

# 1 Uvod

## 1.1 Uvod

– Ta odstavek je potrebno dopolniti!

Diskriminantna analiza se že dolga leta uporablja za določevanje lastnosti, ki poudarjajo razlike med razredi. Definirana je kot optimizacijski problem, ki vključuje kovariančne matrike, ki predstavljajo razpršenost podatkov znotraj posameznega razreda in razpršenost oziroma ločenost posameznih razredov. Diskriminantna analiza pa sama po sebi zahteva, da je ena od teh kovariančnih matrik nesingularna, kar omejuje njeno uporabo na matrikah določenih dimenzij. V nadaljevanju tako preučimo več različnih optimizacijskih kriterijev in poskušamo njihovo uporabo razširiti na vse matrike z uporabo posplošenega singularnega razcepa. Na ta način se izognemo pogoju nesingularnosti, ki ga zahteva diskriminantna analiza. Tako pridemo do posplošene diskriminantne analize, ki jo lahko uporabimo tudi, kadar je ena matrika singularna (v nadaljevanju lahko vidimo, da je matrika singularna, kadar je velikost vzorca manjša, kot pa je dimenzija posamezne meritve. V nadaljevanju bomo testirali učinkovitost posplošene diskriminantne analize in jo, kjer bo to mogoče, primerjali tudi z običajno diskriminantno analizo.

## 1.2 Matematični uvod

Cilj diskriminantne analize je združiti lastnosti originalnih podatkov na način, ki kar najučinkoviteje ločuje med razredi, v katerih so podatki. Pri takšnem združevanju lastnosti podatkov se dimenzija teh podatkov zmanjša tako, da se struktura teh podatkov in določenih razredov kar najbolje ohrani.

Predpostavimo, da so podatki zloženi v matriko  $A \in \mathbb{R}^{m \times n}$ , kjer  $m$  predstavlja dimenzijo posamezne meritve,  $n$  pa predstavlja število meritev oz. podatkov. Denimo, da so podatki v matriki  $A$  iz  $k$  različnih razredov. Tako so stolpci  $a_1, a_2, \dots, a_n$  matrike  $A$  združeni v  $k$  podmatrik, ki predstavljajo razrede, v katerih so podatki:

$$A = [A_1, A_2, \dots, A_k], \text{ kjer } A_i \in \mathbb{R}^{m \times n_i}.$$

Tu število  $n_i$  predstavlja moč indeksne množice razreda  $i$ . To indeksno množico razreda  $i$  označimo z  $N_i$ . Očitno velja tudi

$$\sum_{i=1}^k n_i = n.$$

Matriko  $A$  lahko poleg razdelitve na podmatrike razdelimo tudi na stolpce. Matrika  $A = [a_{i,j}] \in \mathbb{R}^{m \times n}$  je tako sestavljena iz  $n$  posameznih stolpcev, kjer  $i$ -ti stolpec označimo z  $a_i$ :

$$a_i = \begin{bmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{m,i} \end{bmatrix}.$$

Cilj diskriminantne analize je najti linearno preslikavo iz  $\mathbb{R}^m$  v  $\mathbb{R}^\ell$ , ki v novem prostoru kar najboljše poudari razrede, v katerih so podatki. Tu navadno velja  $\ell \leq m - 1$ , torej da je prostor, v katerega ta linearna preslikava slika, manjdimenzionalen kot prvotni prostor. To iskano linearno preslikavo predstavimo z matriko  $G^T \in \mathbb{R}^{\ell \times m}$ . Za preslikavo  $G^T$  torej velja

$$G^T : \mathbb{R}^m \rightarrow \mathbb{R}^\ell.$$

Torej preslikava  $G^T$  nek  $m$ -dimenzionalen vektor preslika v nov vektor v  $\ell$ -dimezionalnem prostoru, v katerem so razredi podatkov poudarjeni, razpršenost podatkov znotraj razredov je zmanjšana, razlike med razredi pa so povečane.

Za nadaljnje izračune moramo definirati tudi centroid  $i$ -tega razreda, ki je izračunan kot povprečje stolpcev v  $i$ -tem razredu,

$$c^{(i)} = \frac{1}{n_i} \sum_{j \in N_i} a_j,$$

in centroid celotnih podatkov, ki je izračunan kot povprečje vseh stolpcev, to je

$$c = \frac{1}{n} \sum_{j=1}^n a_j.$$

Razpršenost podatkov znotraj posameznih razredov, razpršenost vseh podatkov ter razpršenost oziroma razlike med razredi je smiselno predstaviti s pomočjo matrik. Zato v nadaljevanju definiramo matriko

$$S_W = \sum_{i=1}^k \sum_{j \in N_i} (a_j - c^{(i)})(a_j - c^{(i)})^T,$$

ki predstavlja matriko razpršenosti podatkov znotraj razredov, matriko

$$S_B = \sum_{i=1}^k \sum_{j \in N_i} (c^{(i)} - c)(c^{(i)} - c)^T = \sum_{i=1}^k n_i (c^{(i)} - c)(c^{(i)} - c)^T,$$

ki predstavlja matriko razpršenosti oziroma razlik med razredi in matriko

$$S_M = \sum_{j=1}^n (a_j - c)(a_j - c)^T,$$

ki prstavlja matriko celotne razpršenosti podatkov. Vse tri matrike so velikosti  $m \times m$ . S pomočjo preslikave  $G^T$  pa jih preslikamo v matrike velikosti  $\ell \times \ell$  na sledeč način:

$$S_W^\ell = G^T S_W G, \quad S_B^\ell = G^T S_B G, \quad S_M^\ell = G^T S_M G.$$

