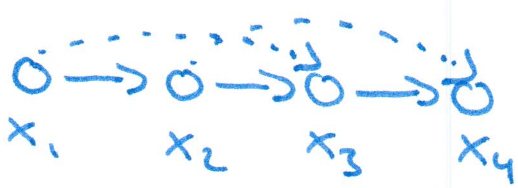
Time Series

- Time series are a sequence of samples and the order is important.

  ⇒ We only look at discrete, regular time intervals.



We talked already about stationary, loosley stationary and autocorrelation.

Autoregression:

$$X_{t+1} = \underbrace{\alpha}_{\text{coefficient}} X_t + \overbrace{Z_t}^{\text{noise } N(0,\sigma_Z)}$$

For which α is this stationary?

$\alpha = 0 \Rightarrow X_{t+1} = Z_t \Rightarrow$ stationary

$\alpha = 1 \Rightarrow X_{t+1} = X_t + Z_t \Rightarrow$ not stationary (random walk).

A

We define B as the operator taking us one step backwards: $X_{t-1} = B \cdot X_t$

$$X_t - \alpha X_{t-1} = Z_t$$

$$X_t - \alpha B X_t = Z_t$$

$$X_t = \frac{Z_t}{(1-\alpha B)}$$

$$\frac{1}{1-\alpha B} = (1-\alpha B)^{-1} \underset{\uparrow}{\approx} 1 + \alpha B + \alpha^2 B^2 + \ldots$$

Taylor expansion

$$\Rightarrow X_t = Z_t + \alpha B Z_t + \alpha^2 B^2 Z_t + \ldots$$

$$= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \ldots$$

$\Rightarrow$ geometric series, converges if $|\alpha| < 1$.

$\Rightarrow$ How do we find α and σ_Z when given a time series?

$$\mathrm{Cov}(X_t, X_{t+k}) \qquad = E[X_t \cdot X_{t+k}]$$

$$X_t = \sum_i \alpha^i \cdot Z_{t-i}$$

$$X_{t+k} = \sum_i \alpha^{i+k} \cdot Z_{t-i+k}$$

$$\Rightarrow \gamma(k) = \sum_i \alpha^i \alpha^{i+k} \cdot \sigma_Z^2$$

$$= \alpha^k \cdot \sigma_Z^2 \cdot \sum_i \alpha^{2i}$$

$$= \sigma_Z^2 \frac{\alpha^k}{1-\alpha^2}$$

$\Rightarrow$ For the autocorrelation we have:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \Rightarrow \rho(k) = \alpha^k$$

©

AR(p):

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t$$

⇒ How can we determine p?

AR(2):

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$$

$$X_{t-1} = \alpha_1 X_{t-2} + \alpha_2 X_{t-3} + Z_{t-1}$$

⇒ X_{t-2} influences X_t directly and also indirectly via X_{t-1}

⇒ Partial autocorrelation function (PACF) measures only the direct effect

⇒ PACF(1): α_1 for AR(1)

PACF(2): α_2 for AR(2)

How do we get those coefficients? Solve Yule-Walker equations (see notebook).

D

Moving Average (MA):

MA(q): $X_t = Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}$

$\Rightarrow$ weighted average of the past q+1 errors.

$$\mathbb{E}[X_t] = 0 \qquad Var[X_t] = \sigma_Z^2 \cdot \sum_{i=0}^{q} \beta_i^2$$

$$\rho_k = \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2}$$

$\Rightarrow$ The covariance is 0 for $k > q$.

$\Rightarrow$ Autocorrelation plot can help to identify q.

ARMA(p,q): Combines AR(p) and MA(q)

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_{t-p} X_{t-p} + Z_t + \beta_1 Z_{t-1}$$
$$+ \ldots + \beta_q Z_{t-q}$$

⇒ once we have established p and q using ACF and PACF plots, we can estimate the α and β coefficients using maximum-likelihood.

