

ZAAWANSOWANE METODY EKSPLORACJI

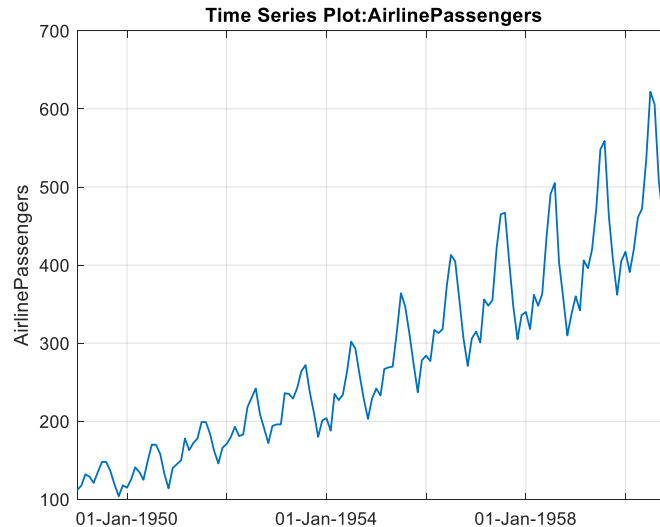
PREDYKCJA SZEREGÓW CZASOWYCH ZE ZŁOŻONĄ SEZONOWOŚCIĄ

Prof. dr hab. inż. Grzegorz Dudek
Wydział Matematyki i Informatyki
Uniwersytet Łódzki

Szereg czasowy to sekwencja danych uporządkowanych w kolejności czasowej.

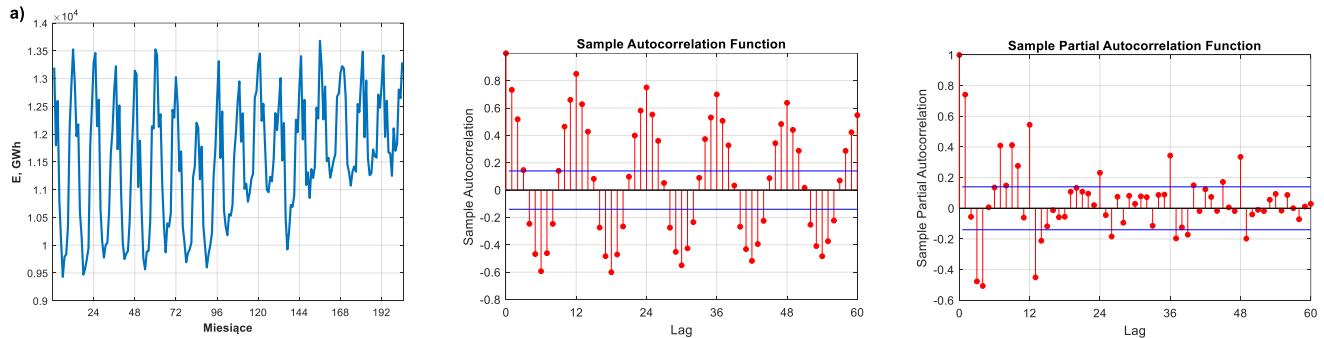
Np. zbiór stanów zmiennej z , uporządkowanych według wartości zmiennej czasowej $t = 1, 2, \dots, N$, charakteryzujących zjawisko (stan obiektu), np. liczba klientów w kolejnych miesiącach, kurs waluty w kolejnych dniach, temperatura atmosferyczna w kolejnych godzinach:

$$\{z_t\}_{t=1}^N$$



Zależność pomiędzy obserwacjami szeregu czasowego odległymi o l (*lag*) jednostek czasu ocenia się za pomocą funkcji autokorelacji (ACF) i autokorelacji cząstkowej (PACF). Ta pierwsza ma postać:

$$ACF(l) = \frac{\hat{\gamma}(l)}{\hat{\gamma}(0)}, \quad \hat{\gamma}(l) = \frac{1}{N} \sum_{j=1}^{N-l} (Z_{j+l} - \bar{Z})(Z_j - \bar{Z}), \quad l = 0, 1, \dots, N-1$$



Wykresy szeregu czasowego miesięcznego zapotrzebowania na energię elektryczną dla Polski i jego autokorelacji i korelacji cząstkowej

Składowe szeregi czasowych (są efektem wpływu różnych czynników na badane zjawisko):

- tendencja rozwojowa (trend), która ujawnia się przez jednokierunkowe i systematyczne zmiany poziomu danego zjawiska, które zachodzą w długim okresie,
- wahania okresowe (sezonowe), czyli rytmiczne wahania o określonym cyklu,
- wahania koniunkturalne, czyli systemowe, falowe wahania obserwowane w dłuższych okresach,
- wahania przypadkowe, czyli nieregularne zakłócenia losowe.

Dekompozycja szeregu umożliwia wyodrębnienie regularnych wzorców i tendencji w szeregu czasowym, czyli trendu i wahań okresowych (*seasonal component*). Składowa wahań koniunkturalnych (cyklicznych), o zmiennej częstotliwości, zwykle łączona jest ze składową trendu. W wyniku dekompozycji identyfikowana jest także składowa wahań przypadkowych (*remainder component*).

Dekompozycję szeregu czasowego można opisać ogólną formułą:

$$Z_t = f(S_t, T_t, R_t),$$

gdzie f reprezentuje pewną funkcję, S_t to składowa sezonowa, T_t to trend długoterminowy, a R_t to zakłócenie losowe.

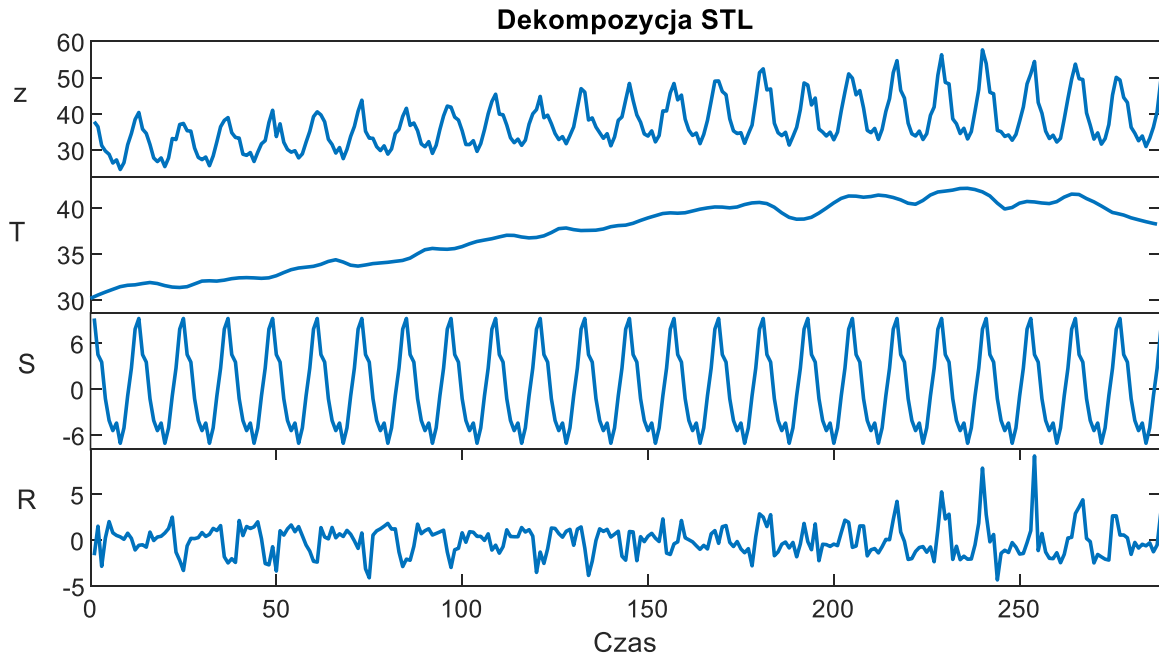
Najbardziej popularne modele dekompozycji to dekompozycja addytywna:

$$Z_t = S_t + T_t + R_t$$

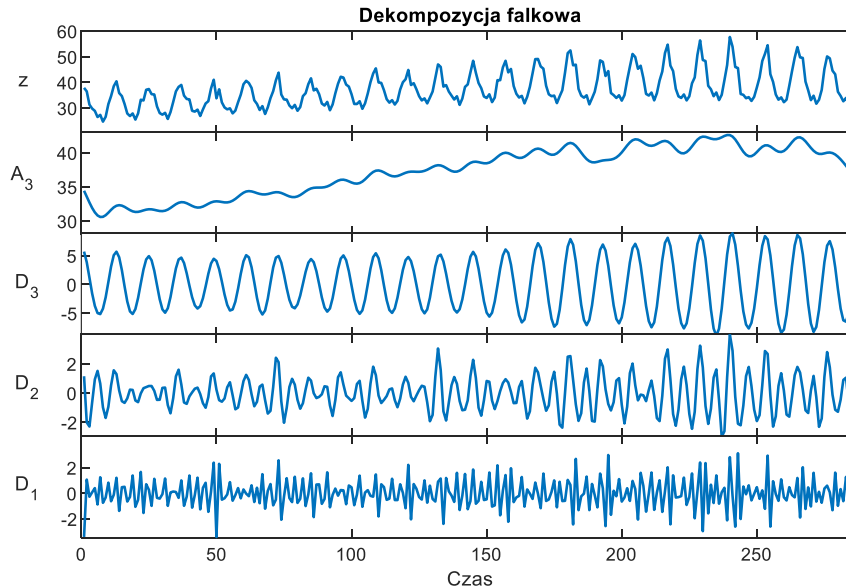
i multiplikatywną

$$Z_t = S_t T_t R_t$$

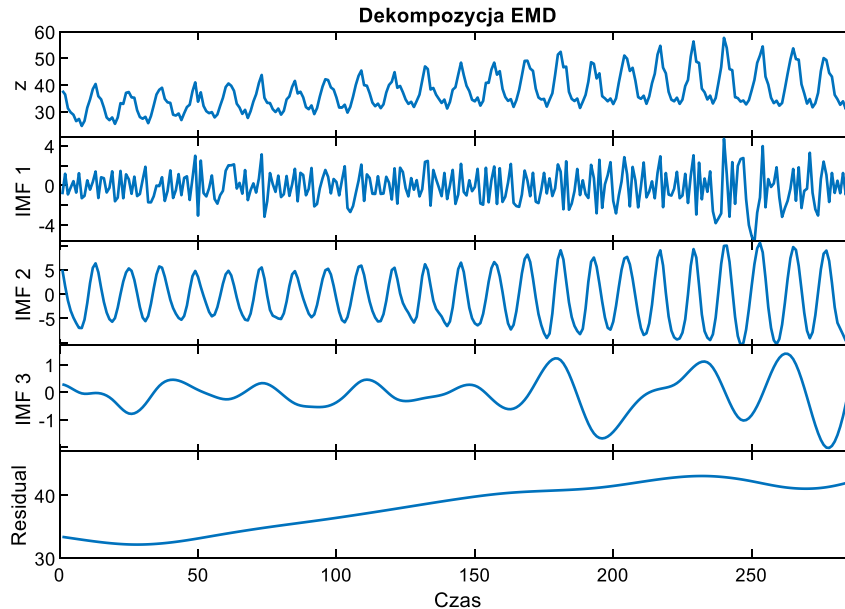
Dekompozycja STL (*Seasonal and Trend decomposition using Loess*) ma charakter addytywny. Procedura dekompozycji STL jest iteracyjna i polega na naprzemiennej estymacji trendu oraz składnika sezonowego za pomocą ważonego lokalnego wygładzania (*Loess*).



