

WYKŁAD: Szeregi czasowe I

Zaawansowane Metody Uczenia Maszynowego

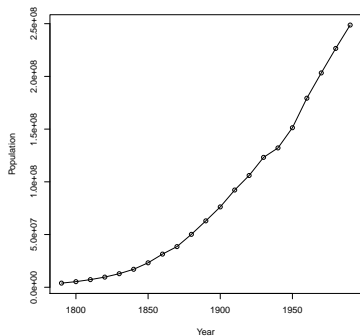
Szereg czasowy

- (X_t) - ciąg zmiennych losowych indeksowany parametrem t (czas). Z reguły $t \in N$ lub $t \in Z$.
- Dotąd rozpatrywaliśmy: (X_t) - ciąg niezależnych zmiennych losowych o tym samym rozkładzie. Teraz dopuszczamy zależność między zmiennymi i różny rozkład zmiennych.
- Dwie podstawowe charakterystyki szeregu: jak zachowuje się **wartość średnia** w funkcji czasu ? Jak **zależność** między dwiema zmiennymi X_s i X_t zależy od odcinka czasu, który upłynął między tymi zdarzeniami?
- Czy mamy do czynienia z 'efektem motyla' (dlaeka przeszłość ma wpływ na terażniejszość) ?

- Zależność między obserwacjami w szeregu czasowym i różnice rozkładów: podstawowe różnice w porównaniu z sytuacją iid. Wpływają one na postać estymatorów i procedury wnioskowania dla szeregów czasowych.
- Szeregi czasowe w R: obiekty klasy `ts` (time series) o strukturze: wartości szeregu czasowego, liczba obserwacji na jednostkę czasu, moment początku i końca obserwacji.
Uwaga. Wiele podstawowych funkcji, jak `acf` i `pacf`, aby obiekt był klasy `ts`

```
library(MASS)
USpop <- ts(data=scan("USPOP.DATA"),
  start=1790, end=1990, frequency=0.1)
# option frequency- no. of obs per time unit,
# in this case unit=1 year,
#frequency=0.1 means 1 observation every 10 years

ts.plot(USpop, gpars=list(xlab="Year", ylab="Population",
type="o"))
```



Wielkość populacji USA w latach 1790-1990. Wyraźny wzrost w czasie i spowolnienie go w latach 1930-40 (wielki kryzys)

Przykłady szeregów czasowych

- **Biały szum** $WN(0, \sigma^2)$

$X_t = \varepsilon_t$: ciąg niezależnych zmiennych losowych o średniej 0 i wariancji σ^2 .

$$EX_t = 0$$

$$\text{Cov}(X_t, X_s) = \sigma^2 I\{t = s\} = \sigma^2 I\{t - s = 0\} = 0$$

$$\rho(X_t, X_s) = \frac{\text{Cov}(X_t, X_{t+h})}{(\text{Var}(X_t)\text{Var}(X_{t+h}))^{1/2}} = I\{t = s\}$$

Siła zależności zależy tylko od $|t - s|$!

- **Proces średniej ruchomej rzędu 1: MA(1)**

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

ε_t : $WN(0, \sigma^2)$

$$EX_t = 0$$

Proces MA(1)

-

$$\text{Cov}(X_s, X_t) = 0 \quad \text{gdy} \quad |t - s| > 0$$

$$\text{Cov}(X_t, X_{t+1}) = \text{Cov}(X_t, X_{t-1}) = \text{Cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t+1} + \theta\varepsilon_t) = \theta\sigma^2$$

$$\text{Cov}(X_t, X_t) = \text{Var}X_t = E(\varepsilon_t + \theta\varepsilon_{t-1})^2 = \sigma^2(1 + \theta^2)$$

- **Proces autoregresyjny rzędu 1 AR(1)**

$$X_t = \phi X_{t-1} + \varepsilon_t,$$

gdzie $|\phi| < 1$ i ε_t : $WN(0, \sigma^2)$, ε_t : niezależny od $X_s, s \leq t - 1$.

Równanie (auto)regresji: X_t - odpowiedź, X_{t-1} : predyktor.

Proces autoregresyjny rzędu 1 AR(1)

$$\begin{aligned} X_t &= \phi X_{t-1} + \varepsilon_t = \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi^2 X_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \phi^2(\phi X_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \sum_{i=0}^k \phi^i \varepsilon_{t-i} + \phi^{k+1} X_{t-(k+1)} \end{aligned} \quad (1)$$

Jeśli $EX_t^2 \leq C$, to ostatni wyraz $\rightarrow 0$ ($|\phi| < 1$) i mamy przedstawienie

$$X_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$$

Proces autoregresyjny rzędu 1 AR(1)

$$\text{Cov}(X_t, X_{t+1}) = \text{Cov}(X_t, \phi X_t + \varepsilon_{t+1}) = \phi \text{Var}(X_t).$$

Analogicznie

$$\text{Cov}(X_t, X_s) = \phi^{|t-s|} \text{Var}(X_t).$$

We wszystkich trzech przypadkach $\text{Cov}(X_s, X_t)$ zależy od $|t - s|$ i średnia jest stała. To proces stacjonarny w szerszym sensie.

