

Przetwarzanie i Analiza Danych - FAQ

Marcin Pietrzykowski

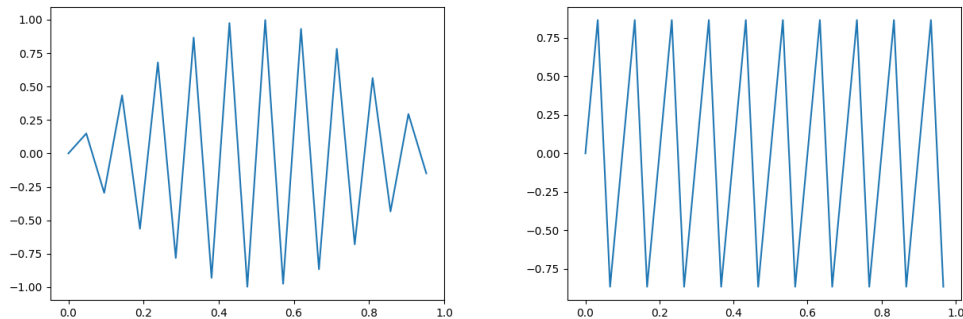
ver. 1.03

Streszczenie

W pliku znajdują się najczęściej zadawane pytania do zadań laboratoryjnych z przedmiotu Przetwarzanie i Analiza Danych. Wraz z prowadzeniem laboratoriów plik będzie systematycznie uzupełniany.

Laboratorium 4 - dyskretyzacja, binaryzacja, kwantyzacja

- Jak powinny wyglądać wykresy zawierające próbkowany sygnał (pkt. 2.3)?



Rysunek 1: (a) $f = 10\text{ Hz}$, $F_s = 21\text{ Hz}$; (b) $f = 10\text{ Hz}$, $F_s = 30\text{ Hz}$.

- **Co powinien przedstawiać obrazek aby przedstawiał „niekształcenia” spowodowane błędnie dobraną częstotliwością (pkt. 2.6)?**
Może między innymi przedstawiać regularny powtarzalny wzór np. dużą ścianę z cegieł (jak w starej kamienicy) lub wieżowiec z dużą liczbą okien.
- **Co powinna zawierać funkcja, która dokona progowania na podstawie histogramu (pkt. 4.3)?**

Progowania można dokonać przy pomocy metody Otsu: https://en.wikipedia.org/wiki/Otsu%27s_method. Można napisać również własną heurystykę np. bazującą na algorytmach poszukiwania minimum funkcji.

Laboratorium 6 - PCA

- Algorytm

1. Normalizacja/standaryzacja zmiennych

$$\bar{x} = x - \text{mean}(x) \quad (1)$$

2. Obliczenie macierzy kowariancji (Dla zbioru Irys macierz kowariancji powinna mieć rozmiar 4×4)

$$\text{cov}(x) = \frac{1}{K} \sum_i^K (x_i - \bar{x})(x_i - \bar{x})^T \quad (2)$$

3. Rozkład macierzy kowariancji na wektory i wartości własne(eig).

$$[w, v] = \text{eig}(\text{cov}(x.T)) \quad (3)$$

4. Sortowanie wartości własnych malejąco

$$v \leftarrow \text{sort}(w) \quad (4)$$

5. Wybór p znaczących wektorów własnych v na podstawie dominujących wartości w

6. Projekcja do nowej podprzestrzeni (funkcji `np.dot`)

$$x \cdot v[:, 1 : p] \quad (5)$$

- Czy podczas implementacji można korzystać z wbudowanych metod takich jak `cov` czy `eig`?

Tak

- Wykres, który otrzymuje posiada odwrócone osie w porównaniu do rezultatu otrzymanego przy pomocy `sklearn.decomposition.PCA`. Co jest tego przyczyną?

Otrzymany wynik jest poprawny. Różnica jest rezultatem tego, że podczas obliczania wartości i wektorów własnych możemy obliczyć lewy wektor własny macierzy spełniający równość:

$$\mathbf{X_L} \mathbf{A} = \lambda_L \mathbf{X_L}, \quad (6)$$

lub prawy wektor własny spełniający równość:

$$\mathbf{A} \mathbf{X}_R = \lambda_R \mathbf{X}_R. \quad (7)$$

Aby otrzymać rezultat zbliżony z wynikiem metody `sklearn.decomposition.PCA`, należy użyć następującej funkcji:

```
w, v = np.linalg.eigh(x_cov, UPLO='U')
```

- **W jaki sposób można dokonać w alg. PCA transformacji odwrotnej?**

Należy wykonać ją w następujący sposób:

$$X_{org} = X_{PCA} \cdot w^T, \quad (8)$$

gdzie w jest macierzą wektorów własnych posortowanych względem odpowiadających im wartością własnym.

- **Po co funkcja `dot` w zadaniu 1(c)?**

Do wygenerowania punktów układających się w podobny kształt jak na Rys. 1. Można to wykonać mnożąc dwie macierze przy pomocy funkcji `dot` np.:

$$X = a_{2 \times 2} \cdot b_{2 \times 100}. \quad (9)$$

- **W jaki sposób wykonać wizualizację podobną do Rys. 1?**

- Punkty zielone - punkty danych X wygenerowane np.: tak jak w punkcie powyższym.
- Punkty czerwone - punkty danych X na których dokonano transformacji PCA (do jednej składowej głównej), a potem dokonano transformacji odwrotnej.
- Wektory - wizualizacja wektorów własnych. Początek wektora jest to średnia wartość zbioru, długość wektora to $3 \cdot \sqrt{\text{wartość własna}}$, a położenie jego grota to początek wektora + wektor własny macierzy.

Laboratorium 11

- `GridSearchCV` - ustawienie parametru `n_jobs = -1` może nigdy się nie zakończyć.