Dekompozycja falkowa rozkłada szereg w procesie iteracyjnym na składowe o malejących częstotliwościach. Na kolejnych etapach dekompozycji szereg jest przetwarzany przez parę filtrów, górnoprzepustowy i dolnoprzepustowy. W wyniku otrzymuje się składową niskoczęstotliwościową, tzw. aproksymację A , reprezentującą trend i składową wysokoczęstotliwościową, tzw. detal D , reprezentującą szczegółowe cechy szeregu. W każdej j -tej iteracji aproksymacja z iteracji poprzedniej A_{j-1} dekomponowana jest na detal D_j i aproksymację A_j .

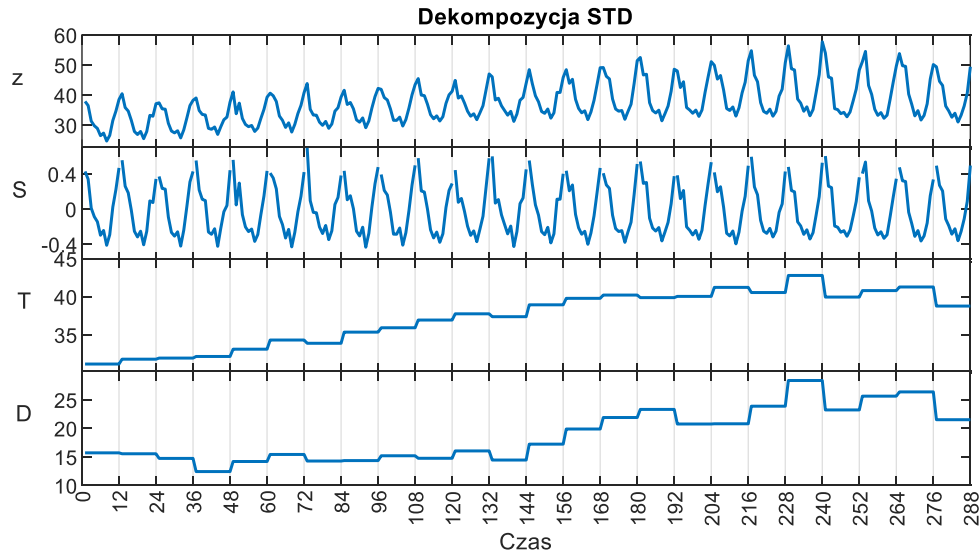


Dekompozycja EMD (*Empirical Mode Decomposition*) pozwala rozłożyć szeregi czasowe w dziedzinie czasu na składowe zwane *Intrinsic Mode Functions* (IMF), które mają zmienną amplitudę i częstotliwość wzdłuż osi czasu. Składowe IMF uzyskuje się poprzez zastosowanie rekurencyjnego algorytmu przesiewowego (*sifting process*). Algorytm ten wyodrębnia lokalne minima i maksima szeregu, a następnie interpoluje je oddzielnie za pomocą splajnów kubicznych.

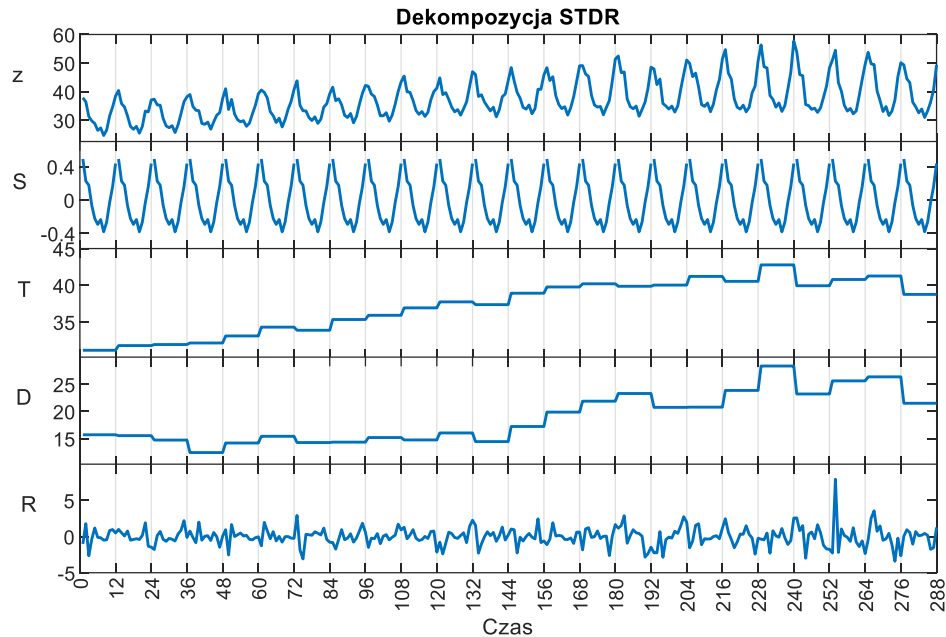


Dekompozycja szeregów czasowych

Dekompozycja STD (na składową sezonową, trend i dyspersję): $Z_t = S_t D_t + T_t$, gdzie D_t jest dyspersją szeregu czasowego. Składowa trendu reprezentuje wartości średnie wyrazów szeregu w kolejnych okresach sezonowych $\bar{Z}_i = \frac{1}{n} \sum_{j=1}^n Z_{i,j}$, składowa dyspersji reprezentuje rozproszenia wyrazów szeregu w kolejnych okresach sezonowych $\tilde{Z}_i = \sqrt{\sum_{j=1}^n (Z_{i,j} - \bar{Z}_i)^2}$, składową sezonową otrzymuje się ze wzoru:

$$S_t = \frac{Z_t - T_t}{D_t}.$$


Dekompozycja STDR (na składowa sezonową, trend, dyspersję i składową resztową). Składową resztową otrzymujemy ze wzoru: $R_t = Z_t - (S'_t D_t + T_t)$, gdzie S'_t jest homogeniczną składową sezonową $\{S'_t\}_{t=1}^N = \underbrace{\{\{\bar{S}_j\}_{j=1}^n, \dots, \{\bar{S}_j\}_{j=1}^n\}}_{N/n \text{ razy}}$, $\bar{S}_j = \frac{1}{n} \sum_{i=1}^{N/n} S_{i,j}$ jest uśrednioną sekwencją sezonową.



ARIMA (Autoregressive Integrated Moving Average)

W wersji sezonowej $SARIMA(p, d, q)(P, D, Q)_m$:

$$\Phi(B^m)\phi(B)(1 - B^m)^D(1 - B)^d z_t = c + \Theta(B^m)\theta(B)\xi_t$$

gdzie z_t są wyrazami szeregu czasowego,

ξ_t to proces szumu białego z zerową średnią i wariancją σ^2 ,

m to długość cyklu sezonowego,

B to operator przesunięcia wstecz ($z_t - z_{t-1} = z_t - Bz_t = (1 - B)z_t$),

D i d to rzędy różnicowania (sezonowego i zwykłego),

$\phi(\cdot), \Phi(\cdot), \theta(\cdot)$, i $\Theta(\cdot)$ to odpowiednio wielomiany stopnia p, q, P i Q ,

a c oznacza stałą.

Wygładzanie wykładnicze (*Exponential Smoothing*)

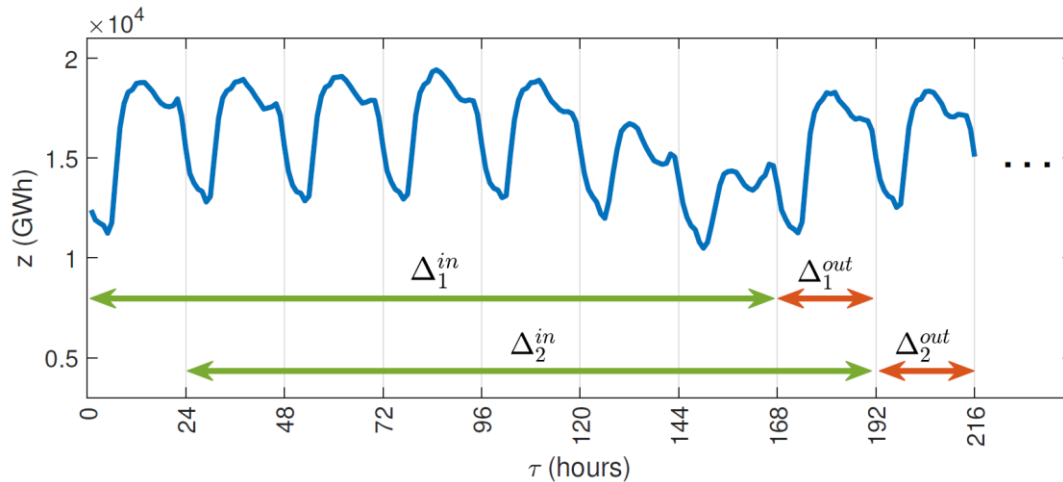
Idea wygładzania wykładniczego polega na przypisaniu obserwacjom z poprzednich okresów wykładniczo zanikających wag. Obserwacje z ostatnich okresów mają dzięki temu większy wpływ na prognozę. Model Holta-Wintersa opiera się na trzech równaniach wygładzających, które reprezentują poziom zmiennej prognozowanej, jej trend i sezonowość.