Procesy stacjonarne w szerszym sensie

(X_t) stacjonarny w szerszym sensie, jeśli

- $EX_t = m$ dla każdego t ;
- $\text{Cov}(X_s, X_t) = \text{Cov}(X_{s+h}, X_{t+h}) = \gamma(t - s)$

$\text{Cov}(X_s, X_t)$ jest funkcją różnicy momentów czasowych.

Funkcja ACF (autokorelacji)

$$\rho(h) = \rho(X_t, X_{t+h}) = \frac{\text{Cov}(X_t, X_{t+h})}{(\text{Var}(X_t)\text{Var}(X_{t+h}))^{1/2}} = \frac{\gamma(h)}{\gamma(0)}$$

Własności funkcji kowariancji $\gamma(h)$

-

$$\gamma(0) = \text{Cov}(X_t, X_t) = \text{Var}(X_t)$$

-

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_{t+h}, X_t) = \gamma(-h)$$

- Nieujemna określoność funkcji $\gamma(\cdot)$: dla dowolnych a_1, \dots, a_k :

$$\sum_{i,j=1}^k \gamma(i-j) a_i a_j \geq 0$$

Procesy stacjonarne w węższym sensie

Własność silniejsza niż stacjonarność w węższym sensie.

Proces (X_t) jest stacjonarny w węższym sensie, jeśli wektor losowy $(X_{t_1}, \dots, X_{t_k})$ ma taki sam rozkład jak $(X_{t_1+h}, \dots, X_{t_k+h})$.

W szczególności takie same rozkłady ma X_s i X_t , zatem pokrywają się ich średnie, oraz takie same rozkłady mają pary (X_s, X_{s+h}) oraz (X_t, X_{t+h}) , zatem pokrywają się ich kowariancje:
stacjonarność w węższym sensie implikuje stacjonarność w szerszym sensie

Sprowadzanie do stacjonarności

Wiele szeregów niestacjonarnych:

dla procesu $X_t = at + \varepsilon_t$, gdzie $\varepsilon_t : WN(0, \sigma^2)$.

$$EX_t = E(at + \varepsilon_t) = at$$

Wartość średnia X_t zależy od t . Jak sprowadzić do stałej wartości oczekiwanej ?

Różnicowanie

$$\Delta(X_t) = X_t - X_{t-1}$$

Dla naszego przykładu

$$\Delta(X_t) = at + \varepsilon_t - (a(t-1) + \varepsilon_{t-1}) = a + \varepsilon_t - \varepsilon_{t-1}$$

Problem związany z różnicowaniem: zmienia się struktura błędu
($\varepsilon_t \rightarrow \varepsilon_t - \varepsilon_{t-1}$)

Sprowadzanie do stacjonarności

Podobnie dla trendu wielomianowego:

$$X_t = a_k t^k + a_{k-1} t^{k-1} + \dots + a_0 + \varepsilon_t$$

mamy

$$E\Delta^k(X_t) = E\Delta \circ \Delta \dots \circ \Delta(X_t) = a_k k!$$

Inna metoda: estymacja trendu i przez jego odjęcie doprowadzenie szeregu do przybliżonej stacjonarności.

Estymacja średniej i funkcji kowariancji

(X_t) : proces stacjonarny o średniej m . Podstawowe estymatory

$$\bar{X}_t = \hat{m} = \frac{1}{t}(X_1 + \dots + X_t)$$

$$\hat{\gamma}(h) = \frac{1}{t - |h|} \sum_{t=1}^{t-|h|} (X_t - \bar{X}_t)(X_{t+|h|} - \bar{X}_t)$$

Dla oszacowania $\gamma(h)$ zastępujemy wartość oczekiwaną w definicji przez średnią możliwych iloczynów $(X_t - \bar{X}_t)(X_{t+|h|} - \bar{X}_t)$ dla wszystkich par $(X_t, X_{t+|h|})$ takich, że $1 \leq t \leq n$, $1 \leq t + |h| \leq n$.



