

Analiza szeregów czasowych

Szereg czasowy to zbiór wartości zarejestrowanych w funkcji czasu $y=f(t|x)$, często szereg czasowy może być również funkcją innych parametrów (oznaczono je jako x).

- 1) Wczytaj plik *electricbill.mat* za pomocą polecenia LOAD. Wczytanie dane zawierają dwie tabele *dat* – zawierającą dane oraz *names* zawierającą nagłówki kolumn. Wczytany zbiór danych stanowi przykład szeregu czasowego opisującego wielkość rachunków za energię elektryczną wyrażoną w \$ wraz z dodatkowymi zmiennymi. Są nimi:

Opis	Nazwa
Numer obserwacji	Num
Rok	Year
Miesiąc	Month
Wielkość rachunku miesięcznego za energię elektryczną wyrażona w \$	Bill
Średnia temperatura za miesiąc	Temp
Liczba dni z dogrzewaniem	HDD
Liczba dni z chłodzeniem	CDD
Liczba członków rodziny	Size
Nowy miernik (zmienna binarna)	Meter
Nowa pompa ciepła 1 (zmienna binarna)	Pump1
Nowa pompa ciepła 2 (zmienna binarna)	Pump2
Łączna opłata (za kWh)	RIDER TOTAL
Obliczona konsumpcja w kWh	Consumption

- 2) Jeśli interesuje się określona kolumna to możesz się do nie odwołać wykonując polecenie:

```
id = strcmpi(names, nazwa_kolumny)  
np. id=strcmpi(names, 'bill') co spowoduje wyznaczenie indeksu kolumny zawierającej informacje o wielkości rachunku.  
A odczytanie kolumny zawierającej temperaturę w poszczególnych miesiącach jako:  
temp = dat(:, strcmpi(names, 'Temp')) ;
```
- 3) Wyodrębnij kolumny Num, Year, Bill, a wczytane dane przedstaw na rysunku/rysunkach. W tym celu wykorzystaj funkcję „plot”. Pamiętaj że jeśli chcesz umieścić kilka wykresów na jednym rysunku to musisz wykonać polecenie `hold on`,
Jako x w wykresach wykorzystaj kolumnę *Num*
- 4) Wyznacz pochodną z wielkości rachunku. Zastanów się jak zinterpretować wykres, co na tej podstawie możemy powiedzieć. Zwróć uwagę na znaki pochodnej.
Na podstawie rys 3 i 4 zastanów się i odpowiedz:
Czy wielkość rachunku ma okres?
Jeśli tak to jaki to okres?
Jaki jest charakter szumu. Addytywny czy multiplikatywny, zastanów się czy to widać?
- 5) Spróbuj podzielić wykres na okresy 12miesięczne i narysuj każdy rok wspólnie na jednym wykresie.
Czy występuje tutaj powtarzalność pomiędzy poszczególnymi latami?
- 6) Analizując wykres z zadania 3 zwróć uwagę że w wykresie brakuje jednej wartości. Spróbuj ją uzupełnić.

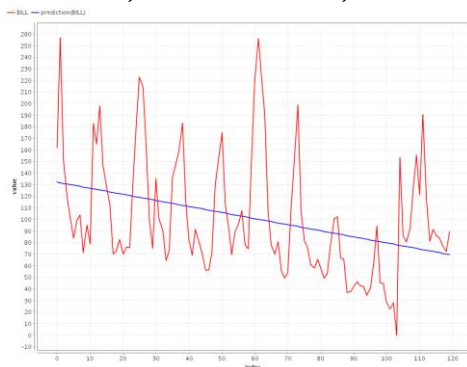
Jaką najprostszą metodą można tutaj uzupełnić brakującą wartość, pamiętając, że jest to szereg czasowy.

UWAGA Wartość brakująca oznaczona jest przez NaN, w celu jej identyfikacji skorzystaj z funkcji `isnan(bill)`, która zwraca binarny indeks zawierający `true` tam gdzie `bill` przyjmuje wartość NaN.



Jak należałoby uzupełnić wartość jeśli odstępy czasowe pomiędzy poszczególnymi próbkami byłyby różne? (obecnie są one wykonywane co 1 miesiąc, a co jeśli byłyby one w różnych odstępach czasu – co miesiąc, co 2 miesiące, co 1.5 miesiąca)

- 7) Wygładzanie i trend. Typowym problemem w analizie serii danych jest wyznaczenie linii trendu. Linia trendu pokazuje nam tendencję tj. czy średnio wartości wzrastają czy maleją. Do wyznaczenia linii trendu wykorzystaj poznaną na poprzednich zajęciach funkcję `polyfit()`, przy rzędzie wielomianu = 1. Uzyskany wynik nanieś na wykres. Do narysowania linii trendu wystarczy że podasz jego wartość w pierwszej i ostatniej chwili czasowej.



- 8) Narysuj linie trendu dla danych składających się z różnych czasookresów tj. np. od lutego 1994 do września 1996 itp.
Czy wybranie dowolnego czasookresu i wyznaczenie dla niego linii trendu ma sens?
Co powinien uwzględniać badany czasookres?
- 9) Zwróć uwagę że wykres `bill` zawiera jedną nietypową wartość (wartość odstającą) spróbuj ją odnaleźć i wygładzić.
- 10) Przewidywanie szeregów czasowych
Częstym problemem jest zbudowanie modelu predykcyjnego, który pozwoliłby przewidywać przyszłe wartości szeregu czasowego. NP. przewidywać zmiany rachunku, przewidywać zużycie energii itp.
W tym celu konieczne jest tzw okienkowanie, napisz więc kod funkcji `windowing`, która pozwalającą na stworzenie zbioru uczącego, która dla zadanego szeregu podzieli go na okna o określonym rozmiarze i na ich podstawie stworzy zbiór uczący.

Następnie otwórz funkcję `validateSeries` w celu przejrzysz jej kod. Spróbuj go zrozumieć. Wykorzystaj ją w celu oceny jakości modelu predykcyjnego.

Dokumentację tej funkcji znajdziesz w helpie.

Uruchom skrypt `validateSeriesScript` który uruchamia funkcję `validateSeries`, rysuje wynik predykcji na ekranie i jednocześnie pozwala na zbadanie wpływu rozmiaru kroku oraz liczby neuronów na uzyskane wyniki. Dobierz najlepszą sieć neuronową najtrafniej przewidywająca wyniki.

Jaka to sieć i jakie ma parametry.?