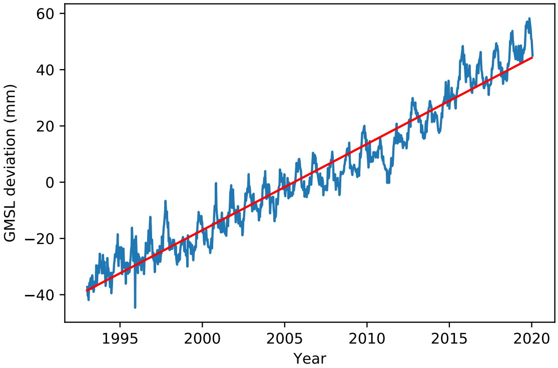
Time Series and Sequence Learning

Discussion seminar for Lectures 1 & 2

Johan Alenlöv, Linköping University

2022-08-30



Standard (linear) regression applied to time series data:

$$y_t = \theta_0 + \theta_1 u_t + \varepsilon_t,$$

where $u_t =$ “time when tth observation was recorded”.

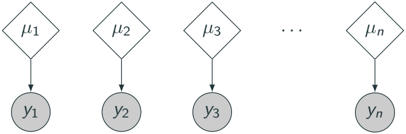
Discussion questions:

1. What does it mean in practice that the observations are **equidistant in time?**
2. For equidistant observations, what can be said about $u_t - u_{t-1}$?
3. If the observations are equidistant, how can we express $u_t = \ldots$?
4. In the simple linear regression model, does it matter if we use u_t or t as input (regressor)? Why/why not? That is, should we write

$$y_t = \theta_0 + \theta_1 u_t + \varepsilon_t \qquad \text{or} \qquad y_t = \theta_0 + \theta_1 t + \varepsilon_t \text{ ?}$$

Classical regression applied to time series data.

ex) $\mu_t = \theta_0 + \theta_1 u_t$



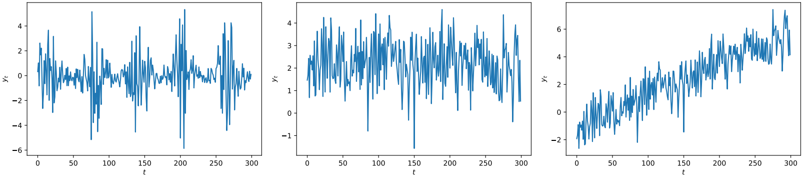
Discussion questions:

1. In the simple linear regression model, $y_t = \theta_0 + \theta_1 u_t + \varepsilon_t$ why do we say that y_t is **independent** of y_{t-1}? If there is a linear trend present, values close in time should take similar values, and thus be dependent, no?
2. Write out an expression for the joint pdf of n consecutive values $p(y_{1:n})$, assuming that y_t is given by the simple linear regression model. (*Tip.* use white board feature in Zoom)
3. Can we model **temporal dependencies** if we allow a more sophisticated (nonlinear) regression model, $\mu_t = f_\theta(t)$?
4. Try to come up with an example of a practical time series application for which it is important to take **statistical temporal dependencies** into account. Discuss why it is important in this application?

Def. A stochastic process $\{y_t\}_{t\geq 1}$ is said to be **(weakly) stationary** if, for all t,

1. $\mathrm{Var}(y_t) < \infty$,
2. $\mu(t) = \mathbb{E}[y_t] = \text{const.}$,
3. The autocovariance function depends only on the time lag,

$$\gamma(t, t+h) =: \gamma(h) \qquad\qquad \text{for all } h.$$

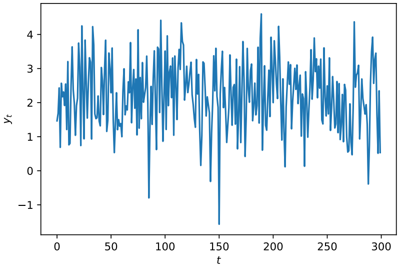


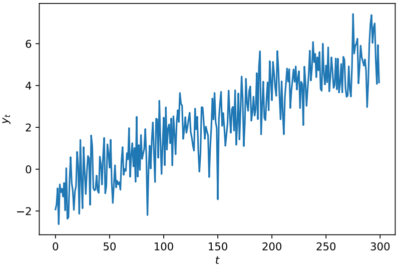
Discussion questions:

1. Which of the time series show above **appear to be** stationary? Why?
2. Assume that the AR(p) model

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots a_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

is stationary. What can be said about the mean function of the process $\mu(t) = \mathbb{E}[y_t]$?





Discussion questions: For each of the two data sets shown above:

1. Is is a good idea to **fit** an AR(p) model

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots a_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

   directly to the data? Why/why not?

2. Try to come up with a suitable procedure for fitting an AR(p) model to the data. Which steps are involved?

An AR(p) model

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots a_p y_{t-p} + \varepsilon_t,$$

is typically estimated using the normal equations:

$$\widehat{\theta} = \left(\Phi^{\mathsf{T}}\Phi\right)^{-1}\Phi^{\mathsf{T}}\mathrm{y}$$

where

$$\mathrm{y} = \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_n \end{pmatrix}, \qquad \Phi = \begin{pmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{n-1} & y_{n-2} & \cdots & y_{n-p} \end{pmatrix}$$

Discussion questions:

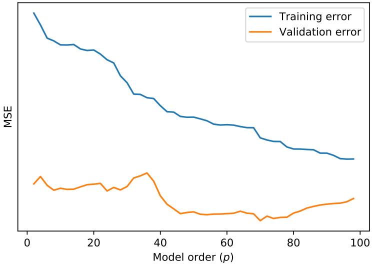
1. In standard linear regression problems we often see a "column of only ones" in the regression matrix. Why is that not the case for the matrix Φ? Which modeling assumptions does this correspond to?
2. Assuming that ε_t is white Gaussian noise, is the estimate shown on the previous slide **equivalent** to maximum likelihood estimation? Why/why not?
3. Consider the standard training procedure: We split the available data into training ($y_{1:n}$) and validation ($y_{n+1:n+m}$). We fit an AR(p) model to the training data and evaluate it by computing the error on the validation data. Is the training MSE always smaller than the validation MSE? Why/why not?

lecture2d - AR(p) models, validation

An AR(p) model is essentially a standard linear regression model, with p input variables corresponding to the p most recent values of the process.

Discussion questions:

1. Comparing the task of **order selection** for AR models, with the task of **input selection** for linear regression, what do these tasks have in common? How do they differ? Is one task in general simpler than the other?
2. In practice, we are not always lucky enough to get error curves looking like the school book examples. Consider the figure to the right — qualitatively, how does this differ from the "expected" behavior? Based on this figure, which order would you pick for your model and why?



Discussion questions:

1. Why is the autocorrelation function at lag 0 always equal to one, $\rho(t,t) = 1$? How can this be seen from the mathematical definition? What is the intuitive explanation?
2. For a **stationary** first order AR model, $y_t = ay_{t-1} + \varepsilon_t$, what can be said about the variance $\mathrm{Var}(y_t)$ as $a \to 1$?
3. For a **stationary** first order AR model, $y_t = ay_{t-1} + \varepsilon_t$, the ACF at lag h is given by $\rho(h) = a^h$. What does this imply if a is negative? Give an intuitive explanation for this behavior. (*Hint.* Think about even and odd values of h.)