	Wersja addytywna	Wersja multiplikatywna
Poziom	$L_t = \alpha(z_t - S_{t-m}) + (1 + \alpha)(L_{t-1} + b_{t-1})$	$L_t = \alpha \frac{z_t}{S_{t-m}} + (1 - \alpha)(L_{t-1} + b_{t-1})$
Trend	$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$	$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$
Sezonowość	$S_t = \gamma(z_t - L_{t-1} - b_{t-1}) + (1 - \gamma)S_{t-m}$	$S_t = \gamma \frac{z_t}{L_t} + (1 - \gamma)S_{t-m}$
Prognoza	$\hat{z}_{t+h t} = L_t + hb_t + S_{t+h-m(k+1)}$	$\hat{z}_{t+h t} = (L_t + hb_t)S_{t+h-m(k+1)}$

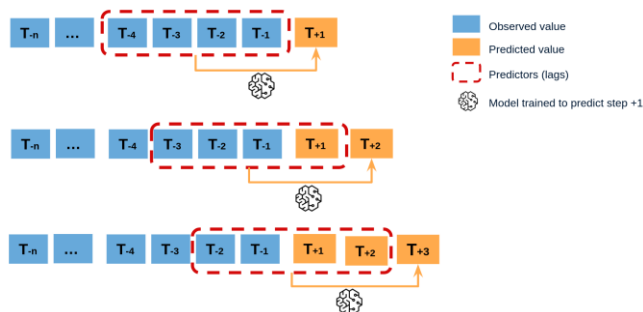
gdzie $\alpha, \beta, \gamma \in [0, 1]$ to współczynniki wygładzania, h to horyzont prognozy, $k = \text{int}((h - 1)/m)$.



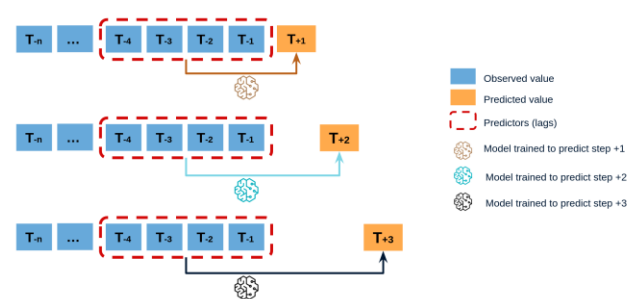
$\mathbf{x}_i \in \Delta_i^{in}$ to wektor predyktorów, $\mathbf{y}_i \in \Delta_i^{out}$ to wektor prognoz
Model uczy się na zbiorze treningowym $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$



Podójście rekurencyjne *Recursive multi-step prediction*

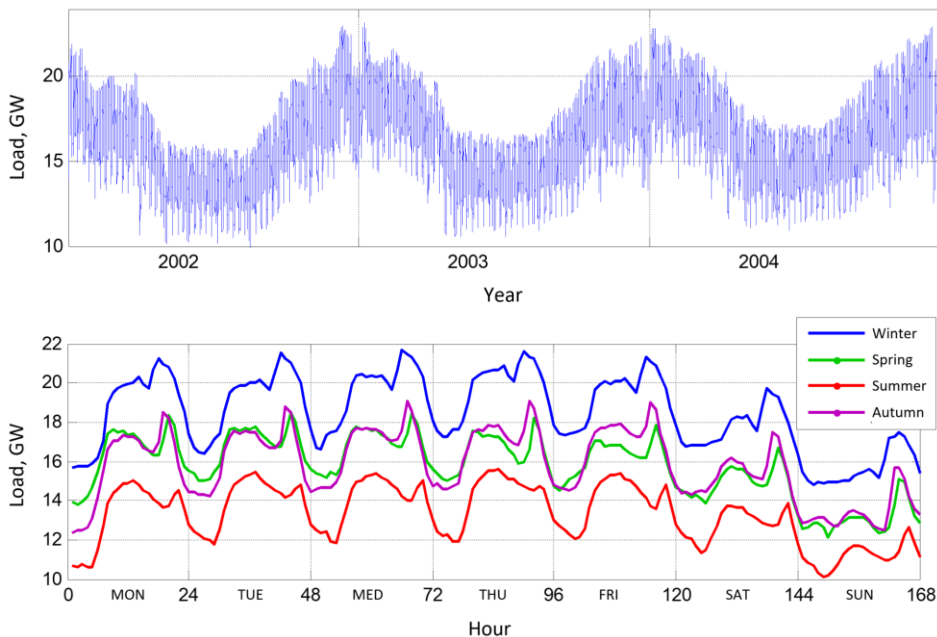


Podójście bezpośrednie *Direct multi-step prediction*



Źródło: <https://cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html>

Predykcja szeregu czasowego z wieloma cyklami sezonowymi w horyzoncie τ na podstawie historii (*univariate approach*)



Obciążenie polskiego systemu elektroenergetycznego z cyklami: rocznym, tygodniowym i dobowym

Metody konwencjonalne

- ARIMA
- Wygładzanie wykładnicze
- Regresja liniowa
- ...

Metody inteligencji obliczeniowej i uczenia maszynowego

- Sieci neuronowe (płytke, głębokie, rekurencyjne)
- Logika rozmyta i systemy neuronowo-rozmyte
- Modele oparte na drzewach decyzyjnych
- Maszyna wektorów nośnych
- Metody grupowania
- ...

Cel

Odfiltrować trend i cykle dłuższe od podstawowego (dobowego), unifikacja danych wejściowych i wyjściowych, sprowadzenie do stacjonarności

Wzorce wejściowe

$\mathbf{x}_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,T}]^T$ reprezentują sekwencje szeregu czasowego $\{P_t\}$ w okresie sezonowym i poprzedzającym okres prognozy: $\mathbf{P}_i = [P_{i,1} \ P_{i,2} \ \dots \ P_{i,T}]^T$

$$x_{i,t} = \frac{P_{i,t} - \bar{P}_i}{\sqrt{\sum_{j=1}^T (P_{i,j} - \bar{P}_i)^2}}$$

Wzorce \mathbf{x}_i są znormalizowanymi wektorami \mathbf{P}_i (mają jednostkową długość, zerową średnią i taką samą wariancję)

Wzorce wyjściowe

$\mathbf{y}_i = [y_{i,1} \ y_{i,2} \ \dots \ y_{i,T}]^T$ reprezentują prognozowane sekwencje szeregu czasowego w okresie $i + \tau$:
 $\mathbf{P}_{i+\tau} = [P_{i+\tau,1} \ P_{i+\tau,2} \ \dots \ P_{i+\tau,T}]^T$, gdzie τ jest horyzontem prognozy

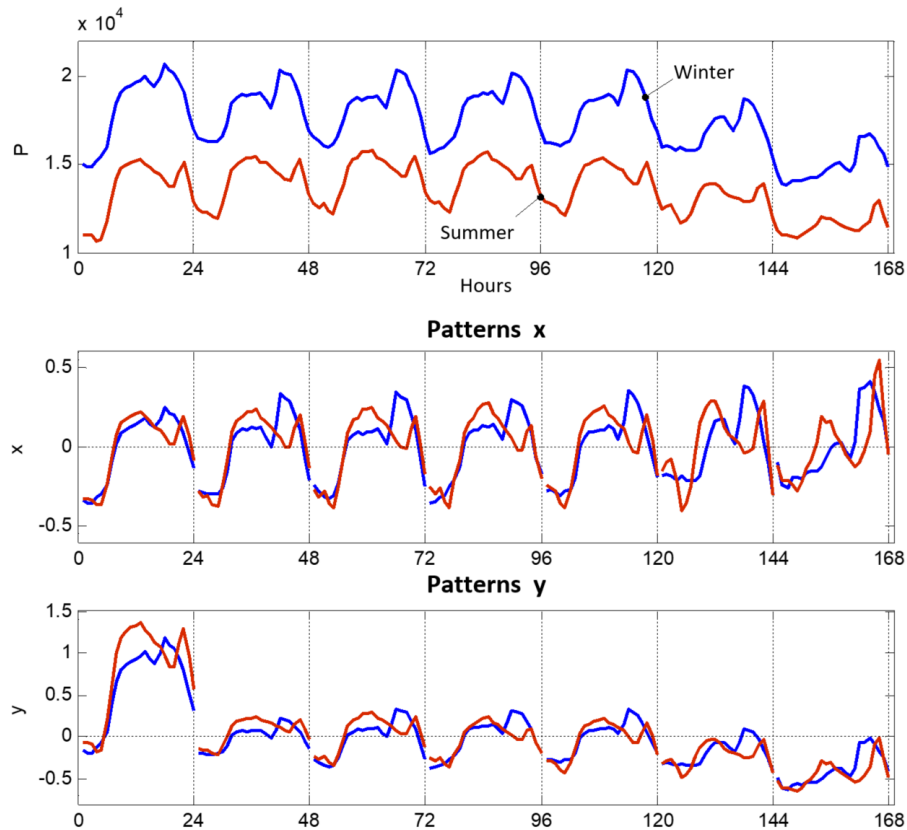
$$y_{i,t} = \frac{P_{i+\tau,t} - \bar{P}_i}{\sqrt{\sum_{j=1}^T (P_{i,j} - \bar{P}_i)^2}}$$

gdzie zmienne kodowe \bar{P}_i i $\sqrt{\sum_{j=1}^T (P_{i,j} - \bar{P}_i)^2}$ wyznaczone są dla okresu i

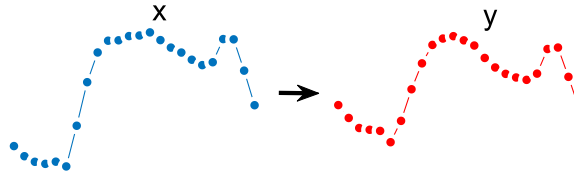
Inne definicje wzorców:

- Dudek G.: **Pattern Similarity-based Methods for Short-term Load Forecasting – Part 1: Principles**. Applied Soft Computing, vol. 37, pp. 277-287, 2015.
- Dudek G. **Similarity-based approaches to short-term load forecasting**. In: Forecasting Models: Methods and Applications. pp. 161-178, iConcept Press 2010.

