{\sum_{i=1}^{t-1} D_i}{t-1}$
- Derived based on minimizing MSE over all historical data points



*MSE - Mean Squared Error

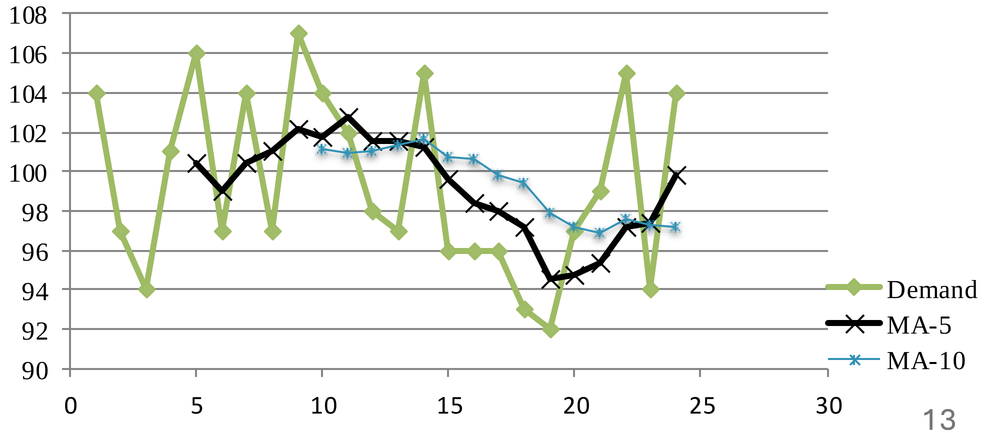
- Historical demand: D_t
- Forecast: y_t

Moving Average (MA)

- Average only the most recent data points

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

- Known as *simple moving average forecast of* ***order*** *N*.
- Term "moving" means the window of data points moves
- Can respond to change in process



Weighted Moving Average

- Historical demand: D_t
- Forecast: y_t

- A generalization of MA with weights

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

Any systematic way to determine the weights?

e.g.,

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

The weight of D_i is $w_i = \frac{c_i}{\sum_{i=t-N}^{t-1} c_i}$ and $\sum_{i=t-N}^{t-1} w_i = 1$.

Exponential Smoothing (Single)

- Historical demand: D_t
- Forecast: y_t

- Adjust forecast based on the recent data point

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

Smoothing factor α : $0 < \alpha < 1$

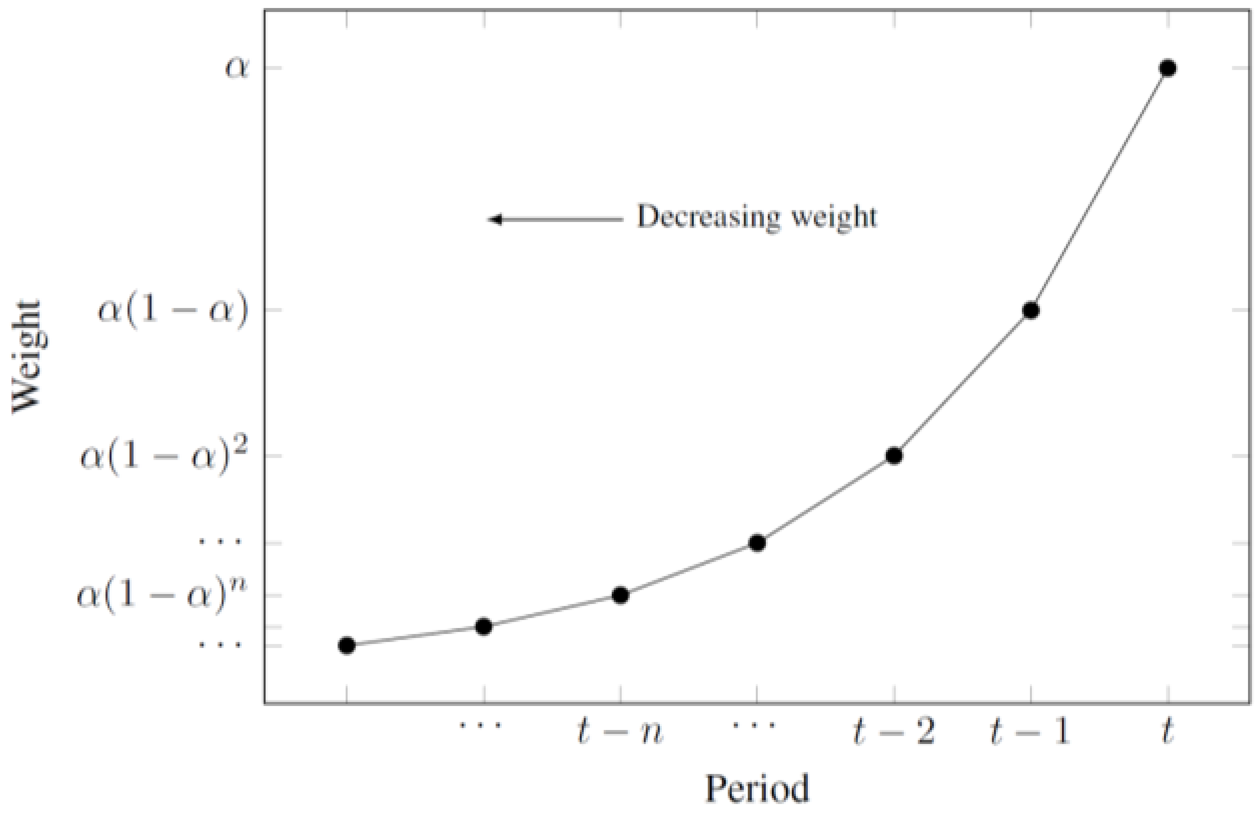
$$\boxed{y_{t-1} = \alpha D_{t-2} + (1-\alpha) y_{t-2}}$$