Iz danih matrik razpršenosti podatkov bi radi tvorili kriterij kvalitete strukture razredov. Kriterij kvalitete strukture razredov mora imeti visoko vrednost,

kadar so razredi, v katerih so podatki, strnjeni in dobro ločeni med seboj. Opazimo lahko, da  $\text{sled}(S_W)$  predstavlja, kako skupaj so si podatki v posameznem razredu, saj velja

$$\begin{aligned}\text{sled}(S_W) &= \sum_{t=1}^m \left[ \sum_{i=1}^k \sum_{j \in N_i} (a_{t,j} - c_t^{(i)})^2 \right] = \sum_{i=1}^k \sum_{j \in N_i} \left[ \sum_{t=1}^m (a_{t,j} - c_t^{(i)})^2 \right] \\ &= \sum_{i=1}^k \sum_{j \in N_i} \|a_j - c^{(i)}\|_2^2.\end{aligned}$$

Podobno  $\text{sled}(S_B)$  predstavlja ločenost med razredi, saj velja

$$\begin{aligned}\text{sled}(S_B) &= \sum_{t=1}^m \left[ \sum_{i=1}^k \sum_{j \in N_i} (c_t^{(i)} - c_t)^2 \right] = \sum_{i=1}^k \sum_{j \in N_i} \left[ \sum_{t=1}^m (c_t^{(i)} - c_t)^2 \right] \\ &= \sum_{i=1}^k \sum_{j \in N_i} \|c^{(i)} - c\|_2^2 = \sum_{i=1}^k n_i \|c^{(i)} - c\|_2^2.\end{aligned}$$

Optimalna preslikava  $G^T$  tako maksimizira  $\text{sled}(S_B^\ell)$  in minimizira  $\text{sled}(S_W^\ell)$ . Smislen maksimizacijski kriterij se tako zdi

$$\text{sled}(G^T S_B G) / \text{sled}(G^T S_W G),$$

ki pa ga zaradi lažjega računanja aproksimiramo kar s kriterijem

$$\text{sled}((S_W^\ell)^{-1} S_B^\ell).$$

Kljub temu, da je ta optimizacijski kriterij lažje izračunljiv ima svoje pomanjkljivosti. Opazimo lahko, da kriterija ne moremo uporabiti, ko je matrika  $S_W^\ell$  singularna, torej kadar je njena determinanta enaka 0.

– **Zelo narobe, razmisli za rešitev!** –

Ker pa za determinanto matrike velja

$$\det(S_W^\ell) = \det(G^T S_W G) = \det(G^T) \cdot \det(S_W) \cdot \det(G),$$

je  $\det(S_W^\ell)$  enaka 0 kadar je  $\det(S_W)$  enaka 0, torej kadar je matrika  $S_W$  singularna.

–

Do te situacije pa lahko pride kar precej pogosto. Matrika  $S_W \in \mathbb{R}^{m \times m}$  je singularna namreč v vseh primerih, ko za matriko  $A \in \mathbb{R}^{m \times n}$  velja  $m > n$ , saj je potem po definiciji sestavljena kot vsota  $n$  matrik z rangom 1.  $m \times m$  dimenzionalna matrika  $S_W$  ima tako rang manjši ali enak  $n$ , iz česar sledi, da je njena determinanta enaka 0. Na primer, do tega problema pride v primeru, ko je pridobivanje podatkov drago oz. zahtevno in so pridobljeni podatki visokih dimenzij (dimenzija posameznega podatka je večja od števila vseh pridobljenih podatkov).

Obstaja več načinov, kako aplicirati diskriminantno analizo na matriki  $A \in \mathbb{R}^{m \times n}$  z  $m > n$ . V grobem jih ločimo na tiste, kjer dimenzijo podatkov zmanjšamo v dveh korakih, in na tiste, kjer dimenzijo podatkov zmanjšamo v enem koraku. Pri prvem načinu se faza diskriminantne analize nadaljuje v fazo, v kateri zanemarimo oblike posameznih razredov. Najpopularnejša metoda za prvi del tega procesa je zmanjšanje ranga matrike s pomočjo singularnega razcepa. To je tudi glavno orodje metode imenovane metoda glavnih komponent. Kakorkoli, celotna predstava dvostopenjskih načinov je precej občutljiva na zmanjšanje dimenzije v prvi fazi. V diplomskem delu se bomo osredotočili na način, ki posploši diskriminantno analizo tako, da teoretično optimalno zmanjša dimenzijo podatkov brez da bi uvedel dodaten korak. V ta namen bomo obravnavali kriterij

$$\text{sled}((S_2^\ell)^{-1} S_1^\ell),$$

kjer matriki  $S_2$  in  $S_1$  predstavljata poljubno matriko izmed  $S_W$ ,  $S_B$  in  $S_M$ . Kadar je matrika  $S_2$  nesingularna, klasična diskriminantna analiza predstavi svojo rešitev s pomočjo posplošenega problema lastnih vrednosti. S prestrukturiranjem problema tako, da uporabimo posplošeni singularni razcep, pa lahko razširimo uporabnost diskriminantne analize tudi na primer, ko je matrika  $S_2$  singularna.