Reprezentacja szeregów czasowych



Model prognostyczny odwzorowuje $X \rightarrow Y$

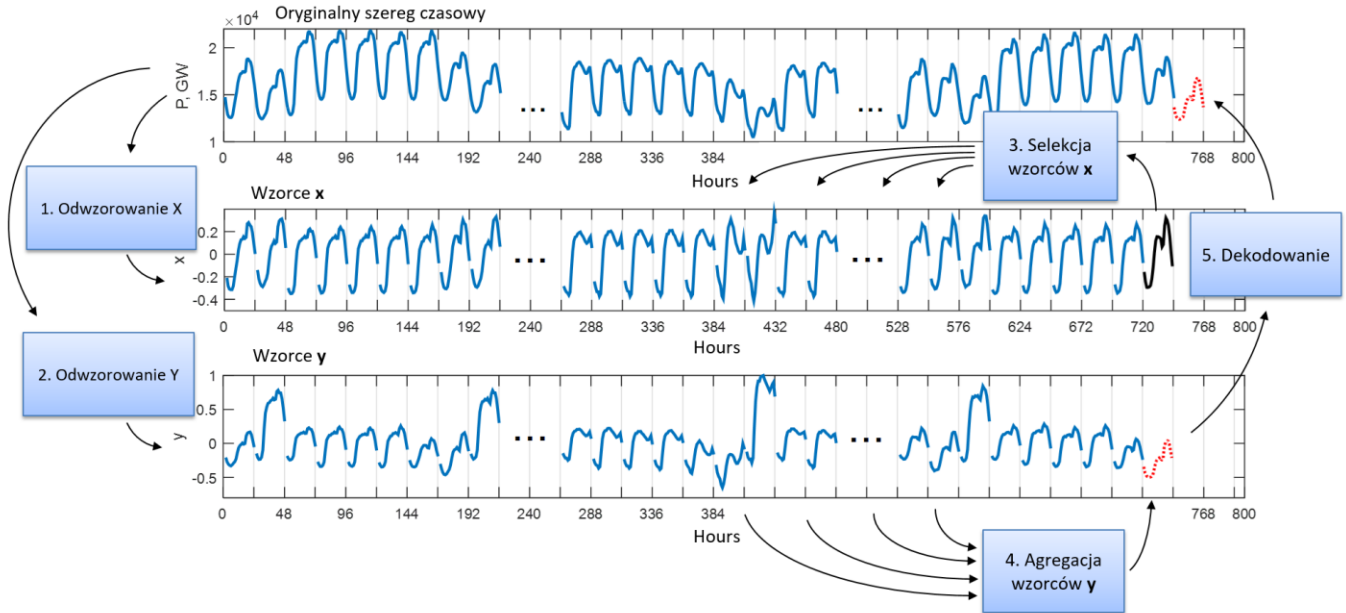


Wyjściem modelu jest prognoza wzorca **y** (lub jego składowej).

Prognozę **P** obliczamy z (dekodowanie):

$$\widehat{P}_{i+\tau,t} = \widehat{y}_{i,t} \sqrt{\sum_{j=1}^T (P_{i,j} - \overline{P}_i)^2} + \overline{P}_i$$

Idea metod opartych na podobieństwie wzorców



Założenie

Jeśli historyczny wzorzec wejściowy \mathbf{x}_a jest podobny do wzorca \mathbf{x}_b reprezentującego okres poprzedzający okres prognozy, to wzorzec wyjściowy \mathbf{y}_a (sparowany z \mathbf{x}_a) jest podobny do wzorca prognozowanego \mathbf{y}_b

Ponieważ wzorce \mathbf{y} są zależne od dnia tygodnia, selekcja podobnych wzorców \mathbf{x} odbywa się w grupach dni tygodnia

Idea metod opartych na podobieństwie wzorców

Weryfikacja założenia

H_0 : Zaobserwowane różnice w liczebnościach występowania jednostek zbiorowości w wyróżnionych kategoriach wartości cech D_x i D_y spowodowane są losowym charakterem próby.
gdzie: D_x – odległość pomiędzy \mathbf{x}_a i \mathbf{x}_b , D_y – odległość pomiędzy \mathbf{y}_a i \mathbf{y}_b

Weryfikacja – test χ^2 test
w oparciu na tablicy
korelacyjnej obrazującej
empiryczny łączny rozkład
cech D_x i D_y .

D_x	D_y							
	0,00-0,47	0,47-0,61	0,61-0,98	0,98-1,36	1,36-1,92	1,92-2,64	2,64-3,78	> 3,78
0,00-0,32	86488	11452	11866	23380	10796	2446	1288	116
0,32-0,44	41892	32916	16838	23344	20522	9802	1846	672
0,44-0,51	15560	48878	23484	23596	20220	12198	2542	1354
0,51-0,59	3440	40054	34316	24282	23038	15496	4062	3144
0,59-0,70	430	13460	37774	21018	27052	28896	11578	7624
0,70-0,83	16	990	16670	13962	21994	38992	32006	23202
0,83-1,00	4	76	5170	10386	12322	23536	42726	53612
> 1,00	2	6	1714	7864	11888	16466	51784	58108

O sile zależności pomiędzy cechami świadczy wartość współczynnika zbieżności V Cramera i współczynnika korelacji liniowej ρ Pearsona

Metody uczenia maszynowego – memory-based models, indukcyjne modele regresyjne uczone pod nadzorem na zbiorze wzorców $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$

Metody regresji nieparametrycznej – funkcja regresji modelowana jest lokalnie w sąsiedztwie wzorca wejściowego \mathbf{x}

$$\mathbf{y} = m(\mathbf{x}) + \xi, \quad m(\mathbf{x}) = \frac{\sum_{i=1}^N w(\mathbf{x}_i) \mathbf{y}_i}{\sum_{i=1}^N w(\mathbf{x}_i)}$$

gdzie: $\xi \sim N(0, \sigma^2)$ – błąd,

$w(\mathbf{x}_i)$ – waga i -tego wzorca \mathbf{y} (zależna od podobieństwa/odległości \mathbf{x} od \mathbf{x}_i),

$m(\mathbf{x}) = E(Y \mid \mathbf{X} = \mathbf{x})$ – krzywa regresji

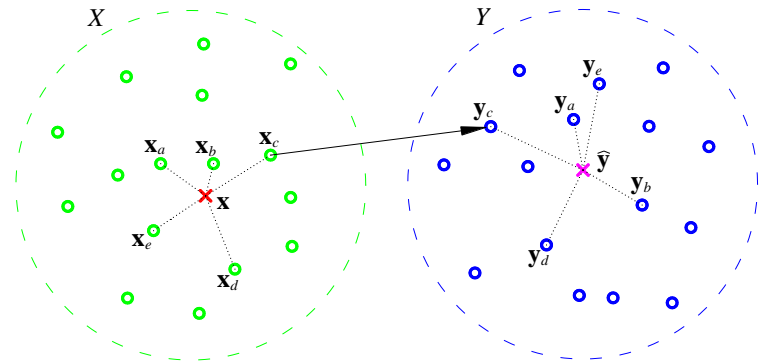
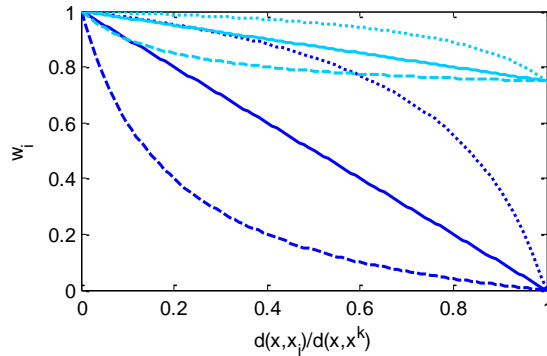
$$m(\mathbf{x}) = \frac{\sum_{i \in \Theta_k(\mathbf{x})} w_i(\mathbf{x}_i) \mathbf{y}_i}{\sum_{i \in \Theta_k(\mathbf{x})} w_i(\mathbf{x}_i)}, \quad w_i(\mathbf{x}_i) = p \left(\frac{1 - \frac{d(\mathbf{x}, \mathbf{x}_i)}{d(\mathbf{x}, \mathbf{x}^k)}}{1 + \lambda \frac{d(\mathbf{x}, \mathbf{x}_i)}{d(\mathbf{x}, \mathbf{x}^k)}} - 1 \right) + 1$$

gdzie:

$\Theta_k(\mathbf{x})$ – zbiór k najbliższych sąsiadów \mathbf{x}

$d(\mathbf{x}, \mathbf{x}^k)$ – odległość pomiędzy \mathbf{x} i jego k -tym sąsiadem \mathbf{x}^k

$p \in [0, 1]$, $\lambda \geq -1$ – parametry



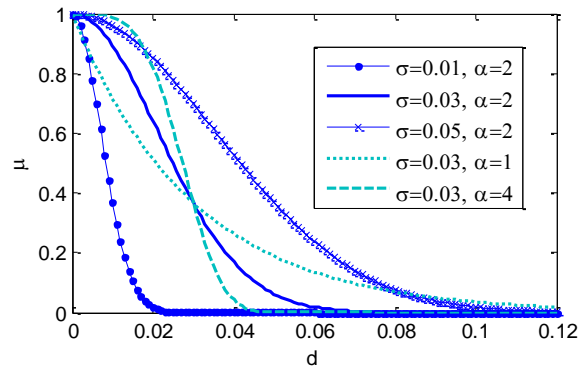
Model rozmytego sąsiedztwa - Fuzzy neighborhood model (FNM)

$$m(\mathbf{x}) = \frac{\sum_{i=1}^N \mu(\mathbf{x}, \mathbf{x}_i) y_i}{\sum_{i=1}^N \mu(\mathbf{x}, \mathbf{x}_i)}, \quad \mu(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\left(\frac{d(\mathbf{x}, \mathbf{x}_i)}{\sigma}\right)^\alpha\right)$$

gdzie:

$\mu(\mathbf{x}, \mathbf{x}_i)$ – funkcja przynależności \mathbf{x}_i do otoczenia \mathbf{x}

$\sigma > 0$, $\alpha \geq 0$ – parametry



Model estymacji jądrowej - Kernel estimation-based model (KEM)