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

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

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

Exponential Smoothing (Single)



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

- The weight of D_{t-i-1} is $\alpha_i = \alpha(1-\alpha)^i$
- Sum over all weights $\sum_{i=0}^{\infty} \alpha_i = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i = 1$ (Deducted from Geometric series property)

$$\boxed{\text{Smoothing factor } \alpha : \ 0 < \alpha < 1}$$

- Weight α_i can be approximated by an exponential function $f(i) = \alpha e^{-\alpha i} \to$ *Exponential smoothing*

Double Exponential Smoothing

- Historical demand: D_t
- Forecast: y_t

Double exponential smoothing can be used to forecast demands with a linear trend (non-stationary)
Demand model (base demand + slope · time length):

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

The predictor consists of base and slope:

Period t

Observe $D_t \rightarrow$ update $I_t, S_t \rightarrow$ make forecast y_{t+1}

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

$$I_t = \alpha D_t + (1-\alpha)\left(I_{t-1} + S_{t-1}\right)$$

$$S_t = \beta\left(I_t - I_{t-1}\right) + (1-\beta) S_{t-1}$$

Update previous forecast with recent observation.

$\boxed{\alpha \text{ is the smoothing constant and } \beta \text{ is the trend constant}}$

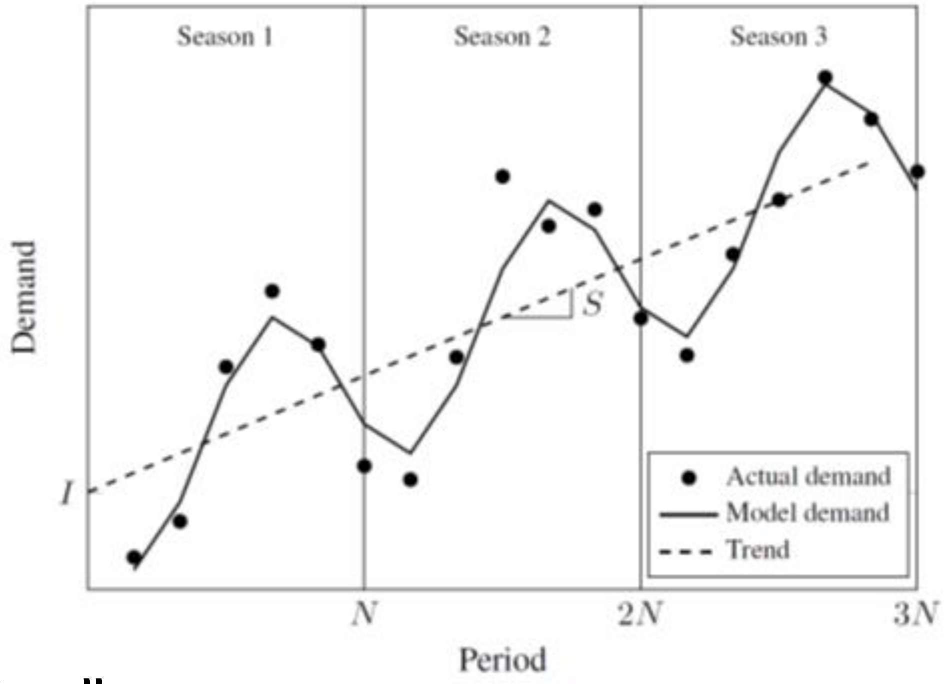
Triple Exponential Smoothing

Demand model

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

The predictor

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



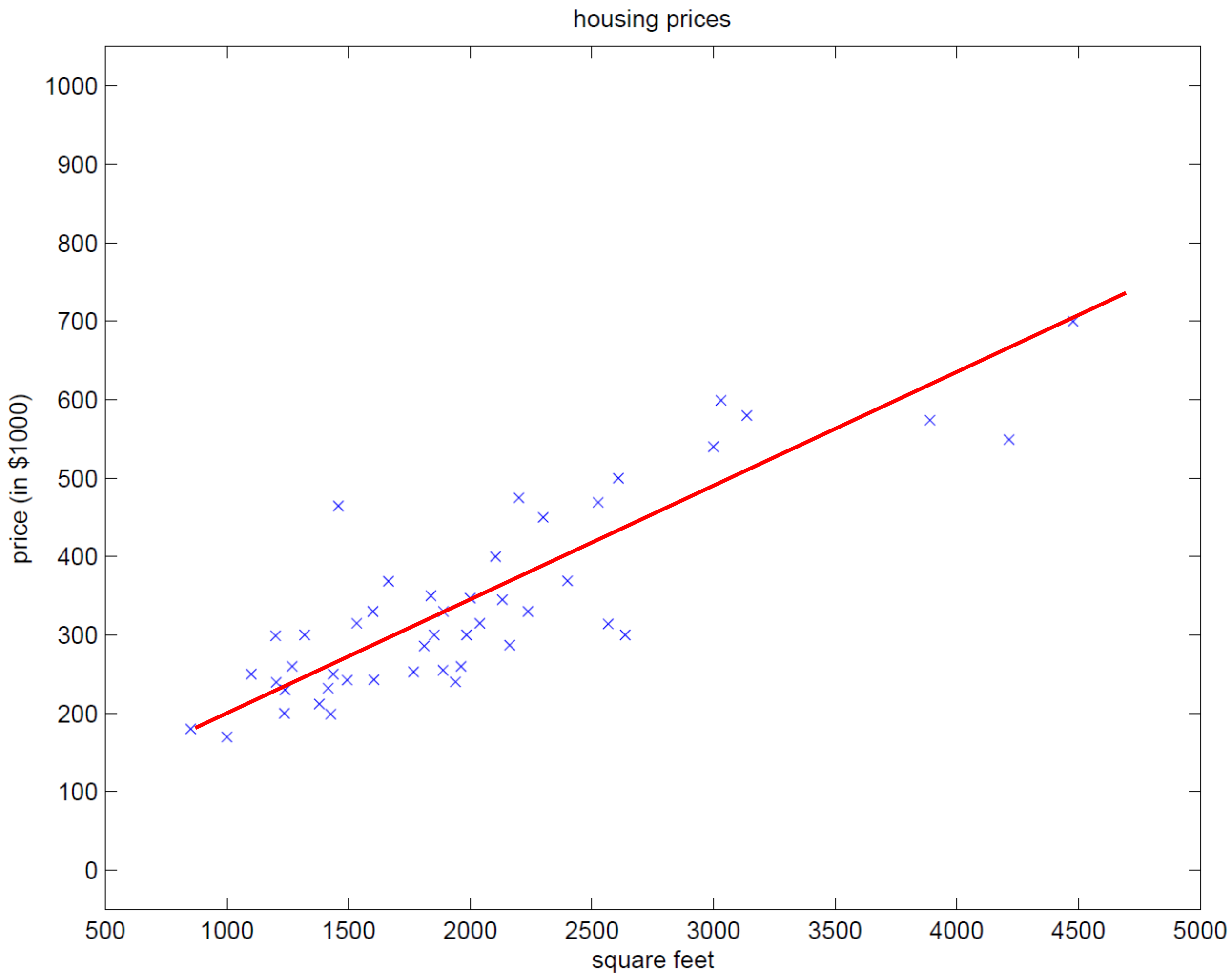
Basic idea is to “de-trend” and “de-seasonalize”