## 2 Matematična priprava - posplošeni singularni razcep

Originalna definicija posplošenega singularnega razcepa (Van Loan) je sledeča.

**Izrek 1** (Posplošeni singularni izrek (Van Loan)). *Za matriki  $K_A \in \mathbb{R}^{p \times m}$  z  $p \geq m$  in  $K_B \in \mathbb{R}^{n \times m}$  obstajata ortogonalni matriki  $U \in \mathbb{R}^{p \times p}$  in  $V \in \mathbb{R}^{n \times n}$  ter nesingularna matrika  $X \in \mathbb{R}^{m \times m}$ , da velja*

$$U^T K_A X = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_m \\ \hline & & 0_{p-m,m} \end{bmatrix} \quad \text{in} \quad V^T K_B X = \Sigma_{B_q},$$

kjer je  $q = \min(n, m)$ ,

$$\Sigma_{B_q} = \begin{bmatrix} \beta_1 & & & & \\ & \ddots & & & \\ & & \beta_q & & \\ \hline & & 0_{n-q,q} & & \\ & & & 0_{n-q,m-q} & \end{bmatrix},$$

$\alpha_i \geq 0$  za  $1 \leq i \leq m$ ,  $\beta_i \geq 0$  za  $1 \leq i \leq q$  in  $\beta_1 \geq \beta_2 \geq \dots \geq \beta_q$ .

*Dokaz.* Iz matrik  $K_A$  in  $K_B$  tvorimo združeno  $(p+n) \times m$  matriko  $K = \begin{bmatrix} K_A \\ K_B \end{bmatrix}$ , na kateri naredimo singularni razcep. Iz singularnega razcepa dobimo matriki

$Q \in \mathbb{R}^{(p+n) \times (p+n)}$  in matriko  $Z_1 \in \mathbb{R}^{m \times m}$ , tako da velja

$$Q^T \begin{bmatrix} K_A \\ K_B \end{bmatrix} Z_1 = \begin{bmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_m \\ \hline & & & 0_{p+n-m, m} \end{bmatrix}, \quad (1)$$

kjer za  $k = \text{rang}(K)$  velja  $\gamma_1 \geq \dots \geq \gamma_k > \gamma_{k+1} = \dots = \gamma_m = 0$ .

Matriko  $Z_1$  razdelimo na dve matriki, matriko  $Z_{11} \in \mathbb{R}^{m \times k}$ , ki je sestavljena iz prvih  $k$  stolpcev matrike  $Z_1$  in matriko  $Z_{12} \in \mathbb{R}^{m \times (m-k)}$ , ki je sestavljena iz preostalih  $m - k$  stolpcev matrike  $Z_1$ . Pišemo

$$Z_1 = [Z_{11} \ Z_{12}].$$

Po predpostavki velja  $p \geq m$  in ker je očitno tudi  $m \geq k$  sledi  $p \geq m \geq k$ . Sedaj definirajmo matriko

$$D := \text{diag}(\gamma_1, \dots, \gamma_k) \in \mathbb{R}^{k \times k}.$$

Tako iz zgornje enačbe (1) dobimo

$$\begin{bmatrix} K_A Z_{11} & K_A Z_{12} \\ K_B Z_{11} & K_B Z_{12} \end{bmatrix} = Q \begin{bmatrix} D & 0_{k, m-k} \\ 0_{p+n-k, k} & 0_{p+n-k, m-k} \end{bmatrix}, \quad (2)$$

od koder sledi

$$\begin{bmatrix} K_A Z_{11} \\ K_B Z_{11} \end{bmatrix} = Q \begin{bmatrix} D \\ 0 \end{bmatrix}.$$

V kolikor še matriko  $Q$  razdelimo na podmatrike na naslednji način

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix},$$

kjer je matrika  $Q_{11} \in \mathbb{R}^{p \times k}$ , matrika  $Q_{12} \in \mathbb{R}^{p \times (p+n-k)}$ , matrika  $Q_{21} \in \mathbb{R}^{n \times k}$  in matrika matrika  $Q_{22} \in \mathbb{R}^{n \times (p+n-k)}$ , ugotovimo, da je

$$Q \begin{bmatrix} D \\ 0 \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} D \\ 0 \end{bmatrix} = \begin{bmatrix} Q_{11} D \\ Q_{21} D \end{bmatrix}.$$

Iz tega neposredno sledi enakost

$$K_A Z_{11} = Q_{11} D \implies K_A Z_{11} D^{-1} = Q_{11} =: K_{A_1} \in \mathbb{R}^{p \times k}$$

in enakost

$$K_B Z_{11} = Q_{21} D \implies K_B Z_{11} D^{-1} = Q_{21} =: K_{B_1} \in \mathbb{R}^{n \times k}.$$

Sedaj singularni razcep naredimo na matriki  $K_{B_1}$ . Za matriko  $K_{B_1}$  vemo, da ima isti rang kot matrika  $K_B$ , saj velja, da je matrika  $Z_{11}$  polnega ranga (je namreč podmatrika ortogonalne matrike  $Z$ ) in vemo, da je matrika  $D^{-1}$  polnega ranga.