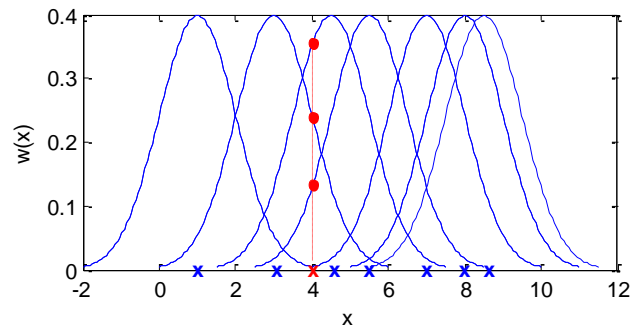
Estymator Nadarayi-Watsona – przypadek wielowymiarowy z jądrami produktowymi:

$$m(\mathbf{x}) = \frac{\sum_{i=1}^N \prod_{j=1}^T K\left(\frac{x_j - x_{i,j}}{h_j}\right) \mathbf{y}_i}{\sum_{i=1}^N \prod_{j=1}^T K\left(\frac{x_j - x_{i,j}}{h_j}\right)}, \quad K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

gdzie:

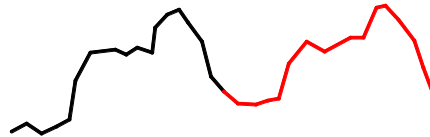
$K(u)$ – jądro normalne,

h_j – szerokość jądra

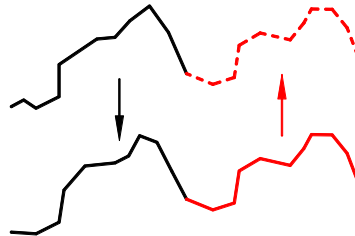


Procedura prognostyczna FP1

1. Połączenie sparowanych wzorców \mathbf{x} i \mathbf{y} we wzorcu \mathbf{z} .

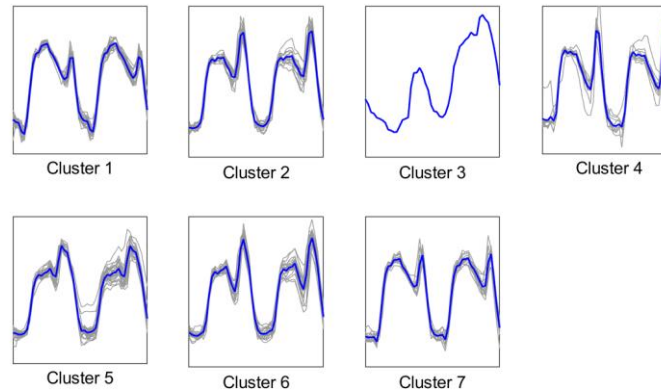


2. Grupowanie wzorców \mathbf{z} .
3. Prezentacja wzorca testowego \mathbf{x}^* i przydzielenie go do najbliższej grupy.
4. Rekonstrukcja obrazu \mathbf{y} sparowanego z \mathbf{x}^* na podstawie środka najbliższej grupy: $m(\mathbf{x}) = \mathbf{m}_{y,i^*}$.

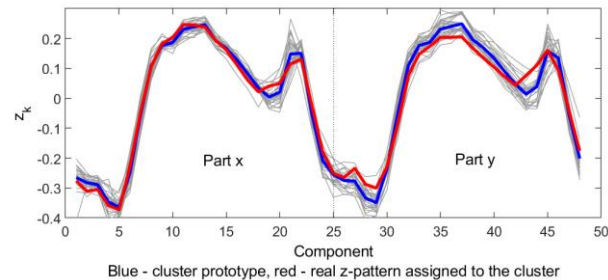


Procedura prognostyczna FP1

Grupy wzorców z



Rozpoznawany wzorec x i zrekonstruowany prognozowany wzorec y



Procedura prognostyczna FP2

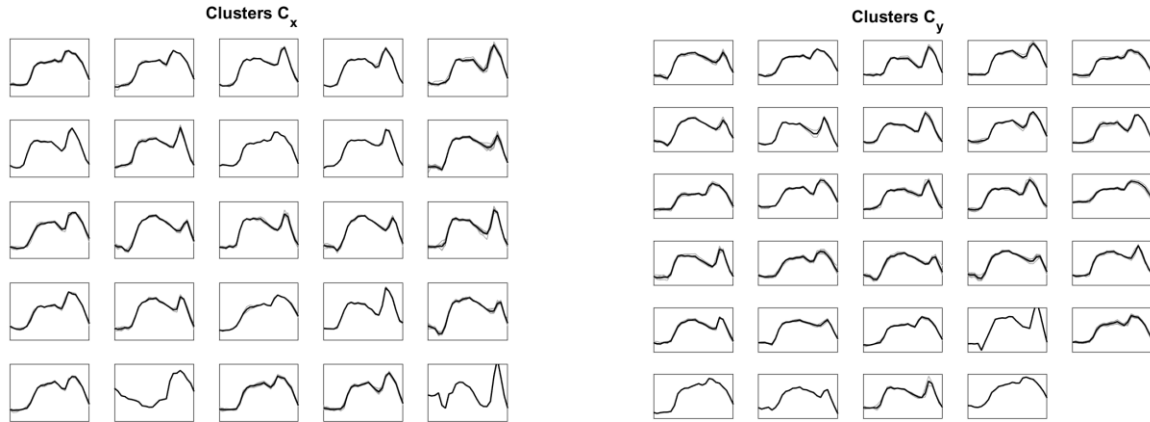
1. Grupowanie wzorców \mathbf{x} .
2. Grupowanie wzorców \mathbf{y} .
3. Estymacja prawdopodobieństw warunkowych $P(C_{y,j} | C_{x,i})$.
4. Prezentacja obrazu testowego \mathbf{x}^* i przydzielenie go do najbliższej grupy C_x .
5. Rekonstrukcja obrazu \mathbf{y} sparowanego z \mathbf{x}^* na podstawie środków grup (centroidów \mathbf{m}) C_y i prawdopodobieństw $P(C_{y,j} | C_{x,i})$

$$m(\mathbf{x}) = \frac{\sum_{l=1}^L P(C_{y,l} | C_{x,i^*}) \mathbf{m}_{y,l}}{\sum_{l=1}^L P(C_{y,l} | C_{x,i^*})}$$

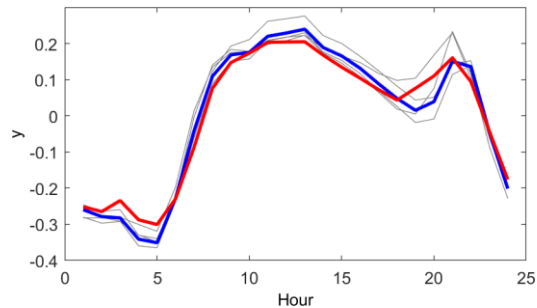
Modele prognostyczne

Modele oparte na grupowaniu wzorców

Procedura prognostyczna FP2



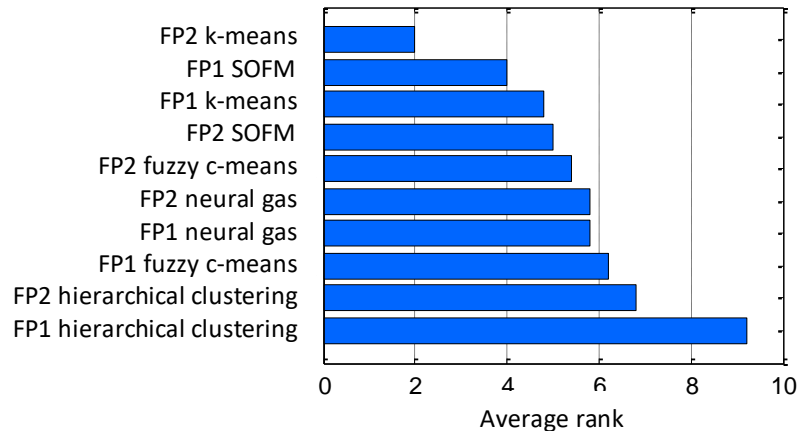
Zrekonstruowany wzorzec y



Gray - prototypes of clusters of non-zero conditional probabilities $P(C_{y,i} | C_{x,k})$,
 blue - forecasted y-pattern, red - real y-pattern

Algorytmy grupowania

- k-means
- fuzzy c-means
- hierarchical clustering
- SOFM - self-organizing feature map
- neural gas

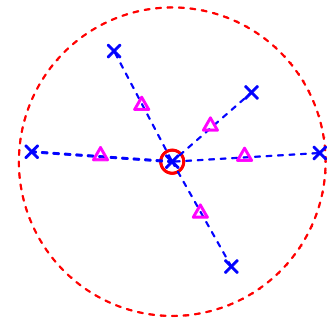


Sztuczny system immunologiczny AIS1

Trening (kreacja pamięci immunologicznej)

1. Załadowanie zbioru antygenów uczących $\mathbf{z} = [\mathbf{x}^T \mathbf{y}^T]^T$.
2. Generacja populacji przeciwciał $\mathbf{v} = [\mathbf{p}^T \mathbf{q}^T]^T$.
3. Wyznaczenie powinowactw przeciwciał względem antygenów.
4. Ocena przeciwciał.
5. Wykonuj do spełnienia warunku stopu (pętla selekcji klonalnej).
 - 5.1. Wykonuj dla każdego przeciwciała.
 - 5.1.1. Klonowanie.
 - 5.1.2. Hipermutacja klonalna.
 - 5.2. Wyznaczenie powinowactw klonów względem antygenów.
 - 5.3. Ocena klonów.
 - 5.4. Selekcja klonalna.

$$v_{k,t}^j = v_{k,t} + \eta_{k,t}^j (z_{j,t} - v_{k,t})$$



Test

6. Prezentacja antygeny testowego z łańcuchem \mathbf{x} i identyfikacja aktywowanych przeciwciał.
7. Wyznaczenie prognozy na podstawie łańcuchów \mathbf{q} aktywowanych przeciwciał.

$$\hat{m}(\mathbf{x}) = \frac{\sum_{k \in \Xi} a(\mathbf{p}_k, \mathbf{x}) \mathbf{q}_k}{\sum_{k \in \Xi} a(\mathbf{p}_k, \mathbf{x})}$$

Sztuczny system immunologiczny AIS2

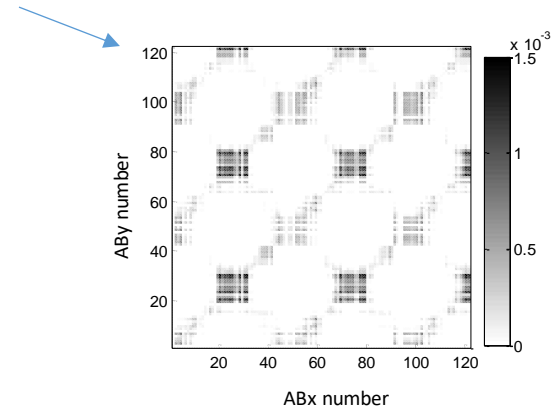
Trening (kreacja pamięci immunologicznej)

1. Załadowanie zbioru antygenów uczących typu x i y .
2. Generacja populacji przeciwciał typu x i y (PCx i PCy).
3. Wyznaczenie promieni krosreaktywności przeciwciał PCx .
4. Wyznaczenie promieni krosreaktywności przeciwciał PCy .
5. Wyznaczenie prawdopodobieństw warunkowych $P(PCy_k | PCx_j)$.