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

Supervised Learning Model

Linear Regression

- “Best fitting line”
- Example
  - One factor



Linear Regression

- For any given $(\boldsymbol{x}, y)$

$$y = h_\theta(x) + \epsilon$$

- (Linear) Hypothesis Function:

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

- Least-square estimation for θ:

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

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

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

m data points

n factors

LSE - Convex Function

- Cost function

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

- Write out each term

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

$$\frac{\partial^2 J}{\partial^2 \theta} = 2X^T X$$ Positive Semidefinite

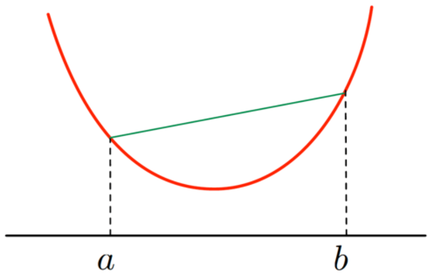
$J(\theta)$ Convex

- To minimize convex function $J(\theta)$ → let $\frac{\partial J}{\partial \theta} = 0$

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

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

*Convex Function

Convex Function	Second derivative	PSD
 A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if for all $a, b \in \mathbb{R}^d$ and $0 < \theta < 1$, $f(\theta a + (1-\theta)b) \leq \theta f(a) + (1-\theta)f(b).$	For a function $f : \mathbb{R}^d \to \mathbb{R}$, • the first derivative is a vector with d entries: $\nabla f(z) = \begin{pmatrix} \frac{\partial f}{\partial z_1} \\ \vdots \\ \frac{\partial f}{\partial z_d} \end{pmatrix}$ • the second derivative is a $d \times d$ matrix, the **Hessian** $H(z)$: $H_{jk} = \frac{\partial^2 f}{\partial z_j \partial z_k}$	A symmetric matrix M is **positive semidefinite (psd)** if: $x^T M x \geq 0$ for all vectors x

$$\text{If } f(x) \text{ is convex and } \partial f(x') = 0, \text{ then}$$
$$x' \text{ is a global minimum.}$$

