Szeregi czasowe

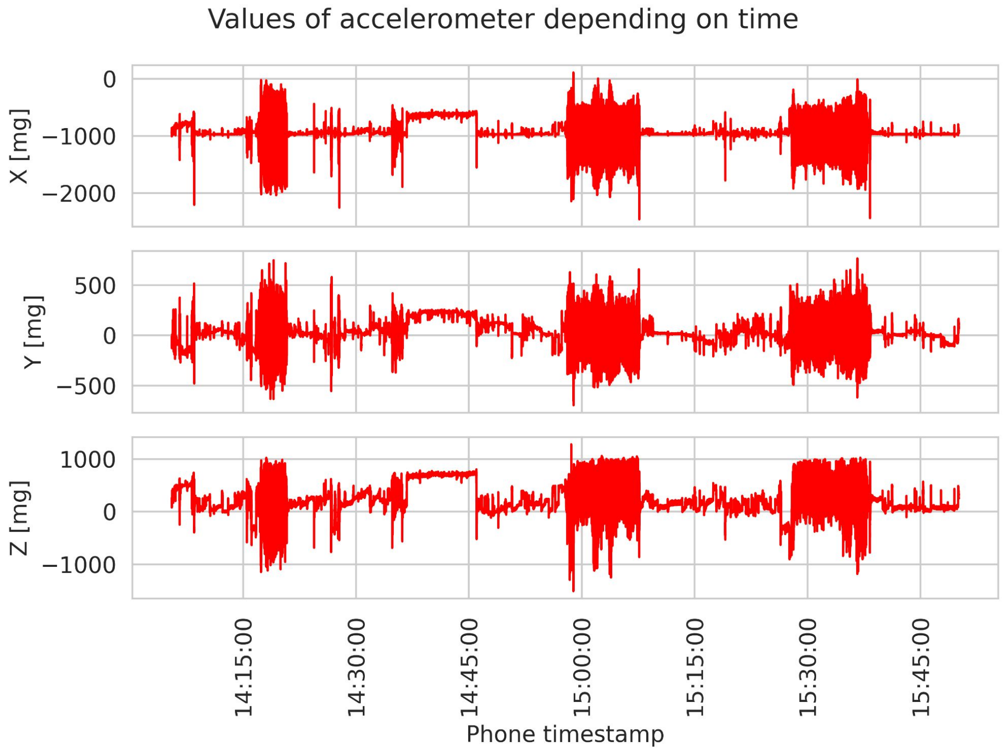
Wykład w ramach przedmiotu “Nauczanie maszynowe”
4 grudnia 2024