Test

6. Prezentacja antygeny x i detekcja aktywowanych przeciwciał x .
7. Wyznaczenie prognozy na podstawie $P(PCy_k | PCx_j)$, aktywowanych przeciwciał typu x (powinowactw) i przeciwciał typu y .

$$\hat{m}(\mathbf{x}) = \sum_{j=1}^N w_j(\mathbf{x}) \mathbf{q}_j$$
$$w_j(\mathbf{x}) = \frac{\sum_{i \in \Xi} P(ABy_j | ABx_i) a(\mathbf{p}_i, \mathbf{x})}{\sum_{k=1}^N \sum_{i \in \Xi} P(ABy_k | ABx_i) a(\mathbf{p}_i, \mathbf{x})}$$



Sztuczny system immunologiczny AIS3

Trening (kreacja pamięci immunologicznej)

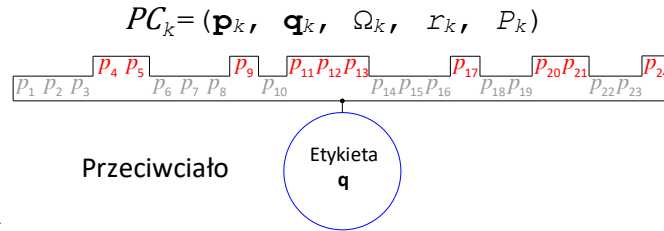
1. Załadowanie zbioru antygenów uczących.
2. Generacja populacji początkowej przeciwciał.
3. Wykonuj dla każdego przeciwciała.
 - 3.1. Wykonuj do spełnienia warunku stopu (selekcja klonalna).
 - 3.1.1. Klonowanie.
 - 3.1.2. Hipermutacja klonalna.
 - 3.1.3. Obliczenie promieni krosreaktywności i powinowactw klonów do antygenów.
 - 3.1.4. Obliczenie wpisów do etykiet klonów.
 - 3.1.5. Ocena klonów.
 - 3.1.6. Selekcja najlepszego klonu i zastąpienie nim przeciwciała macierzystego.

Test

4. Prezentacja antygeny testowego i obliczenie powinowactw przeciwciał.
5. Wyznaczenie wpisu do etykiety antygeny (prognozy).

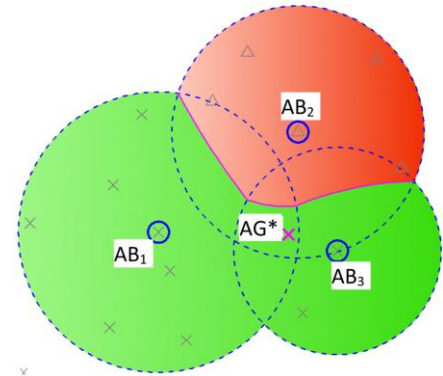
Cechy AIS3

- Przeciwciała – jednostki rozpoznawczo-predykcyjne ze zmiennymi paratopami (lokalna selekcja cech)

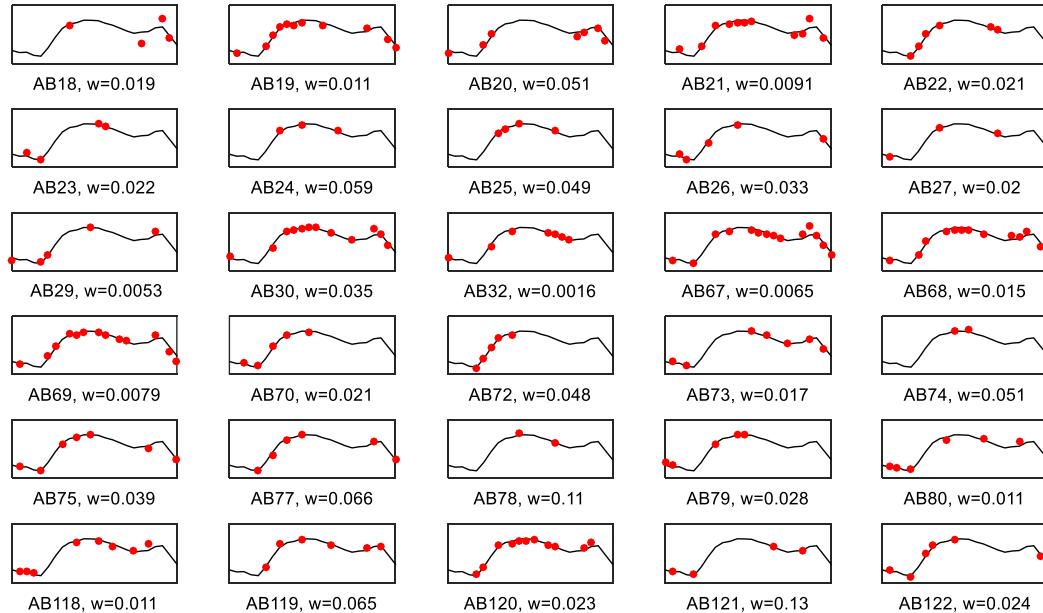


- W procesie kreacji pamięci immunologicznej:
 - kształtowane są paratopy Ω (lokalna selekcja cech)
 - adaptowane są progi krosreaktywności r
 - adaptowane są etykiety \mathbf{q}
- Końcowa decyzja podejmowana jest kolektywnie przez przeciwciała kompetentne w ich obszarach rozpoznania (komitet słabych uczniów)
- Funkcja regresji

$$\hat{m}(\mathbf{x}) = \frac{\sum_{k \in \Xi} P_k a(\mathbf{p}_k, \mathbf{x}, \Omega_k) \mathbf{q}_k}{\sum_{k \in \Xi} P_k a(\mathbf{p}_k, \mathbf{x}, \Omega_k)}$$



AIS3 – proces rozpoznania



Dane

Cztery szeregi czasowe zużycia energii elektrycznej – dane godzinowe (PL) lub półgodzinne (FR, GB, VC), trzyletni okres danych

Dane treningowe – pierwsze 2 lata, dane testowe – ostatni rok

Problem prognostyczny

Prognoza na następne τ dni

Uczenie i optymalizacja modeli

Kroswalidacja (leave-one-out cross-validation), grid search (tournament searching, genetic alg., ...)

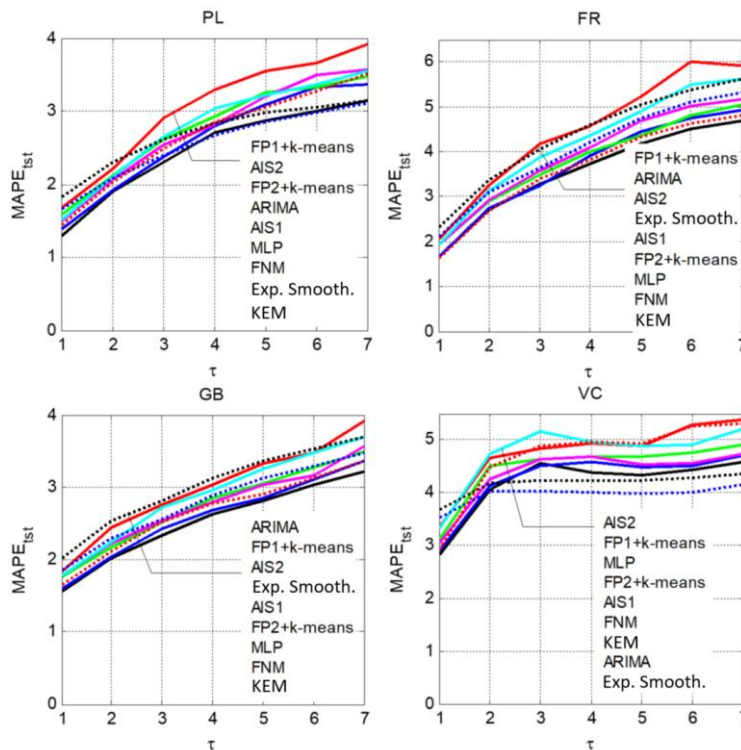
Błąd

$$MAPE = \frac{100}{M} \sum_{j=1}^M \left| \frac{P_j - \hat{P}_j}{P_j} \right|$$

Błędy modeli i ich rozstępy międzykwartylowe dla $\tau = 1$

Model	PL		FR		GB		VC	
	$MAPE_{tst}$	IQR_{tst}	$MAPE_{tst}$	IQR_{tst}	$MAPE_{tst}$	IQR_{tst}	$MAPE_{tst}$	IQR_{tst}
FNM	1.38	1.38	1.67	1.71	1.60	1.66	2.91	2.67
KEM	1.30	1.30	1.66	1.67	1.55	1.63	2.82	2.56
FP1 <i>k</i> -means	1.69	1.64	2.05	2.17	1.84	1.88	3.34	3.01
FP2 <i>k</i> -means	1.59	1.51	1.94	2.05	1.76	1.84	3.13	2.94
AIS1	1.50	1.50	1.93	1.95	1.77	1.84	3.04	2.75
AIS2	1.50	1.51	1.93	1.96	1.78	1.87	3.33	2.93
AIS3	1.51	1.49	1.79	1.81	1.67	1.73	3.13	2.75
ARIMA	1.82	1.71	2.32	2.53	2.02	2.07	3.67	3.42
Exp. Smooth.	1.66	1.57	2.10	2.29	1.85	1.84	3.52	3.35
Naïve	3.43	3.42	5.05	5.96	3.52	3.82	4.88	4.55
MLP+patterns	1.44	1.41	1.64	1.70	1.65	1.70	2.92	2.69

