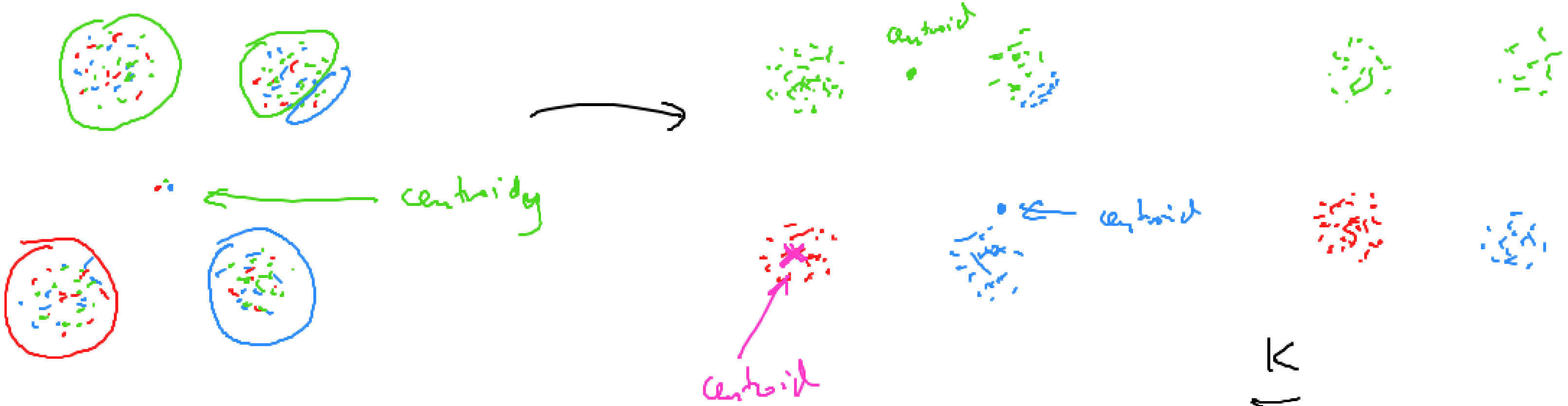
K-means

K=3



$$\sum_{j=1}^{K} W(C_j)$$

Hierarchical clustering (Grupowanie hierarchiczne)

$x_1, x_2, \dots, x_n \in \mathbb{R}^p$

1) Zaczynamy od n grup $\{x_1\}, \{x_2\}, \dots, \{x_n\}$

   · ustalamy pewną miarę (nie)podobieństwa między obserwacjami, np. odległość euklidesową

2) Dla $i = n, n-1, \dots, 2$:

   a) obliczamy podobieństwa pomiędzy każdą parą spośród i grup

   b) sklejamy ze sobą dwie najbardziej podobne grupy, otrzymując $i-1$ grup

Do a) potrzebujemy ustalić, jak liczymy podobieństwo między grupami. Np.

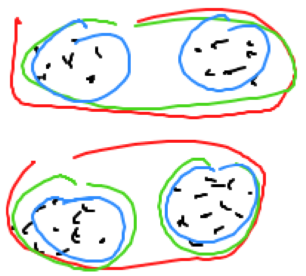
$K_1, K_2 \subset \{x_1, \dots, x_n\}$

· $d(K_1, K_2) = \max\limits_{x \in K_1,\, y \in K_2} d(x,y)$ kompletne

· $d(K_1, K_2) = \min\limits_{x \in K_1,\, y \in K_2} d(x,y)$ pojedyncze (single)

· $d(K_1, K_2) = \frac{1}{|K_1||K_2|} \sum\limits_{x \in K_1,\, y \in K_2} d(x,y)$ średnie

'linkage' (metoda połączenia)



PCA (principal component analysis)

metoda zmniejszenia wymiaru

$$X = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \\ x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \vdots \\ \phi_{p1} \end{bmatrix} \qquad x_1, x_2, \dots, x_n \in \mathbb{R}^p$$

$$x_1 = (x_{11}, \dots, x_{1p})$$

pierwsze współrzędne: znajdujemy $(\phi_{11}, \phi_{21}, \dots, \phi_{p1}) \in \mathbb{R}^p$, aby

$$Z_{1j} = X_{j1}\phi_{11} + X_{j2}\phi_{21} + \dots + X_{jp}\phi_{p1}, \quad j = 1, \dots, n$$

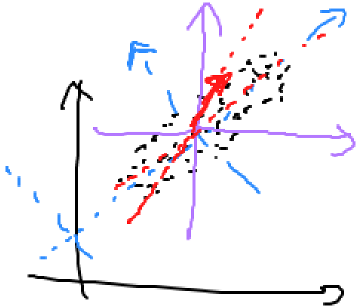
miały możliwie ~~dużą~~ największą wariancję, przy założeniu $\|(\phi_{11}, \dots, \phi_{p1})\|_2 = 1$.

