LAB. 2. Analiza systemów statycznych.
Wybór zmiennych objaśniających metodą Hellwiga

Z wstępnie określonego (licznego) zbioru zmiennych objaśniających należy wybrać te, które w sposób optymalny opisują modelowane zjawisko.

W modelu powinny znaleźć się zmienne objaśniające silnie skorelowane ze zmienną objaśnianą i jednocześnie stosunkowo słabo skorelowane między sobą.

Z pierwotnego zbioru zmiennych objaśniających należy dokonać wyboru tych, które w efekcie końcowym wniosą najwięcej informacji o modelowanym zjawisku.

Metoda Hellwiga

1. Utworzenie zbioru, którego elementami są zestawy zmiennych objaśniających otrzymane w wyniku kombinacji zbioru początkowego. Liczba elementów zbioru: $2^n - 1$, gdzie n to początkowa liczba zmiennych objaśniających.

2. Dla każdej zmiennej w każdej kombinacji obliczenie indywidualnej pojemności nośnika informacji h_{kj}:

$$h_{kj} = \frac{r_{0j}^2}{\sum |r_{kj}|},$$

gdzie:
r_{0j} — współczynnik korelacji j–tej zmiennej objaśniającej ze zmienną objaśnianą,
r_{kj} — współczynnik korelacji j–tej zmiennej objaśniającej z pozostałymi zmiennymi objaśniającymi występującymi w k–tej kombinacji.

3. Obliczenie integralnej pojemności nośników informacji H_k dla k–tej kombinacji zmiennych objaśniających:

$$H_k = \sum h_{kj}.$$

4. Zmiennymi objaśniającymi do modelu będą zmienne tej kombinacji, której integralna pojemność nośnika w informacji ma wartość maksymalną.

Przykład

Pierwotny zbiór zmiennych objaśniających ($n = 2$):

$$X = \{x_1, x_2\}.$$

Na podstawie zebranych wcześniej danych mogą zostać wyznaczone współczynniki korelacji między zmiennymi objaśniającymi R oraz wektor R_0, który przedstawia stopień skorelowania zmiennej objaśnianej ze zmiennymi objaśniającymi:

$$R = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix} \qquad R_0 = \begin{pmatrix} 0.8 \\ -0.2 \end{pmatrix}$$

Liczba kombinacji zmiennych objaśniających wynosi $2^n - 1 = 2^2 - 1 = 3$:
$K_1 = \{x_1\}$,
$K_2 = \{x_2\}$,
$K_3 = \{x_1, x_2\}$.

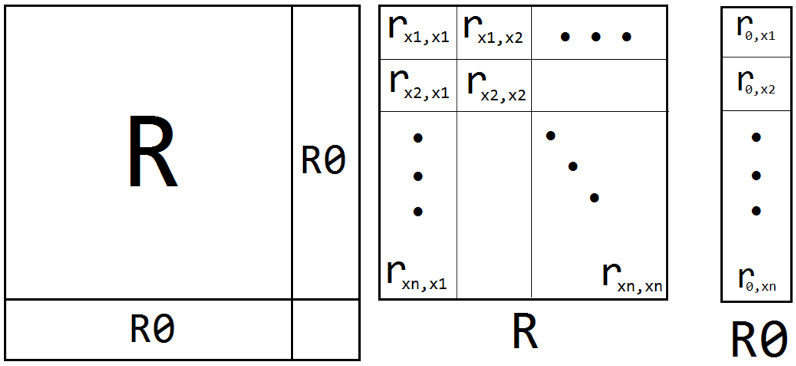
Obliczenie indywidualnej pojemność nośnika informacji dla każdej ze zmiennej j występującej w poszczególnych kombinacjach.
$\rightarrow$ Kombinacja $K_1 = \{x_1\}$:
$r_{01} = 0.8$, $r_{11} = 1$,
$h_{11} = \frac{0.8^2}{1} = 0.64$,
$H_1 = \mathbf{0.64}$.
$\rightarrow$ Kombinacja $K_2 = \{x_2\}$:
$r_{02} = -0.2$, $r_{22} = 1$,
$h_{22} = \frac{(-0.2)^2}{1} = 0.04$,
$H_2 = \mathbf{0.04}$.
$\rightarrow$ Kombinacja $K_3 = \{x_1, x_2\}$:
$r_{01} = 0.8$, $r_{02} = -0.2$,
$h_{31} = \frac{0.8^2}{1+0.6} = 0.4$,
$h_{32} = \frac{(-0.2)^2}{1+0.6} = 0.025$,
$H_3 = h_{31} + h_{32} = 0.4 + 0.025 = \mathbf{0.425}$