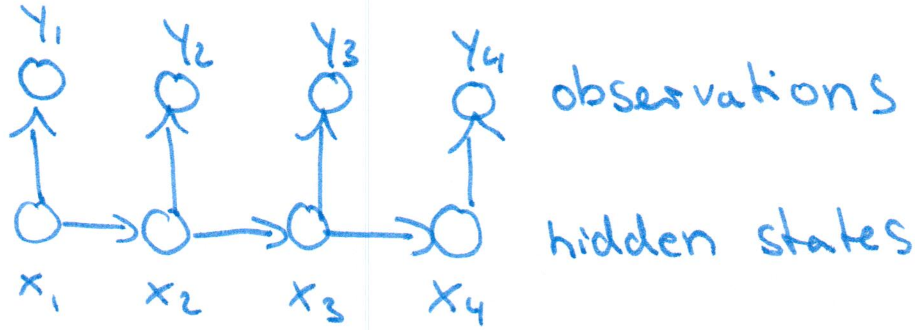
Box-Jenkins procedure:

- load the data
- plot ACF and PACF
- specify and fit ARMA/MA/AR model
- check goodness of fit
- generate forecasts.

⇒ I omitted the ARIMA model here, please see the notebook for further details and examples.

Lecture 18 - Hidden Markov Model

Hidden Markov Model:



A HMM is a discrete time, discrete state Markov chain with hidden states and observations.

$\Rightarrow$ HMMs can represent long range dependencies between observations via the hidden states.

Example: Estimate underlying heath state (hidden), given symptom descriptions (observations).

X: {heathy, sick} (states)

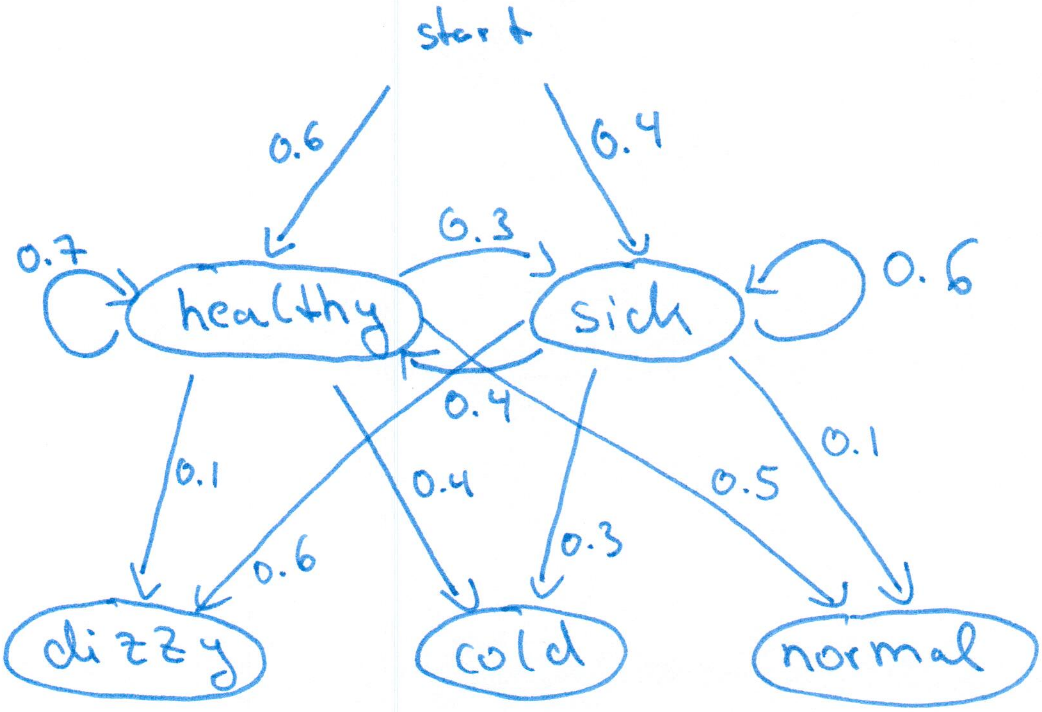
Y: {cold, dizzy, normal} (observations)

transition probabilities:

	healthy	sick
healty	0.7	0.3
sick	0.4	0.6

emission probabilities:

	cold	dizzy	normal
healthy	0.4	0.1	0.5
sick	0.3	0.6	0.1



we can compute the joint probability:

$$p(X_{1:T}, Y_{1:T}) = p(X_{1:T}) \cdot p(Y_{1:T} | X_{1:T})$$

$$= \left[p(X_1) \cdot \prod_{t=2}^{T} p(X_t | X_{t-1}) \right] \cdot$$

↳ transition prob.

$$\left[\prod_{t=1}^{T} p(Y_t | X_t) \right]$$

↳ emission prob.