Označimo  $r = \text{rang}(K_B) = \text{rang}(K_{B_1})$ . Iz singularnega razcepa za matriko  $K_{B_1}$  dobimo ortogonalni matriki  $V \in \mathbb{R}^{n \times n}$  in  $Z_2 \in \mathbb{R}^{k \times k}$ , da velja

$$V^T K_{B_1} Z_2 = \Sigma_{B_t}, \quad (3)$$

kjer je  $t = \min\{n, k\}$ ,  $\Sigma_{B_t} = \begin{bmatrix} \beta_1 & & & & & \\ & \ddots & & & & \\ & & \beta_t & & & \\ \hline & & & 0_{t, k-t} & & \\ & 0_{n-t, t} & & & 0_{n-t, k-t} & \end{bmatrix}$  in velja  $\beta_1 \geq \beta_2 \geq$

$$\dots \geq \beta_r > \beta_{r+1} = \dots = \beta_t = 0.$$

Iz enačbe (2) sledi, da je

$$K_B Z_{12} = 0_{n, m-k}.$$

Opazimo, da velja tudi

$$\begin{aligned} V^T K_B Z_1 \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix} &= V^T K_B [Z_{11} \ Z_{12}] \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix} = \\ V^T K_B [Z_{11} D^{-1} Z_2 \ Z_{12}] &= [V^T K_B Z_{11} D^{-1} Z_2 \ V^T K_B Z_{12}] = \\ [V^T K_{B_1} Z_2 \ 0_{n, m-k}] &= [\Sigma_{B_t} \ 0_{n, m-k}] = \\ \begin{bmatrix} \beta_1 & & & & & \\ & \ddots & & & & \\ & & \beta_t & & & \\ \hline & & & 0_{t, m-t} & & \\ & 0_{n-t, t} & & & 0_{n-t, m-t} & \end{bmatrix}. \end{aligned}$$

Če za  $q = \min\{n, m\}$  dodatno definiramo še  $\beta_{t+1} = \dots = \beta_q = 0$ , dobimo ravno

$$V^T K_B Z_1 \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix} = \begin{bmatrix} \beta_1 & & & & & \\ & \ddots & & & & \\ & & \beta_q & & & \\ \hline & & & 0_{q, m-q} & & \\ & 0_{n-q, q} & & & 0_{n-q, m-q} & \end{bmatrix},$$

kar je pa ravno matrika  $\Sigma_{B_q}$  iz izreka. Matriko  $X$  tako definiramo na sledeči način

$$X := Z_1 \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix}.$$

Dokazati pa moramo tudi, da zgornja matrika  $X$  ustreza tudi enačbi iz izreka za matriko  $K_A$ , torej, da obstaja tudi taka matrika  $U$ , da velja

$$U^T K_A X = \begin{bmatrix} \alpha_1 & & & & & \\ & \ddots & & & & \\ & & \alpha_m & & & \\ \hline & & & 0_{p-m, m} & & \end{bmatrix}.$$

Ker je matrika  $Q$  ortogonalna, dodatno velja  $K_{A_1}^T K_{A_1} + K_{B_1}^T K_{B_1} = I_k$ , kjer je  $I_k$  identična matrika dimenzije  $k \times k$ . To enakost lahko pokažemo tako, da razpišemo spodnjo enačbo

$$\begin{aligned} Q^T Q &= \begin{bmatrix} Q_{11}^T & Q_{21}^T \\ Q_{12}^T & Q_{22}^T \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} Q_{11}^T Q_{11} + Q_{21}^T Q_{21} & Q_{11}^T Q_{12} + Q_{21}^T Q_{22} \\ Q_{12}^T Q_{11} + Q_{22}^T Q_{21} & Q_{12}^T Q_{12} + Q_{22}^T Q_{22} \end{bmatrix} \\ &= \begin{bmatrix} K_{A_1}^T K_{A_1} + K_{B_1}^T K_{B_1} & Q_{11}^T Q_{12} + Q_{21}^T Q_{22} \\ Q_{12}^T Q_{11} + Q_{22}^T Q_{21} & Q_{12}^T Q_{12} + Q_{22}^T Q_{22} \end{bmatrix} = I = \begin{bmatrix} I_k & 0 \\ 0 & I_{p+n-k} \end{bmatrix}. \end{aligned}$$

Definirajmo matriko  $G$ , ki jo dobimo s preoblikovanjem enačbe (3):

$$G := K_{B_1} Z_2 = V \Sigma_{B_t} \in \mathbb{R}^{n \times k}.$$

Za matriko  $K_{A_1} Z_2$  izračunamo razširjen QR razcep,  $K_{A_1} Z_2 = UR$ , kjer je  $U \in \mathbb{R}^{p \times p}$  ortogonalna in  $R \in \mathbb{R}^{p \times k}$  zgornja trapezna matrika. Tak razcep lahko naredimo na primer z uporabo Householderjevih zrcaljenj.