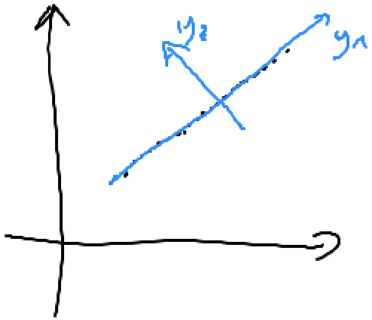
Przy założeniu, że $\sum_{j=1}^{n} X_{jk} = 0$ (średnia zero) dla $k = 1, \dots, p$;

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{p} x_{ji}\phi_{i1} \right)^2 = \sum_{j=1}^{n} \cancel{\langle X_j \cdot \phi_1, X_j \phi_1 \rangle} =$$

$$= \langle X\phi_1, X\phi_1 \rangle \qquad X\phi_1 = \left(\sum_{i=1}^{p} X_{ji}\phi_{i1} \right)_{j=1}^{n}$$

$$\| \quad \sum_{j=1}^{n} \left(\sum_{i=1}^{p} X_{ji}\phi_{i1} \right)^2$$





max: $\langle X\phi_1, X\phi_1\rangle = \langle \underbrace{X^T X}_{\text{(symetryczna) nieujemnie określona}}\phi_1, \phi_1\rangle$

$(AB)^T = B^T A^T$

ϕ_1 – wektor własny $X^T X$ odpowiadający największej wartości własnej

$(X^T X)^T = X^T X^{TT} = X^T X$

$\langle X^T X y, y\rangle = \langle Xy, Xy\rangle \geq 0$

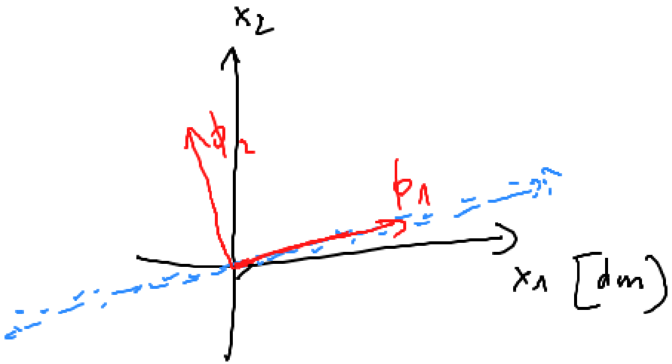
druga współrzędna: szukamy $\phi \in \mathbb{R}^p$, $\|\phi\|_2 = 1$, $\langle \phi, \phi_1\rangle = 0$ taki by
ϕ_2

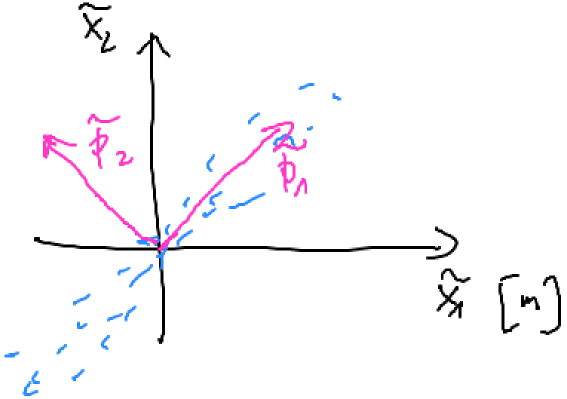
$\langle X^T X \phi, \phi\rangle$ – maksymalne

ϕ_2 – wektor własny $X^T X$ odp. drugiej największej wartości własnej

itd.

Dostajemy nowy układ współrzędnych $(\phi_1, \phi_2, \dots, \phi_p)$





Postaci $\phi_1, \dots, \phi_p$ mogą zależeć od początkowej normalizacji: (początkowa jednostka zera)

$x_1 \approx 10^5$

$x_2 \approx 10^2$

$x_3 \approx 1$