Common inference problem: compute the prob. of a hidden state, given the observations.

<u>Forward algorithm:</u>

$$p(X_t | Y_{1:t}) = \sum_{X_{t-1}} p(X_t, X_{t-1} | Y_{1:t-1}, Y_t)$$

↳ marginalize over all possible states.

$$\propto \sum_{X_{t-1}} p(Y_t | X_t, X_{t-1}, Y_{1:t-1})$$

$$\cdot p(X_t | X_{t-1}, Y_{1:t-1}) \cdot p(X_{t-1} | Y_{1:t-1})$$

$$= \sum_{X_{t-1}} p(Y_t | X_t) \cdot p(X_t | X_{t-1}) \cdot p(X_{t-1} | Y_{1:t-1})$$

↳emission prob

transition prob.

same problem for the previous step!

We define $\alpha_t(X_t) = p(X_t | Y_{1:t})$

$$\Rightarrow \alpha_t(X_t) = \sum_{X_{t-1}} p(Y_t | X_t) \cdot p(X_t | X_{t-1}) \cdot \alpha_{t-1}(X_{t-1})$$

For $\alpha_1(X_1)$ we have:

$$\alpha_1(X_1) = p(X_1 | Y_1) \propto p(Y_1 | X_1) \cdot p(X_1)$$

↳emission prob.

↳start prob.

We can also try to predict future hidden states, given the observations we have seen so far:

$$p(X_{t+2} | Y_{1:t}) = \sum_{X_{t+1}} \sum_{X_t} p(X_{t+2}, X_{t+1}, X_t | Y_{1:t})$$

$$= \sum_{X_{t+1}} \sum_{X_t} p(X_{t+2} | X_{t+1}) \cdot p(X_{t+1} | X_t) \cdot p(X_t | Y_{1:t})$$

Predicting Observations:

$$p(Y_{t+h} \mid Y_{1:t}) = \sum_{X_{t+h}} p(Y_{t+h} \mid X_{t+h}) \cdot p(X_{t+h} \mid Y_{1:t})$$

Forward-Backward Algorithm:

Goal: Compute $p(X_t \mid Y_{1:T})$ with $T > t$

$\Rightarrow$ use past and future to infer the present.

$$p(X_t \mid Y_{1:T}) = p(X_t \mid Y_{1:t}, Y_{t+1:T})$$

$$\propto \underbrace{p(Y_{t+1:T} \mid X_t, \cancel{Y_{1:t}})}_{\substack{\beta_t(X_t) \\ \hookrightarrow \text{backward}}} \cdot \underbrace{p(X_t \mid Y_{1:t})}_{\substack{\alpha_t(X_t) \\ \hookrightarrow \text{forward}}}$$

Backward Algorithm:

$$\beta_{t-1}(X_{t-1}) = p(Y_{t:T} \mid X_{t-1})$$

$$= \sum_{X_t} p(X_t, Y_t, Y_{t+1:T} \mid X_{t-1})$$

$$= \sum_{X_t} p(Y_{t+1:T} \mid X_t, \cancel{Y_t}, \cancel{X_{t-1}}) \cdot p(X_t, Y_t \mid X_{t-1})$$

$$= \sum_{X_t} \underbrace{p(Y_{t+1:T} \mid X_t)}_{\beta_t(X_t)} \cdot p(Y_t \mid X_t, \cancel{X_{t-1}})$$

↳ emission prob.

$$\cdot\, p(X_t \mid X_{t-1})$$

↳ transition prob.

⇒ We need to know the terminal time value, then we can compute backwards.

$$\beta_T(X_T) = p(Y_{T+1:T} \mid X_T) = p(\emptyset \mid X_T) = 1$$

⇒ probability of a non-event.