Opazimo lahko, da so stolpci matrike  $K_{A_1} Z_2$  medsebojno ortogonalni, saj velja

$$\begin{aligned} (K_{A_1} Z_2)^T (K_{A_1} Z_2) &= Z_2^T K_{A_1}^T K_{A_1} Z_2 = Z_2^T (I_k - K_{B_1}^T K_{B_1}) Z_2 = \\ &= Z_2^T Z_2 - Z_2^T K_{B_1}^T K_{B_1} Z_2 = I_k - G^T G = I_k - \Sigma_{B_t}^T V^T V \Sigma_{B_t} \\ &= I_k - \begin{bmatrix} \beta_1^2 & & & & \\ & \ddots & & & \\ & & \beta_t^2 & & \\ & & & & 0_{t,k-t} \\ 0_{k-t,k-t} & & & & 0_{t,k-t} \end{bmatrix} = \begin{bmatrix} 1 - \beta_1^2 & & & & \\ & \ddots & & & \\ & & 1 - \beta_t^2 & & \\ & & & & 0_{t,k-t} \\ 0_{k-t,k-t} & & & & I_{t,k-t} \end{bmatrix} \\ &= \text{diag}(1 - \beta_1^2, \dots, 1 - \beta_k^2), \end{aligned}$$

kjer smo dodatno definirali še  $\beta_{t+1} = \dots = \beta_k = 0$ . Iz tega sledi, da je matrika  $R$  oblike

$$R = \begin{bmatrix} \sqrt{1 - \beta_1^2} & & & \\ & \ddots & & \\ & & \sqrt{1 - \beta_k^2} & \\ 0_{p-k,k} & & & \end{bmatrix},$$

saj velja

$$(K_{A_1} Z_2)^T (K_{A_1} Z_2) = R^T U^T U R = R^T R = \text{diag}(1 - \beta_1^2, \dots, 1 - \beta_k^2).$$

Velja tudi

$$\begin{aligned}
U^T K_A Z_1 \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix} &= U^T K_A [Z_{11} \ Z_{12}] \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix} = \\
U^T K_A [Z_{11} D^{-1} Z_2 \ Z_{12}] &= [U^T K_A Z_{11} D^{-1} Z_2 \ U^T K_A Z_{12}] = \\
[U^T K_{A_1} Z_2 \ 0_{p,m-k}] &= [U^T U R \ 0_{p,m-k}] = [R \ 0_{p,m-k}] = \\
&\begin{bmatrix} \alpha_i & & & & & \\ & \ddots & & & & \\ & & \alpha_k & & & \\ & & & & 0_{k,m-k} & \\ \cdots & & & & & \\ & 0_{p-k,k} & & & & 0_{p-k,m-k} \end{bmatrix},
\end{aligned}$$

kjer smo dodatno definirali  $\alpha_i = \sqrt{1 - \beta_i^2}$  za  $i = 1, \dots, k$  in  $\alpha_{k+1} = \dots = \alpha_m$ . S tem smo pokazali, da matrika

$$X = Z_1 \begin{bmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{bmatrix}$$

zadošča tako razcepu matrike  $K_A$  kot tudi razcepu matrike  $K_B$  iz izreka in s tem dokazali izrek.  $\square$

Problem tega izreka pa je, da se ga ne da uporabiti, kadar dimenzije matrike  $K_A$  niso ustrezne. Zaradi tega pretirano zavezujočega pogoja se odločita C.C. Paige in M.A. Saunders ta posplošeni singularni izrek še dodatno posplošiti. Tako dobimo naslednji izrek:

**Izrek 2** (Posplošeni singularni razcep (Paige in Saunders)). *Naj bosta dani matriki  $K_A \in \mathbb{R}^{p \times m}$  in  $K_B \in \mathbb{R}^{n \times m}$ . Potem za  $K = \begin{bmatrix} K_A \\ K_B \end{bmatrix}$  in  $k = \text{rang}(K)$  obstajajo ortogonalne matrike  $U \in \mathbb{R}^{p \times p}$ ,  $V \in \mathbb{R}^{n \times n}$ ,  $W \in \mathbb{R}^{k \times k}$  in  $Q \in \mathbb{R}^{m \times m}$ , da velja*

$$U^T K_A Q = \Sigma_A \begin{bmatrix} W^T R, & 0 \end{bmatrix} \quad \text{in} \quad V^T K_B Q = \Sigma_B \begin{bmatrix} W^T R, & 0 \end{bmatrix}, \quad (4)$$

kjer sta

$$\Sigma_A = \begin{bmatrix} I_A & & \\ & D_A & \\ & & 0_A \end{bmatrix} \quad \text{in} \quad \Sigma_B = \begin{bmatrix} 0_B & & \\ & D_B & \\ & & I_B \end{bmatrix},$$

$R \in \mathbb{R}^{k \times k}$  je nesingularna matrika, matriki  $I_A \in \mathbb{R}^{r \times r}$  in  $I_B \in \mathbb{R}^{(k-r-s) \times (k-r-s)}$  identični matriki, kjer je

$$r = \text{rang}(K) - \text{rang}(K_B) \quad \text{in} \quad s = \text{rang}(K_A) + \text{rang}(K_B) - \text{rang}(K),$$

$0_A \in \mathbb{R}^{(p-r-s) \times (k-r-s)}$  in  $0_B \in \mathbb{R}^{(n-k+r) \times r}$  ničelni matriki, ki imata lahko tudi ničelno število vrstic ali stolpcev, matriki  $D_A = \text{diag}(\alpha_{r+1}, \dots, \alpha_{r+s})$  in  $D_B = \text{diag}(\beta_{r+1}, \dots, \beta_{r+s})$  pa diagonalni matriki, ki zadoščata pogoju

$$1 > \alpha_{r+1} \geq \dots \geq \alpha_{r+s} > 0 \quad \text{in} \quad 0 < \beta_{r+1} \leq \dots \leq \beta_{r+s} < 1$$

pri  $\alpha_i^2 + \beta_i^2 = 1$  za  $i = r+1, \dots, r+s$ .