Kamil Książek

Uniwersytet Jagielloński

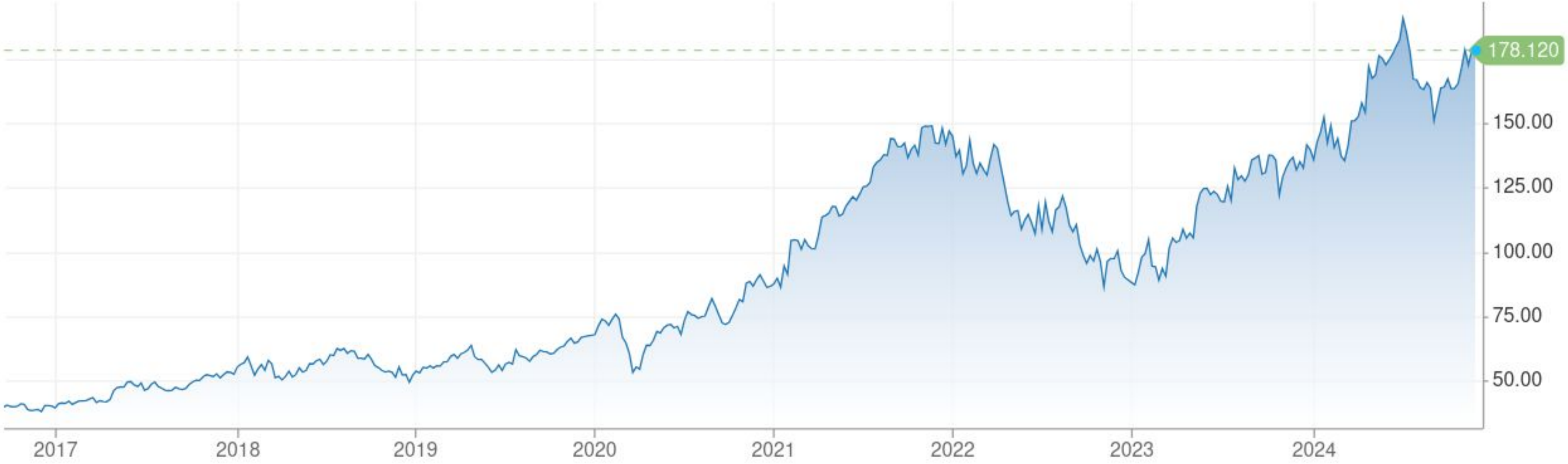
Szereg czasowy

Szeregiem czasowym nazywamy dane zawierające wartości w różnych taktach, najczęściej w równych odstępach czasu. Jeśli w każdym takcie występuje jedna wartość, mówimy o **jednowymiarowym szeregu czasowym** (ang. *univariate time series*), a jeśli wiele wartości, to o **wielowymiarowym szeregu czasowym** (ang. *multivariate time series*).



Zastosowania szeregów czasowych

- **ekonomia i finanse**: dane dot. bezrobocia, zajętości łóżek w szpitalach, kursów walut, cen akcji na giełdzie, itd.;
- **środowisko**: opady deszczu, jakość powietrza, itd.;
- **medycyna**: odczyty EKG, EEG, itd.



Dane giełdowe spółki Alphabet, źródło: CNBC

Zadania dot. szeregów czasowych

group of machine
gmum
learning research

- **prognozowanie przyszłych wartości** (ang. *forecasting*) na podstawie danych historycznych**;**
- **klasyfikacja;**
- **imputacja;**
- **wykrywanie anomalii**

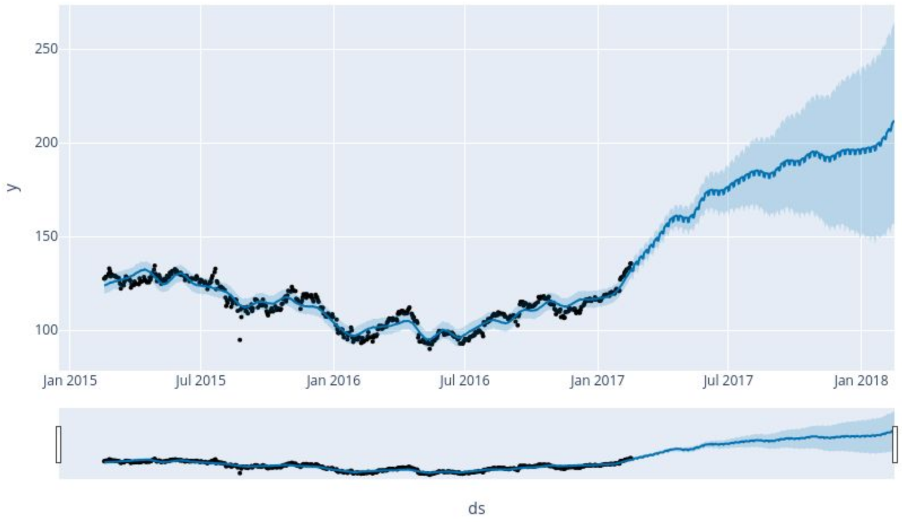
i inne…

Prognoza danych finansowych firmy Apple przy użyciu metody Prophet, źródło: h1ros.github.io

Rozkład szeregu czasowego

Jednym ze sposobów wyrażenia szeregu czasowego jest jego rozkład (**dekompozycja**) na kilka składowych:

- **trend T(t)** – długoterminowe przesunięcia średniej;
- **efekty sezonowe S(t)** – cykliczne fluktuacje związane z godzinami, dniami tygodnia, miesiącami, itd.;
- **cykle C(t)** – inne cykliczne fluktuacje (np. związane z cyklami biznesowymi);
- **rezydua E(t)** – inne losowe lub systematyczne fluktuacje.