<u>Viterbi algorithm</u>: Compute the most probable sequence of states:

$$X^* = \underset{X_{1:T}}{\operatorname{argmax}}\; p(X_{1:T} \mid Y_{1:T})$$

$$\delta_t(X_t) = \max_{X_1,\dots,X_{t-1}} p(X_{1:t-1}, X_t \mid Y_{1:t})$$

$\Rightarrow$ probability of ending up in X_t if we take the most probable path given the observations.

$$\delta_t(X_t) = \max_{X_{t-1}} \delta_{t-1}(X_{t-1}) \cdot \underbrace{p(X_t \mid X_{t-1})}_{\text{transition}} \cdot \underbrace{p(Y_t \mid X_t)}_{\text{emission}}$$

$$\delta_1(X_1) = p(Y_1 \mid X_1) \cdot \underbrace{p(X_1)}_{\text{start prob.}}$$

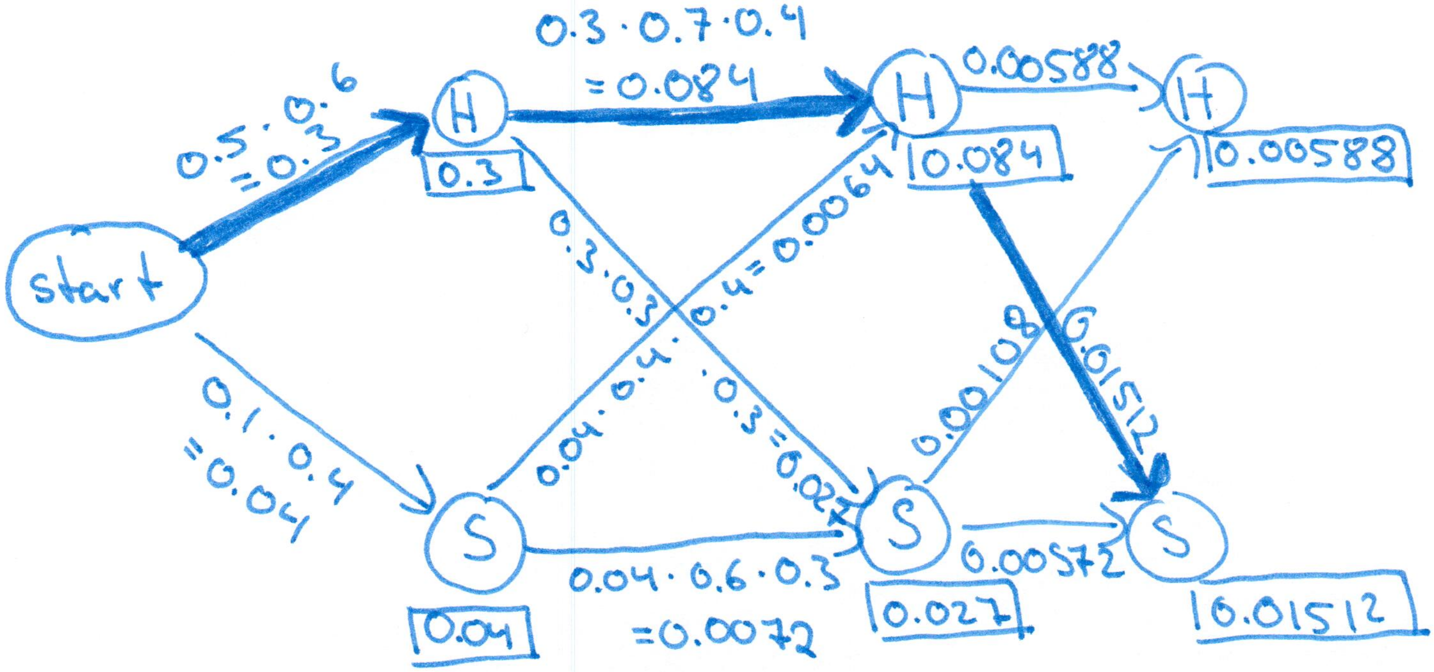
We also need backpointers:

$$a_t(X_t) = \operatorname*{argmax}_{X_t} \delta_{t-1}(X_{t-1}) \cdot p(X_t \mid X_{t-1}) \cdot p(Y_t \mid X_t)$$

$\Rightarrow a_t(X_t)$ tells us the most likely previous state on the most probable path to X_t.

Example:

observations: normal, cold, dizzy



⇒ most prob. state path is H H S.

⇒ Need to be careful about numerical underflow when implementing.