*Dokaz.* Izračunamo singularni razcep matrike  $K$ . Tako dobimo ortogonalni matriki  $P \in \mathbb{R}^{(p+n) \times (p+n)}$  in  $Q \in \mathbb{R}^{m \times m}$ , da velja

$$PKQ^T = \begin{bmatrix} R & 0_{k, m-k} \\ 0_{n+p-k, k} & 0_{n+p-k, m-k} \end{bmatrix}, \quad (5)$$

kjer je  $R$  diagonalna matrika ranga  $k$ . Matriki  $Q$  in  $P$  razdelimo na podmatrike

$$Q = [Q_1 \ Q_2] \quad \text{in} \quad P = [P_1 \ P_2] = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix},$$

kjer je matrika  $Q_1 \in \mathbb{R}^{m \times k}$  sestavljena iz prvih  $k$  stolpcev matrike  $Q$ , matrika  $Q_2 \in \mathbb{R}^{m \times (m-k)}$  pa iz preostalih  $m-k$  stolpcev matrike  $Q$ , podmatrike matrike  $P$  pa so sledečih dimenzij:  $P_{11} \in \mathbb{R}^{p \times k}$ ,  $P_{12} \in \mathbb{R}^{p \times (p+n-k)}$ ,  $P_{21} \in \mathbb{R}^{n \times k}$  in  $P_{22} \in \mathbb{R}^{n \times (p+n-k)}$ .

Ker je  $P$  ortogonalna matrika velja  $\|P\|_2 \leq 1$  in posledično tudi  $\|P_{11}\|_2 \leq \|P\|_2 \leq 1$ . Posledično nobena singularna vrednost matrike  $P_{11}$  ni večja od 1.

Singularni razcep podobno kot na matriki  $K$  naredimo tudi na matriki  $P_{11}$ . Tako dobimo ortogonalni matriki  $U \in \mathbb{R}^{m \times m}$  in  $W \in \mathbb{R}^{k \times k}$ , da velja

$$U^T P_{11} W = \Sigma_A,$$

kjer je

$$\Sigma_A = \begin{bmatrix} I_r & & \\ & D_A & \\ & & 0_A \end{bmatrix},$$

kjer je  $r$  geometrična večkratnost lastne vrednosti 1, matrika  $I_r$  identična ma-

trika dimenzije  $r \times r$ , matrika  $D_A = \begin{bmatrix} \alpha_{r+1} & & \\ & \ddots & \\ & & \alpha_{r+s} \end{bmatrix}$  diagonalna matrika,

kjer  $r+s$  predstavlja rang matrike  $P_{11}$  in  $0_A \in \mathbb{R}^{(p-r-s) \times (k-r-s)}$  je ničelna matrika, ki ima lahko nič vrstic ali nič stolpcev.

Na matriki  $P_{21}$  uporabimo (**- DOPOLNI!**) razcep (s Householderjevimi zrcaljenji) in tako dobimo matriko  $V \in \mathbb{R}^{n \times n}$ , da velja

$$V^H P_{21} W = L = (\ell_{ij})_{i,j} = \begin{bmatrix} 0 & 0 \\ 0 & L_1 \end{bmatrix},$$

(**- POVEJ KATERIH DIMENZIJ SO TE NIČELNE MATRIKE!**)

kjer je matrika  $L_1$  spodnjetrokotna z diagonalnimi elementi večjimi od 0. Opa-

zimo lahko, da velja spodnja enakost

$$\begin{bmatrix} U^T & 0_{m,n} \\ 0_{n,m} & V^T \end{bmatrix} \begin{bmatrix} P_{11} \\ P_{21} \end{bmatrix} W = \begin{bmatrix} U^T P_{11} W \\ V^T P_{21} W \end{bmatrix} = \begin{bmatrix} \Sigma_A \\ L \end{bmatrix}.$$

Zgornja matrika  $\begin{bmatrix} \Sigma_A \\ L \end{bmatrix}$  je ortogonalna, saj je produkt ortogonalnih matrik. To implicira obliko matrike  $\begin{bmatrix} \Sigma_A \\ L \end{bmatrix}$ , torej

$$\begin{bmatrix} \Sigma_A \\ L \end{bmatrix} = \begin{bmatrix} I_r & & \\ & D_A & \\ & & 0_{m-r-s, k-r-s} \\ & 0_{p-k+r, r} & \\ & & D_B \\ & & & I_{k-r-s, k-r-s} \end{bmatrix},$$

kjer je matrika  $D_B$  diagonalna matrika  $D_B = \begin{bmatrix} \beta_{r+1} & & \\ & \ddots & \\ & & \beta_{r+s} \end{bmatrix}$ . Za matriko  $L$  pa velja  $L = \Sigma_B = \begin{bmatrix} 0_{p-k+r, r} & & \\ & D_B & \\ & & I_{k-r-s, k-r-s} \end{bmatrix}$ . Iz ortogonalnosti matrike  $\begin{bmatrix} \Sigma_A \\ L \end{bmatrix}$  sledi tudi, da so njeni stolpci ortonormirani, iz česar sledi  $\alpha_{r+i}^2 + \beta_{r+i}^2 = 1$  za  $i = 1, \dots, s$ .