Prognoza danych finansowych firmy Apple przy użyciu metody Prophet, źródło: h1ros.github.io

Rozkład szeregu czasowego

group of machine gmum learning research

Najczęściej tworzy się osobne modele do każdej z tych składowych, a na koniec dodaje się je

$$X(t) = T(t) + S(t) + C(t) + E(t)$$

lub mnoży

$$X(t) = T(t) \cdot S(t) \cdot C(t) \cdot E(t)\,.$$

Przyjrzymy się teraz rodzinie modeli **ARMA**, czyli autoregresywnej średniej kroczącej (ang. *autoregressive moving average*), a później przeanalizujemy model **Neural Prophet**.

Prognozowanie przy użyciu rodziny ARMA

group of machine
gmum
learning research

$$\hat{X}(t) = \sum_{i=1}^{N} a(i) \cdot X(t-i) + \sum_{i=1}^{M} \theta(i) \cdot \epsilon(t-i)\,,$$

gdzie

- $\epsilon(t) = X(t) - \hat{X}(t)$ jest błędem prognozowania,
- $\hat{X}(t)$ jest prognozą modelu dla kroku czasowego *t*,
- $X(t)$ jest rzeczywistą wartością modelu dla kroku czasowego *t*,
- *N* i *M* są hiperparametrami,
- $a(i)$ i $\theta(i)$ są wyuczonymi wagami związanymi ze znaczeniem wcześniejszych kroków czasowych.

Pierwsza suma dotyczy *N* przeszłych wartości szeregu czasowego (człon **autoregresywny**), a druga *M* przeszłych błędów prognozy.

Prognozowanie przy użyciu rodziny ARMA

- Model ARMA zakłada **stacjonarność** szeregu czasowego, czyli mającego stałe własności statystyczne w miarę upływu czasu.

- W przeciwnym wypadku może pomóc różnicowanie, które aproksymuje pochodną szeregu czasowego. Niech $X(t)$ będzie szeregiem czasowym mającym *T* obserwacji. Wówczas szereg

$$Y(t) = X(t) - X(t-1), \;\; t \in \{2, \ldots, T\}$$

będzie posiadał *T–1* obserwacji.

Prognozowanie przy użyciu rodziny ARMA

- Przeprowadzenie *n* rund różnicowania przybliża pochodną *n*-tego stopnia szeregu czasowego i eliminuje wielomianowe trendy do *n*-tego stopnia włącznie.
- **Autoregresywny zintegrowany model średniej kroczącej** (ang. *autoregressive integrated moving average*, ARIMA) przeprowadza *n* rund różnicowania, przybliżając stacjonarność szeregu, a następnie stosuje model ARMA.
- **Sezonowy model ARIMA** (ang. *seasonal ARIMA*, SARIMA) dodatkowo modeluje składową sezonową dla wybranej częstotliwości, np. tygodniowej, za pomocą techniki ARIMA.

Neural Prophet

group of machine gmum learning research

- Przyjrzymy się teraz modelowi Neural Prophet zaproponowanym w 2021 roku m.in. przez naukowców ze Stanforda oraz Facebooka [1]:

$$\hat{X}(t) = T(t) + S(t) + E(t) + F(t) + A(t) + L(t)$$

$T(t)$ – trend w chwili *t*,
$S(t)$ – wpływ sezonowości w chwili *t*,
$E(t)$ – wpływ zdarzeń i świąt w chwili *t*,
$F(t)$ – wpływ zmiennych egzogenicznych na przyszłe predykcje,
$A(t)$ – wpływ autoregresji w chwili *t* w oparciu o przeszłe obserwacje,
$L(t)$ – wpływ opóźnionych obserwacji zmiennych egzogenicznych.