Function of d variables

$$F : \mathbb{R}^d \to \mathbb{R}$$

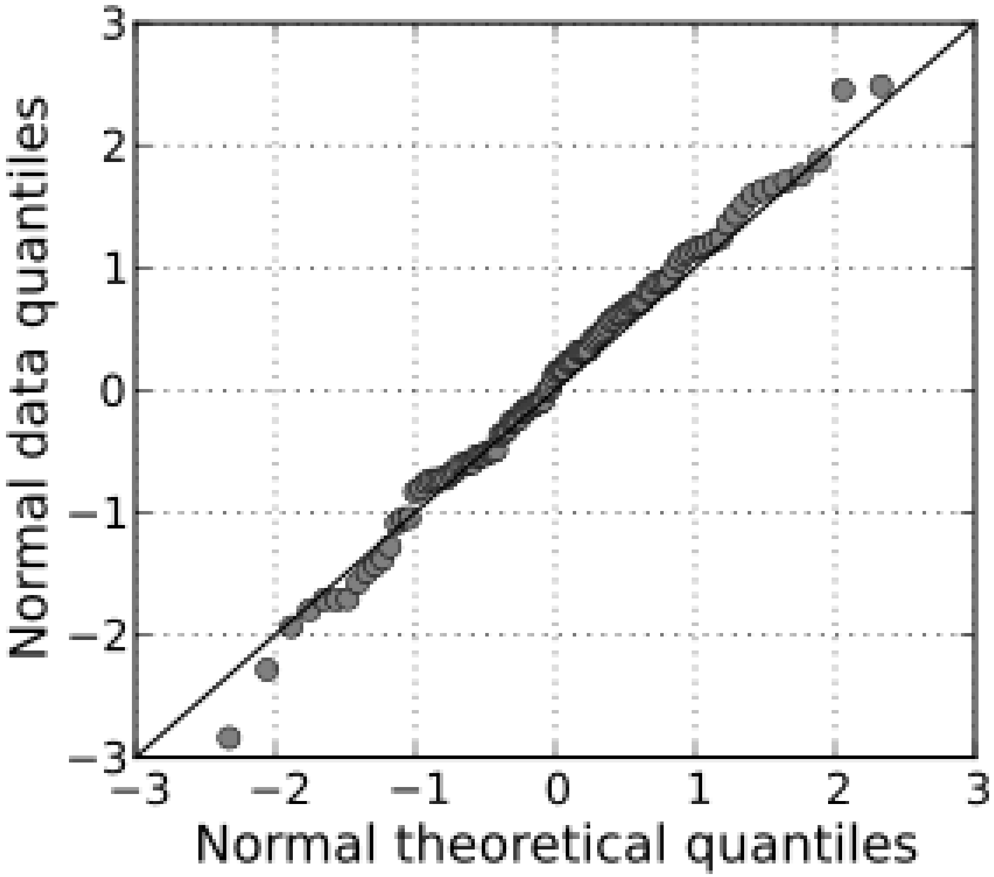
Convex if second derivative matrix is always positive semidefinite

We are going to use this proposition in the future lectures as well.

Residuals

- The residual should be normally distributed

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



Probabilistic Interpretation

- If the error term follows a normal distribution:

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

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

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

Maximum Likelihood Estimation

- Deriving the likelihood function

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

$$L(\theta) = \prod_{i=1}^{m} p\left(y^{(i)}|x^{(i)}; \theta\right)$$

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

Maximum Likelihood Estimation

- Maximum Likelihood Estimator (MLE) -- Log-likelihood Function

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

Equivalent to LSE

Analysis Metrics

Problem Model Algorithm **Analysis**

- Historical demand: D_t
- Forecast: y_t

Actual Demand	Forecast 1	Forecast 2
100	90	90
10	9	15
200	220	190

Which forecast is better?

Analysis Metrics

Problem Model Algorithm **Analysis**

- Historical demand: D_t
- Forecast: y_t

Metrics	Definition	Forecast 1	Forecast 2
~~Forecast error~~	$\cancel{\frac{1}{n}\sum_{t=1}^{n}(y_t - D_t)}$	3	-5
Mean absolute error (MAE) Mean absolute deviation (MAD)	$\frac{1}{n}\sum_{t=1}^{n}\lvert y_t - D_t\rvert$	$\frac{31}{3} = 10.3$	$\frac{25}{3} = 8.3$
Mean Squared error (MSE)	$\frac{1}{n}\sum_{t=1}^{n}(y_t - D_t)^2$	$\frac{501}{3} = 167$	$\frac{225}{3} = 75$
Mean Absolute Percentage Error (MAPE)	$\frac{1}{n}\sum_{t=1}^{n}\left\lvert\frac{y_t - D_t}{D_t}\right\rvert \times 100\%$	10%	$\frac{65}{3}\% = 21.7\%$
Weighted Absolute Percentage Error (WAPE)	$\frac{\sum_{t=1}^{n}\lvert y_t - D_t\rvert}{\sum_{t=1}^{n}\lvert D_t\rvert} \times 100\%$	10%	$\frac{25}{310}\% = 8.1\%$

Take Away

This class:

- Time series forecast model
  - Simple average
  - Moving average
  - Weighted moving average
  - Exponential Smoothing (single, double triple)
- Supervised learning model
  - Linear regression
    - LSE – convex optimization
    - MLE – equivalent to LSE under normal error assumption
- Analysis Metrics

Next class:

- Time series forecast model – ARIMA
- Supervised learning model – gradient boost