Iz enačbe (5) dobimo

$$\begin{bmatrix} K_A \\ K_B \end{bmatrix} Q = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} R & 0_{k, m-k} \\ 0_{n+p-k, k} & 0_{n+p-k, m-k} \end{bmatrix} = \begin{bmatrix} P_{11}R & 0_{m, m-k} \\ P_{21}R & 0_{p, m-k} \end{bmatrix} = \begin{bmatrix} U\Sigma_A W^T R & 0_{m, m-k} \\ V\Sigma_B W^T R & 0_{p, m-k} \end{bmatrix}, \quad (6)$$

iz česar sledi

$$P_{11} = U\Sigma_A W^T \text{ in } P_{21} = V\Sigma_B W^T.$$

Če začetno enačbo iz enačbe (6) pomnožimo iz leve z matriko  $\begin{bmatrix} U^T & 0_{m, n} \\ 0_{n, m} & V^T \end{bmatrix}$  dobimo

$$\begin{bmatrix} U^T & 0_{m, n} \\ 0_{n, m} & V^T \end{bmatrix} \begin{bmatrix} K_A \\ K_B \end{bmatrix} Q = \begin{bmatrix} U^T & 0_{m, n} \\ 0_{n, m} & V^T \end{bmatrix} \begin{bmatrix} U\Sigma_A W^T R & 0_{m, m-k} \\ V\Sigma_B W^T R & 0_{p, m-k} \end{bmatrix},$$

iz česar sledi

$$U^T K_A Q = \Sigma_A \begin{bmatrix} W^T R, & 0 \end{bmatrix} \text{ in } V^T K_B Q = \Sigma_B \begin{bmatrix} W^T R, & 0 \end{bmatrix},$$

s čimer smo dokazali izrek.  $\square$

Iz posplošenega singularnega razcepa, ki sta ga definirala Paige in Saunders neposredno sledi Van Loanova posplošitev singularnega razcepa. S preoblikovanjem enačbe (3), dobimo

$$U^T K_A Q = \begin{bmatrix} \Sigma_A, & 0 \end{bmatrix} \begin{bmatrix} W^T R & 0 \\ 0 & I \end{bmatrix}.$$

Inverz matrike  $\begin{bmatrix} W^T R & 0 \\ 0 & I \end{bmatrix}$  je kar matrika  $\begin{bmatrix} R^{-1} W & 0 \\ 0 & I \end{bmatrix}$ , saj veljata obe enakosti iz definicije inverza, torej

$$\begin{bmatrix} R^{-1} W & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} W^T R & 0 \\ 0 & I \end{bmatrix} = I \quad \text{in} \quad \begin{bmatrix} W^T R & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} R^{-1} W & 0 \\ 0 & I \end{bmatrix} = I.$$

Iz enačbe (3) tako tu očitno sledi

$$U^T K_A Q \begin{bmatrix} R^{-1} W & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} \Sigma_A & 0 \end{bmatrix}.$$

Matriko  $X$  definiramo kot

$$Q \begin{bmatrix} R^{-1} W & 0 \\ 0 & I \end{bmatrix}$$

in dobimo ravno Van Loanovo posplošitev razcepa

$$U^T K_A X = \begin{bmatrix} \Sigma_A & 0 \end{bmatrix}.$$

Podobno lahko pokažemo tudi za matriko  $K_B$ , za katero iz (4) dobimo

$$U^T K_B X = \begin{bmatrix} \Sigma_B & 0 \end{bmatrix}.$$

Tako je definirana matrika  $X$  ravno iskana matrika iz prve posplošitve singularnega razcepa in matriki  $\begin{bmatrix} \Sigma_A & 0 \end{bmatrix}$  in  $\begin{bmatrix} \Sigma_B & 0 \end{bmatrix}$  ravno iskani diagonalni matriki.

Za nadaljnje delo definirajmo matrike

$$H_W := \begin{bmatrix} A_1 - c^{(1)} e^{(1)T}, \dots, A_k - c^{(k)} e^{(k)T} \end{bmatrix},$$

$$H_B := \begin{bmatrix} (c^{(1)} - c) e^{(1)T}, \dots, (c^{(k)} - c) e^{(k)T} \end{bmatrix}$$

in

$$H_M := [a_1 - c, \dots, a_n - c] = A - c e^T = H_W + H_B,$$

kjer velja  $e^{(i)} = (1, \dots, 1)^T \in \mathbb{R}^{n_i \times 1}$  in  $e = (1, \dots, 1)^T \in \mathbb{R}^{n \times 1}$ .

S pomočjo teh matrik lahko definiramo tudi matrike razpršenosti podatkov. Matriko  $S_W$  lahko definiramo kot produkt matrike  $H_W$  z njeno transponiranko, torej

$$S_W = H_W H_W^T,$$

matriko  $S_B$  lahko definiramo na podoben način kot

$$S_B = H_B H_B^T,$$

prav tako pa tudi matriko  $S_M$

$$S_M = H_M H_M^T.$$

S pomočjo tega razcepa lahko tudi na drugačen način pokažemo, da je matrika  $S_W$ , kadar velja  $m > n$ , singularna. Razvidno je namreč, da je ta matrika  $S_W$  definirana kot produkt dveh matrik, kjer je matrika  $H_W$  dimenzije  $m \times n$ , matrika  $H_W^T$  pa dimenzije  $n \times m$ . V kolikor velja  $m > n$ , sta tako ti dve matrike največ ranga  $n$ . Ker pa za rang matrike velja, da je rang produkta dveh matrik navzgor omejen z manjšim izmed rangov teh dveh posameznih matrik ( $\text{rang}(AB) \leq \min(\text{rang}(A), \text{rang}(B))$ ), je posledično tudi matrika  $S_W$  največ ranga  $n$  in torej očitno singularna.