[1] O. Triebe et al, NeuralProphet: Explainable Forecasting at Scale, arXiv 2111.15397, 2021.

Modelowanie trendu w Neural Prophet

- Tradycyjne podejście do modelowania trendu polega na wyrażeniu go jako kombinacji tempa wzrostu *k* oraz przesunięcia *m*. Wówczas trend w czasie t_1 wyraża się poprzez przemnożenie tempa wzrostu z różnicą w czasie, zaczynając od punktu t_0 , wraz z przesunięciem o *m*.

$$T(t_1) = T(t_0) + k \cdot \Delta t = m + k \cdot (t_1 - t_0)$$

- W NeuralProphet trend może być modelowany jako funkcja odcinkowo liniowa, w której dobierane jest n_c punktów, pomiędzy którymi parametry zmiany tempa wzrostu są regularyzowane.

Modelowanie sezonowości w Neural Prophet

- Sezonowość w Neural Prophet jest modelowana przy pomocy **szeregów Fouriera**. Dla każdej sezonowości definiowane jest *k* składników sumy w szeregu o periodyczności *p*.

$$S_p(t) = \sum_{j=1}^{k} \left(a_j \cos\left(\frac{2\pi jt}{p}\right) + b_j \sin\left(\frac{2\pi jt}{p}\right)\right)$$

- Przykładowo, dla rocznej sezonowości, przy danych dziennych *p* = 365.25, zaś dla danych tygodniowych *p* = 52.18. Biblioteka automatycznie aktywuje dzienną, tygodniową i roczną sezonowość w zależności od zakresu i częstotliwości danych. Domyślnie, k = 6 dla sezonowości rocznej, k = 3 dla tygodniowej, a k = 1 dla dziennej.

- Szeregi Fouriera produkują gładkie funkcje, które są proste w interpretacji oraz stabilne w treningu. Niestety, modelują tylko **deterministyczną**, stałą w czasie sezonowość.

Autoregresja

- Autoregresja (AR) jest procesem, w którym przyszłe wartości pewnej zmiennej zależą od jej przeszłych wartości. Liczba przeszłych wartości, do których sięgamy, odnosi się do rzędu *p* w modelu AR(*p*). Współczynnik $\theta[i]$ jest modelowany osobno dla każdej wcześniejszej wartości.
- Niech *c* będzie punktem przesunięcia, $\epsilon[t]$ białym szumem. Wówczas:

$$y[t] = \sum_{i=1}^{p} \theta[i] \cdot y[t-i] + c + \epsilon[t].$$

- Parametr ten powinien być dobierany na podstawie przybliżonej długości istotnego kontekstu. W praktyce często jest to dwukrotność horyzontu prognozy *h*.

Modelowanie autoregresji – Deep AR

- Modelowanie autoregresji w Neural Prophet opiera się na AR-Net [2] i przyjmuje kilka wariantów. Jednym z nich jest **Deep AR**, w którym w tym celu trenowany jest perceptron wielowarstwowy.
- Dane jest *p* wcześniejszych obserwacji szeregu czasowego w wektorze $X \in \mathbb{R}^p$, a próbujemy przewidzieć kolejnych *h* wartości $Y \in \mathbb{R}^h$:

$$X = [Y[t-1], Y[t-2], \ldots, Y[t-p]]$$

$$Y = [A^t[t], A^t[t+1], \ldots, A^t[t+h-1]]$$

[2] O. Triebe et al, AR-Net: A simple Auto-Regressive Neural Network, arXiv preprint 1911.12436, 2019.