- Jakość estymacji $\gamma(h)$ przez $\hat{\gamma}(h)$ zależy od $|h|$: liczymy średnią z $n - |h|$ obserwacji. Dla $h = n - 1$ mamy tylko jedną obserwację ! W praktyce wybiera się h tak, aby

$$h \leq t/3 \quad \text{lub} \quad h \leq \sqrt{t}$$

- Czynniki $1/(t - |h|)$ w definicji $\hat{\gamma}(h)$ zastępuje się często przez $1/t$. Dostajemy wtedy funkcję nieujemnie określoną (tak samo jak $\gamma(h)$): estymator ma tę samą własność co obiekt, który estymujemy.
- Autokorelacja próbkowa definiowana w sposób naturalny

$$ACF(h) = \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Rozkład ACF dla białego szumu

$(X_t) = (\varepsilon_t)$: biały szum (ciąg niezależnych zmiennych losowych).
 $0 < t_1 < \dots < t_h \in N$. Wtedy dla $\hat{\rho}(h) = (\hat{\rho}(t_1), \dots, \hat{\rho}(t_h))$ i
 $\rho(h) = (\rho(t_1), \dots, \rho(t_h))$ mamy

$$n^{1/2}(\hat{\rho}(h) - \rho(h)) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}),$$

Pas ufności dla $H_0 : \rho(h) \equiv 0$

$$CI_{1-\alpha}(t_k) = \left(\hat{\rho}(t_k) \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \right) \quad k = 1, \dots, h$$

Odrzucamy H_0 gdy $0 \notin CI_{1-\alpha}(t_k)$ w więcej niż αh przypadkach (dla więcej niż jednego dla $h = 20$ i $\alpha = 0.05$) lub inaczej w tylu przypadkach

$$|\hat{\rho}(t_k)| \geq \frac{z_{1-\alpha/2}}{\sqrt{n}} \quad k = 1, \dots, h$$

Testy dla białego szumu

Z twierdzenia o rozkładzie dla empirycznego współczynnika autokorelacji wynika, że

$$Q = n \sum_{i=1}^h \hat{\rho}(i)^2 \xrightarrow{\mathcal{D}} \chi_h^2$$

(suma kwadratów współrzędnych lewej strony zbiega do sumy kwadratów współrzędnych prawej strony). W praktyce używa się modyfikacji Ljunga-Boxa

$$Q_{LB} = n(n+1) \sum_{i=1}^h \hat{\rho}(i)^2 / (n-i) \xrightarrow{\mathcal{D}} \chi_h^2$$

Obszar krytyczny w przypadku testowania hipotezy, że proces jest białym szumem przy użyciu tej statystyki ma postać

$$\{Q_{LB} > \chi_{h,1-\alpha}^2\}$$

Prognoza liniowa

X_t : proces stacjonarny w szerszym sensie, $EX_t = 0$. Interesuje nas **optymalna prognoza liniowa** X_{t+1} na podstawie X_1, \dots, X_t , to znaczy taka kombinacja

$$\bar{a}_1 X_t + \bar{a}_2 X_{t-2} + \dots + \bar{a}_t X_1,$$

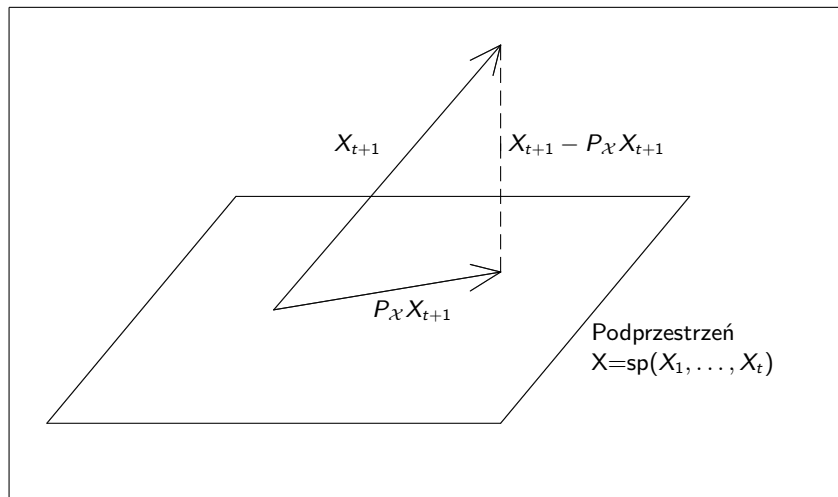
która jest rozwiązaniem problemu minimalizacji

$$(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_t) = \operatorname{argmin}_{a_1, \dots, a_t} E(X_{t+1} - a_1 X_t - a_2 X_{t-2} - \dots - a_t X_1)^2$$

Kombinacja $\bar{a}_1 X_t + \bar{a}_2 X_{t-2} + \dots + \bar{a}_t X_1$ będąca rozwiązaniem problemu jest rzutem ortogonalnym elementu X_{t+1} na podprzestrzeń $\mathcal{X} = \operatorname{sp}(X_1, \dots, X_t)$ rozpiętą na X_1, \dots, X_t :

$$\bar{a}_1 X_t + \bar{a}_2 X_{t-2} + \dots + \bar{a}_t X_1 = P_{\mathcal{X}} X_{t+1}$$