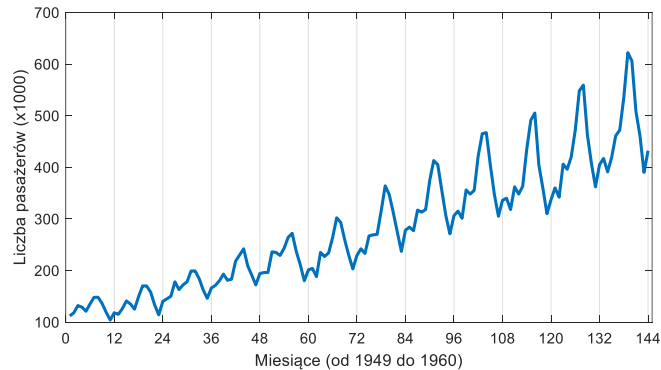
MAPE dla $\tau = 1, 2, \dots, 7$



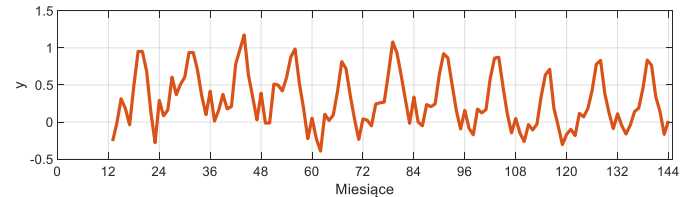
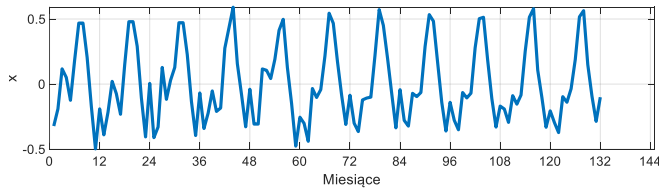
Dane

Liczba pasażerów linii lotniczych – dane miesięczne z okresu od 1949 do 1960

Dane uczące – 1946-1959, dane testowe – ostatni rok



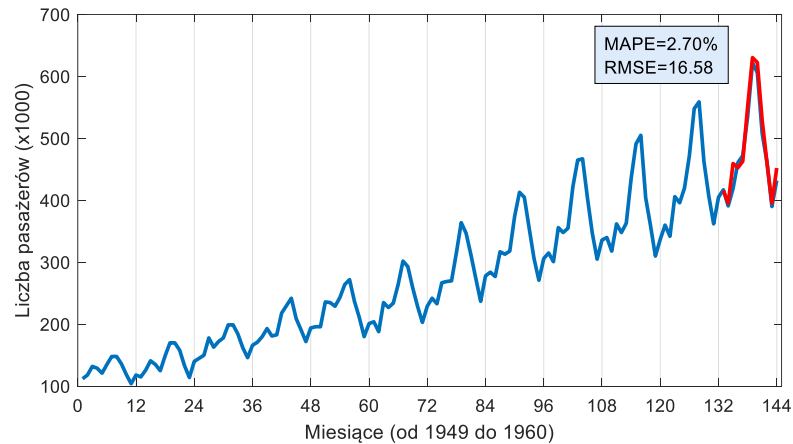
Wzorce x i y



Model rozmytego sąsiedztwa

$$\sigma = 0.09, \alpha = 2, n = 12$$

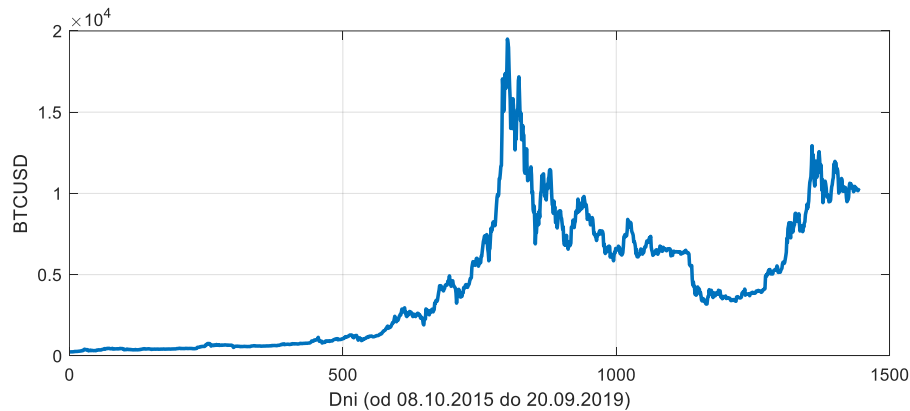
Prognoza



Dane

Dobowy kurs bitcoina w okresie od 08.10.2015 r. do 27.09.2019 r.

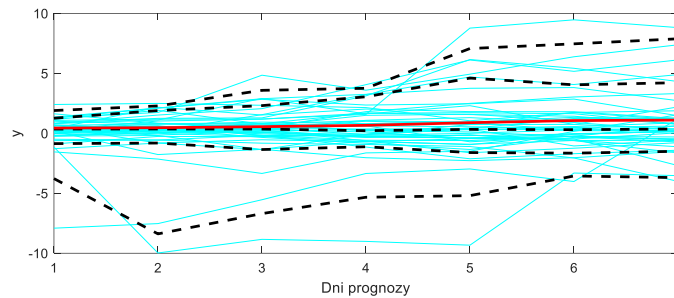
Dane uczące – do 20.09.2019 r., dane testowe – ostatni tydzień



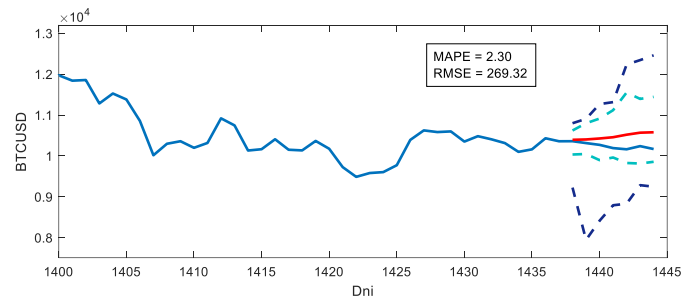
Model rozmytego sąsiedztwa

$$\sigma = 0.6, \alpha = 2, n = 5$$

Wzorce y formujące wzorzec prognozy

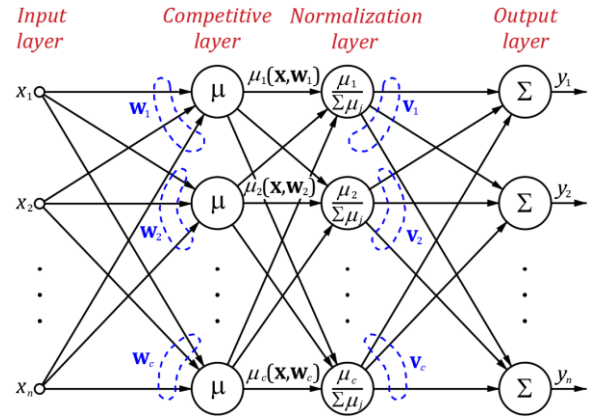


Prognoza



Modele sieci neuronowych oparte na wzorcach

- Wielowarstwowy perceptron (MLP)
- Sieć z radialnymi funkcjami bazowymi (RBFNN)
- Sieć regresji uogólnionej (GRNN)
- Rozmyta sieć z kontrpropagacją (FCPNNs)
- Samoorganizujące się odwzorowanie cech (SOM)



Model	PL		FR		GB		VC	
	MAPE	IQR	MAPE	IQR	MAPE	IQR	MAPE	IQR
MLP	1.45	1.38	1.59	1.64	1.63	1.68	2.99	2.74
RBFNN	1.67	1.53	1.70	1.70	1.84	1.90	3.23	3.05
GRNN	1.38	1.33	1.64	1.71	1.56	1.64	2.83	2.59
FCPNN1	1.71	1.46	1.90	1.95	1.69	1.79	3.18	2.97
FCPNN2	1.63	1.50	1.82	1.86	1.66	1.71	3.22	2.99
SOM1	1.74	1.65	2.10	2.19	1.95	1.98	3.41	3.12
SOM2	1.73	1.53	1.95	2.04	1.78	1.89	3.28	3.08
SOM3	1.99	1.83	2.06	2.18	1.95	2.02	3.63	3.47

Modele regresji liniowej oparte na wzorcach

- Liniowa regresja wieloraka (MLR)
- Regresja krokowa (stepwise regression)
- Regularyzowana regresja najmniejszych kwadratów (ridge regression, lasso)
- Regresja głównych składowych (principal components Regression - PCR)
- Partial Least-Squares Regression (PLSR)

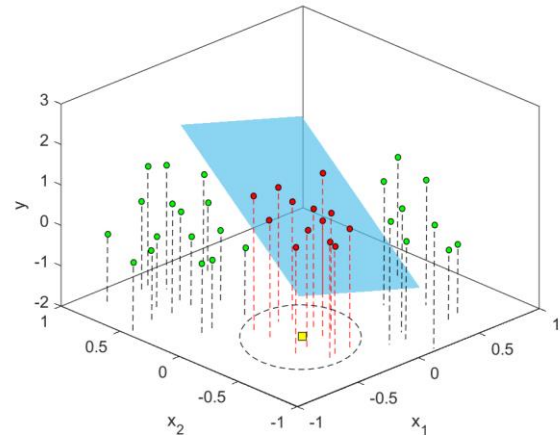


Table 1

Forecast errors and their interquartile ranges in the first example.