– Tu mogoče dopiši glede matrik  $K_A$  in  $K_B$

### 3 Matematična rešitev problema

V tem odstavku prikažemo uporabo posplošenega singularnega razcepa v namen razširjene uporabe posplošene diskriminantne analize.

#### 3.1 Optimizacija optimizacijskega kriterija $J_1 = \text{sled}(S_2^{-1}S_1)$ za nesingularno matriko $S_2$

Tu izhajamo iz optimizacije optimizacijskega kriterija

$$J_1(G) = \text{sled}((G^T S_2 G)^{-1} (G^T S_1 G))$$

z izbiro optimalne preslikave  $G$ , kjer sta matriki  $S_1$  in  $S_2$  izbrani izmed matrik  $S_W$ ,  $S_B$  in  $S_M$ . Ker je tu matrika  $S_2$  nesingularna in je sestavljena kot produkt matrike in transponiranke te matrike je simetrično pozitivno definitna (posledično so vse lastne vrednosti te matrike večje ali enake 0). Za simetrično pozitivno definitno matriko pa obstaja razcep Choleskega; tako vemo da obstaja spodnjetrokotna matrika  $V$  s pozitivnimi elementi na diagonalni, da velja:

$$S_2 = VV^T.$$

#### 3.2 Posplošitev maksimizacijskega kriterija $\text{sled}((S_W^Y)^{-1}S_B^Y)$ za singularno matriko $S_2$

### 4 Algoritem

### 5 Zaključek

### 6 Priloge

**Izrek 3** (Singularni razcep). *Za vsako matriko  $A \in \mathbb{R}^{m \times n}$ , z lastnostjo  $m \geq n$ , obstaja singularni razcep*

$$A = U\Sigma V^T,$$

kjer sta  $U \in \mathbb{R}^{m \times m}$  in  $V \in \mathbb{R}^{n \times n}$  ortogonalni matriki,  $\Sigma \in \mathbb{R}^{m \times n}$  je oblike

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$$

in  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$  so singularne vrednosti matrike  $A$ .

*Dokaz.* Ker je  $A^T A$  simetrična pozitivno semidefinitna matrika, so vse njene lastne vrednosti nenegativne. Označimo in uredimo jih kot

$$\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_n^2 \geq 0.$$

Ustrezni ortonormirani lastni vektorji  $v_1, \dots, v_n$  zadoščajo  $A^T A v_i = \sigma_i^2 v_i$  za  $i = 1, \dots, n$ . Naj bo  $\sigma_r > 0$  in  $\sigma_{r+1} = \dots = \sigma_n = 0$ . Matriko  $V$  razdelimo na  $V_1 = [v_1, \dots, v_r]$  in  $V_2 = [v_{r+1}, \dots, v_n]$ . Iz

$$(AV_2)^T (AV_2) = V_2^T A^T A V_2 = V_2^T [0, \dots, 0] = 0$$

sledi  $AV_2 = 0$ . Sedaj definiramo  $u_i := \frac{1}{\sigma_i} A v_i$  za  $i = 1, \dots, r$ . Vektorji  $u_1, \dots, u_r$  so ortonormirani, saj je

$$u_i^T u_j = \frac{1}{\sigma_i \sigma_j} v_i^T A^T A v_j = \frac{\sigma_j}{\sigma_i} v_i^T v_j = \delta_{ij}, \quad i, j = 1, \dots, r,$$

kjer smo v zapisu uprabilili t.i. *Kroneckerjev delta*, definiran z  $\delta_{ij} = 1$  za  $i = j$  in  $\delta_{ij} = 0$  za  $i \neq j$ . Označimo  $U_1 = [u_1 \ \dots \ u_r]$  in dopolnimo z  $U_2 = [u_{r+1} \ \dots \ u_n]$ , da je  $U = [U_1 \ U_2]$  ortogonalna matrika. Matrika  $U^T A V$  ima obliko

$$U^T A V = \begin{bmatrix} U_1^T A V_1 & U_1^T A V_2 \\ U_2^T A V_1 & U_2^T A V_2 \end{bmatrix}.$$

Desna bloka sta zaradi  $AV_2 = 0$  enaka 0. Za  $i = 1, \dots, r$  in  $k = 1, \dots, m$  velja

$$u_k^T A v_i = \sigma_i u_k^T u_i = \sigma_i \delta_{ik},$$

torej  $U_2^T A V_1 = 0$  in  $U_1^T A V_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$ . Dobimo singularni razcep  $A = U \Sigma V^T$ , kjer je  $S = \text{diag}(\sigma_1, \dots, \sigma_r)$  in

$$\Sigma = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}.$$

□

V primeru, ko velja  $n > m$ , dobimo singularni razcep za  $A \in \mathbb{R}^{m \times n}$  tako, da transponiramo singularni razcep  $A^T$ .

## 7 Viri