Modelowanie autoregresji – Deep AR

- Na wejście do sieci dane jest p ostatnich obserwacji. Sieć składa się z *l* warstw ukrytych o wymiarze *d*. Ostatnia warstwa ma na wyjściu h logitów **nieprzekształcanych funkcją aktywacji** oraz **pozbawionych wektora obciążenia** (ang. bias):

$$a_1 = ReLU(W_1 X + b_1)$$
$$a_i = ReLU(W_i a_{i-1} + b_i), \;\; i \in \{2, \ldots, l\}$$
$$Y = W_{l+1} a_l \, ,$$

gdzie $b \in \mathbb{R}^d$, zaś rozmiar warstw ukrytych $W_i \in \mathbb{R}^{d \times d}, i \in \{2, \ldots, l\}$, z kolei $W_1 \in \mathbb{R}^{d \times p}$ oraz $W_{l+1} \in \mathbb{R}^{h \times d}$.

Opóźniony regresor (ang. lagged regressor)

group of machine
gmum
learning research

- Opóźnione regresory są zmiennymi, które wpływają na zmienną docelową w oparciu o ich własne wartości historyczne (opóźnione), a nie przyszłe wartości.
- Przykłady:
  – Zapotrzebowanie na energię w zależności od temperatury w poprzednich dniach.
  – Ceny jednych towarów w zależności od innych, np. ceny ropy z poprzednich kilku dni/tygodni.
- W Neural Prophet tworzony jest osobny moduł pod każdą z tych zmiennych towarzyszących i stosowany jest model AR-Net.

Przyszły regresor (ang. future regressor)

group of machine gmum learning research

- Przyszłe regresory są dodatkowymi zewnętrznymi cechami, które wpływają na wartość przewidywanej zmiennej. W przeciwieństwie do opóźnionych regresorów, **ich wartości muszą być znane z góry (lub przewidziane)** dla wszystkich kolejnych kroków czasowych, dla których chcemy dokonać predykcji.

- Przykład: nadchodzące warunki pogodowe (temperatura, opady deszczu).

Wpływ wydarzeń i świąt

group of machine
gmum
learning research

- Wybrane wydarzenia i święta, choć zachodzą sporadycznie, mają wpływ na wartość przewidywanej zmiennej.

- Mogą one zostać oznaczone zmienną binarną lub różnymi wartościami w przedziale $[t_e - n, t_e + m]$ wokół zdarzenia t_e przez $n + m$ dni. Przykładowo, przed świętami Bożego Narodzenia można ustawić okno $[-1, 0]$, dzięki któremu każdy kolejny dzień będzie miał wyrażony inny wpływ świąt na wynik.

Metryki

- Jedną z częściej stosowanych metryk do oceny błędu regresji jest **pierwiastek błędu średniokwadratowego** (ang. *root mean square error, RMSE*):

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}\ ,$$

gdzie *N* jest liczbą próbek, y_i jest rzeczywistą wartością i-tej próbki, zaś $\hat{y}_i$ predykcją modelu dla i-tej próbki.

Wśród innych można wymienić np. *mean absolute percentage error (MAPE)* lub *mean absolute deviation (MAD).*

group of machine
gmum
learning research

Tyle teorii, teraz spróbujmy zmierzyć się z rzeczywistym zbiorem danych i treningiem sieci neuronowej...

Źródła:

- G. E. P. Box i in., Time Series Analysis, Forecasting and Control, Fifth Edition, Wiley Series in Probability and Statistics, 2016.
- R. Weber, materiały do kursu szeregów czasowych, University of Cambridge, 2017.
- A. Geron, Uczenie maszynowe z użyciem Scikit-Learn, Keras i TensorFlow. Wydanie III, Helion, 2023.
- O. Triebe et al, NeuralProphet: Explainable Forecasting at Scale, arXiv 2111.15397, 2021.
- O. Triebe et al, AR-Net: A simple Auto-Regressive Neural Network, arXiv preprint 1911.12436, 2019.
- Dokumentacja Scikit-learn.