Linear model	January		July		Average	
	MAPE _{test}	IQR _{test}	MAPE _{test}	IQR _{test}	MAPE _{test}	IQR _{test}
MLR	2.37	2.44	2.63	2.42	2.50	2.45
Stepwise	1.52	1.44	1.14	1.20	1.33	1.28
Ridge	1.59	1.50	1.23	1.23	1.41	1.29
Lasso	1.51	1.39	1.06	1.02	1.28	1.18
PCR	1.36	1.21	0.94	0.99	1.15	1.09
PLSR	1.18	1.29	1.00	1.03	1.09	1.14

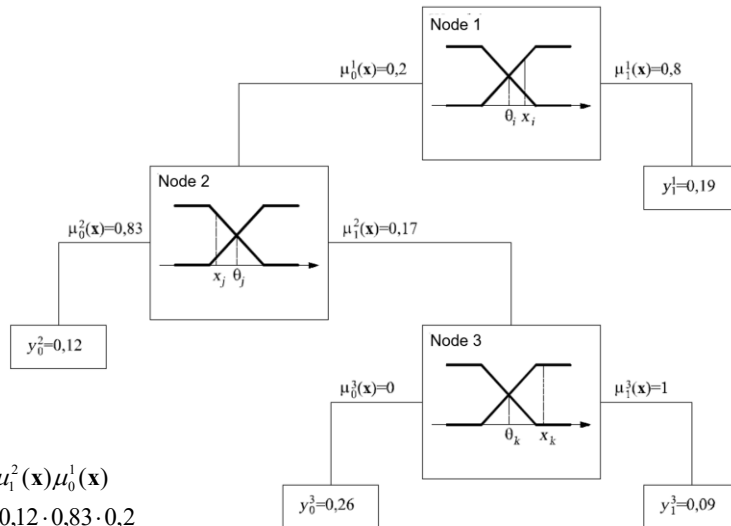
Table 2

Forecast errors and their interquartile ranges (MAPE_{test}/IQR_{test}) in the second example.

Model	PL	FR	GB	VC
PCR	1.35/1.33**	1.71/1.78	1.60/1.68**	3.00/2.70
PLSR	1.34/1.32**	1.57/1.61*	1.54/1.61*	2.83/2.60**
ARIMA	1.82/1.71	2.32/2.53	2.02/2.07	3.67/3.42
ES	1.66/1.57	2.10/2.29	1.85/1.84	3.52/3.35
MLP	1.44/1.41	1.64/1.70**	1.65/1.70**	2.92/2.69
N-WE	1.30/1.30*	1.66/1.67	1.55/1.63*	2.82/2.56*
Naïve	3.43/3.42	5.05/5.96	3.52/3.82	4.88/4.55

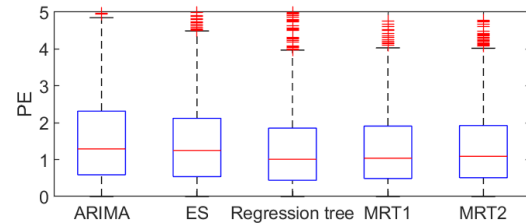
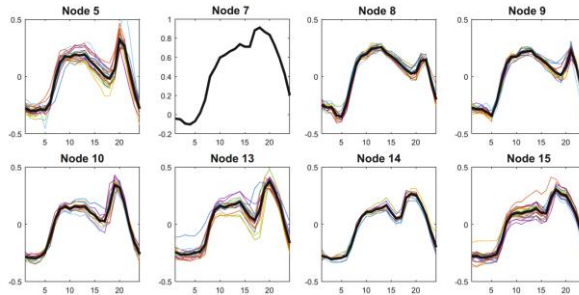
Drzewa regresyjne oparte na wzorcach

- Rozmyte drzewo regresyjne



$$\begin{aligned}
 y(x) &= y_1^1 \mu_1^1(x) + y_0^2 \mu_0^2(x) \mu_0^1(x) + y_0^3 \mu_0^3(x) \mu_1^2(x) \mu_0^1(x) \\
 &\quad + y_1^3 \mu_1^3(x) \mu_1^2(x) \mu_0^1(x) = 0,19 \cdot 0,8 + 0,12 \cdot 0,83 \cdot 0,2 \\
 &\quad + 0,26 \cdot 0 \cdot 0,17 \cdot 0,2 + 0,09 \cdot 1 \cdot 0,17 \cdot 0,2 = 0,175
 \end{aligned}$$

- Wielowymiarowe drzewa regresyjne



- Lasy losowe (random forest)

Model	January		July		Mean	
	$MAPE_{Ist}$	IQR	$MAPE_{Ist}$	IQR	$MAPE_{Ist}$	IQR
RF	1.42	1.39	0.92	0.98	1.16	1.17
CART	1.70	1.58	1.16	1.17	1.42	1.39
Fuzzy CART	1.62	1.47	1.13	1.12	1.37	1.35
ARIMA	2.64	2.34	1.21	1.24	1.91	1.67
ES	2.35	1.88	1.19	1.30	1.76	1.56
ANN	1.32	1.30	0.97	1.01	1.14	1.15
Naïve	6.37	5.36	1.29	1.20	3.78	3.82

Dodatkowe zmienne wejściowe (egzogeniczne)

$$m(\mathbf{x}^*) = \frac{\sum_{t=1}^n \prod_{i=1}^m \mu(\mathbf{z}_i^*, \mathbf{z}_{i,t}) \mu(\mathbf{x}^*, \mathbf{x}_t) \mathbf{y}_t}{\sum_{t=1}^n \prod_{i=1}^m \mu(\mathbf{z}_i^*, \mathbf{z}_{i,t}) \mu(\mathbf{x}^*, \mathbf{x}_t)}$$

$$\mu(\mathbf{z}^*, \mathbf{z}_t) = \exp \left[- \left(\frac{d(\mathbf{z}^*, \mathbf{z}_t)}{\sigma_z} \right)^{\alpha_z} \right]$$

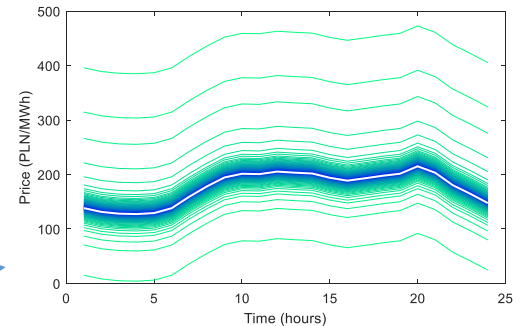
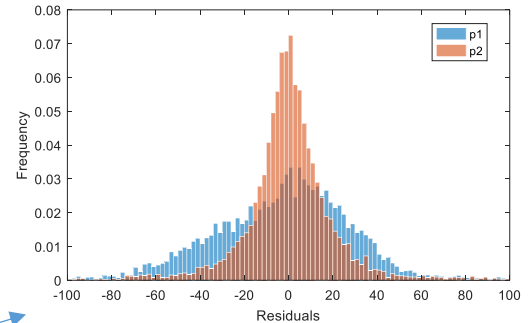
Kontekst \mathbf{z} – wektor zmiennej skorelowanej ze zmienną prognozowaną, np. kontekst temperaturowy może wyrażać temperatury atmosferyczne w okresie dobowym: $\mathbf{z} = [T_1 \ T_2 \ \dots \ T_{24}]$.

Założenie:

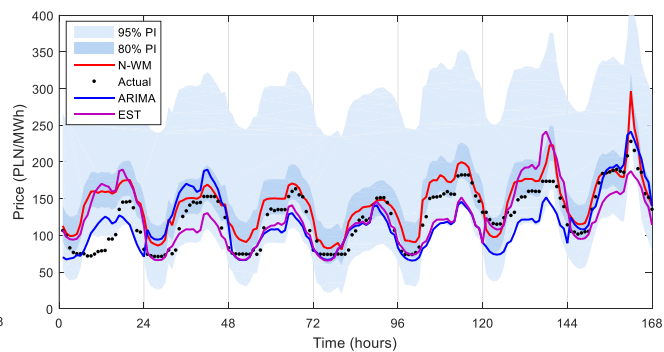
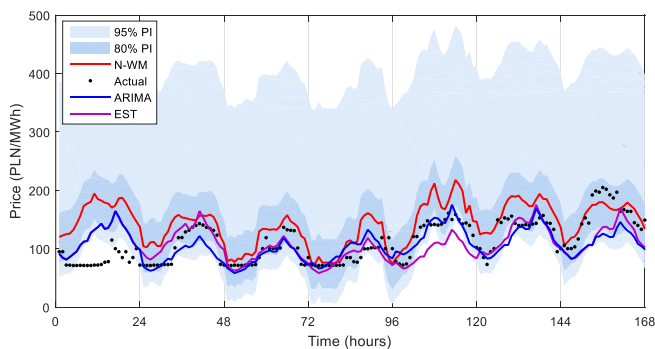
Wzorzec \mathbf{y}_t z kontekstem \mathbf{z}_t , który jest bardziej podobny do kontekstu \mathbf{z}^* prognozowanego wzorca \mathbf{y}^* ma większy wpływ na prognozę (większa waga) niż wzorce o kontekstach mniej podobnych.

Prognozowanie probabilistyczne

1. Model prognostyczny generuje prognozy punktowe.
2. Aby wyznaczyć prognozy probabilistyczne (w kwantylach), prognozyje się dane historyczne.
3. Wyznacza się rozkład reszt modelu i kwantyle dla danych historycznych.
4. Zakłada się, że dla nowych zadań prognostycznych rozkład błędów będzie podobny jak dla danych historycznych.
5. Kwantyle reszt dla $p = 0.01, 0.02, \dots, 0.99$ dodaje się do prognoz punktowych.



Prognozowanie probabilistyczne



Przykłady prognoz dla rynku energii: ceny na rynku bilansującym (po lewej) i na giełdzie energii (po lewej)

- Podejście oparte na reprezentacji szeregów czasowych za pomocą wzorców umożliwia prognozowanie szeregów czasowych ze złożoną sezonowością
- Podejście oparte na reprezentacji szeregów czasowych za pomocą wzorców upraszcza problem prognostyczny – wzorce przenoszą informacje o kształcie cyklu sezonowego, unifikują dane wejściowe i wyjściowe, filtrują trend i dodatkowe cykle sezonowe
- Metody oparte na podobieństwie wzorców konstruuje krzywą regresji agregując wzorce \mathbf{y} z wagami wyrażającymi podobieństwo pomiędzy wzorcami \mathbf{x}
- Modele oparte na podobieństwie wzorców są proste – liczba hiperparametrów jest niewielka, procedura optymalizacji jest prosta, własności generalizacji są lepsze
- Modele oparte na podobieństwie wzorców prognozują od razu cały wektor \mathbf{y} . Liczba wyjść modelu nie wpływa na liczbę jego parametrów, nie komplikuje modelu i problemu optymalizacyjnego
- Proponowane modele są konkurencyjne w dokładności prognoz do innych popularnych modeli prognostycznych (ARIMA, wykładanie wykładnicze, sieci neuronowe)