Prognoza liniowa



Rysunek: Prognoza $P_X X_{t+1}$ jest prostopadłym rzutem X_{t+1}

Równania prognozy liniowej

Prostopadłość zmiennych losowych: $X \perp Y$ o średniej 0 $\iff EXY = 0$.

$$X_{t+1} - P_{\mathcal{X}}X_{t+1} \perp X_j \quad j = 1, \dots, t$$

$$E(X_{t+1} - \bar{a}_1X_t - \bar{a}_2X_{t-1} - \dots - \bar{a}_tX_1)X_j = 0 \quad j = 1, \dots, t$$

$$EX_{t+1}X_j = \sum_{i=1}^t \bar{a}_i EX_{t+1-i}X_j$$

$$\gamma(t+1-j) = \sum_{i=1}^t \bar{a}_i \gamma(t+1-i-j)$$

$$j := t+1-j$$

$$\gamma(j) = \sum_{i=1}^t a_i \gamma(j-i) \quad j = 1, \dots, t$$

Równania prognozy liniowej

W postaci macierzowej

$$\gamma = \Gamma \bar{\mathbf{a}},$$

gdzie $\gamma = (\gamma(1), \dots, \gamma(t))'$, $\bar{\mathbf{a}} = (a_1, \dots, a_t)'$ a $\Gamma = (\gamma(i-j))_{i,j \leq t}$. Jeśli Γ jest odwracalna (zmiennie X_1, \dots, X_t nie są liniowo zależne), to

$$\bar{\mathbf{a}} = \Gamma^{-1} \gamma$$

(równania Yule'a-Walkera). Ich odpowiedniki empiryczne

$$\hat{\mathbf{a}} = \hat{\Gamma}^{-1} \hat{\gamma},$$

gdzie $\hat{\Gamma} = (\hat{\gamma}(i-j))_{i,j \leq t}$, $\hat{\gamma} = (\hat{\gamma}(1), \dots, \hat{\gamma}(t))'$.

Uwaga Prognoza liniowa jest nieefektywna gdy zależność X_{t+1} od X_1, \dots, X_t nie jest liniowa

Współczynnik korelacji częściowej

Istotną rolę w identyfikacji szeregów czasowych odgrywa współczynnik korelacji częściowej PACF (partial autocorrelation coefficient)

$$\alpha(t) = \rho(X_1 - P_{\mathcal{X}}X_1, X_{h+1} - P_{\mathcal{X}}X_{t+1}),$$

gdzie $P_{\mathcal{X}}X_1$ jest rzutem prostopadłym elementu X_1 na \mathcal{X} : przestrzeń rozpiętą na elementach X_2, \dots, X_t .

Liniowe procesy ARMA

Trzy podstawowe klasy procesów liniowych (X_t zależy liniowo od przeszłych obserwacji):

- Procesy średniej ruchomej rzędu q $MA(q)$;
- Procesy autoregresyjne rzędu p $AR(p)$
- Procesy $ARMA(p, q)$ (ogólnienie dwóch poprzednich klas)

Proces średniej ruchomej rzędu q $MA(q)$ (przefiltrowany biały szum)

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

ε_t - $WN(0, \sigma^2)$. Proces stacjonarny i taki, że $\gamma(h) = 0$ dla $|h| > q$.

Liniowe procesy ARMA

Proces autoregresyjny rzędu p

AR(p)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

ε_t - WN($0, \sigma^2$) taki, że ε_t jest niezależne od X_s dla $s < t$.

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} + \dots - \phi_p X_{t-p} = \varepsilon_t.$$

Proces stacjonarny spełniający powyższe równanie istnieje, gdy

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$$

nie ma pierwiastków dla $z \in \mathbb{C} : |z| \leq 1$.

Jak identyfikować procesy MA(q) i AR(p) ?

Jak identyfikować procesy $MA(q)$ i $AR(p)$?

- Proces stacjonarny i taki, że $\gamma(h) = 0$ dla $|h| > q \iff X_t$ jest $MA(q)$. Identyfikacja procesu na podstawie empirycznej funkcji ACF
- $AR(p)$ ma własność:

$$P_{\mathcal{X}} X_{t+1} = \phi_1 X_{t+1-1} + \phi_2 X_{t+1-2} + \dots + \phi_p X_{t+1-p},$$

gdzie $\mathcal{X} = sp(X_t, \dots, X_{t+1-s})$. Stąd wynika, że

$$PACF(h) = 0 \quad \text{dla } |h| > p$$

Identyfikacja procesu na podstawie empirycznej funkcji PACF