Największą wartość pojemności integralnej ma kombinacja $K_1 = \{x_1\}$ i wynosi $H_1 = 0.64$. Oznacza to, że pierwotny zbiór zmiennych objaśniających $X_\{x_1, x_2\}$ został zredukowany do $X = \{x_1\}$.

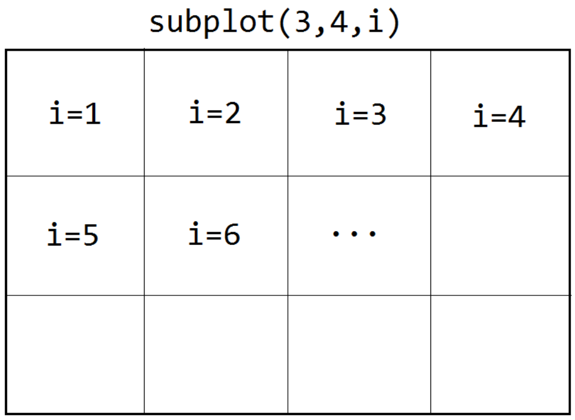
Zadanie
Wyznacz zredukowaną liczbę zmiennych objaśniających dla poniższych danych pomiarowych.

x_1	x_2	x_3	x_4	y
6	0.4	15	70	14
14	0.6	13	40	17
17	0.4	15	80	14.5
14.5	0.7	11	50	20
20	1	10	40	21.6
21.6	1.2	10	50	23
23	1	7	80	24.5
24.5	1.5	6	100	28
28	1.5	8	110	26.4
26.4	1.7	5	80	29

1. (1 punkt) Wyznacz R oraz R_0 wykorzystując funkcje *corrcoef*.



2. (2 punkty) Narysuj wykresy punktowe (w jednym oknie - funkcja *subplot*) poszczególnych zmiennych objaśniających między sobą (x_1, x_2), (x_1, x_3) ... oraz między zmienną objaśnianą (x_1, y), (x_2, y) .... W podpisie wykresu wstaw wartość współczynnika korelacji (funkcja *title*).



np. aby narysować i podpisać wykres przedstawiający (x_1, x_2):

```
subplot(3,4,1)
plot(X(:,1), X(:,2), '*r')
title(['Korelacja x1 z x2 wynosi: ' num2str(R(1,2))])
```

3. (6 punktów) Zapisz wszystkie możliwe kombinacje zmiennych objaśniających i wyznacz indywidualne i integralne pojemności nośników informacji.
np. aby wyznaczyć indywidualne pojemności i integralną pojemność dla kombinacji piątej ($k = 5$) zmiennych (x_1, x_2) należy:

- obliczyć indywidualną pojemność h_{51}:

$$h_{51} = \frac{r_{0,x1}^2}{|r_{x1,x1}| + |r_{x1,x2}|}$$

```
MATLAB: h_51 = (R0(1)^2)/(abs(R(1,1))+abs(R(1,2)))
```

- obliczyć indywidualną pojemność h_{52},

$$h_{52} = \frac{r_{0,x2}^2}{|r_{x2,x1}| + |r_{x2,x2}|}$$

```
MATLAB: h_52 = (R0(2)^2)/(abs(R(2,1))+abs(R(2,2)))
```

- obliczyć integralną pojemność H_5,

$$H_5 = h_{51} + h_{52}$$

```
MATLAB: H_5 = h_51 + h_52
```

4. (1 punkt) Określ ostateczny zbiór zmiennych objaśniających (wyszukanie kombinacji o maksymalnej pojemności integralnej).