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. Na ta način dobimo posplošeno diskriminantno analizo, ki jo lahko uporabimo tudi, kadar je ena matrika nesingularna (v nadaljevanju lahko vidimo, da je matrika nesingularna kadar je velikost vzorca manjša kot pa dimenzija posamezne meritve. V nadaljevanju bom testiral učinkovitost posplošene diskriminantne analize in jo, kjer bo to mogoče, primerjal tudi z običajno diskriminantno analizo.

1.2 Matematični uvod

Cilj diskriminantne analize je združevati 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 na način, ki najbolje ohranja strukture določenih razredov.

Tu 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:

$$A = \begin{pmatrix} A_1, & A_2, & \dots, & A_k \end{pmatrix}, \text{ kjer } A_i \in \mathbb{R}^{m\times n_i}.$$

Tu število n_i predstavlja moč indeksne množice razreda i. To indeksno množico razreda i označujemo 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 najbolje 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:

$$G^T : \mathbb{R}^m \to \mathbb{R}^\ell,$$

. Torej preslikava G^T nek m-dimenzionalen vektor preslika v nov vektor v ℓ-dimezionalnem prostoru (navadno velja $\ell \leq m$), 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 pradstavlja matriko celotne razpršenosti podatkov.

Med zgoraj definiranimi matrikami velja tudi enakost: $S_M = S_W + S_B$. – **DOKAŽI (pozneje, ko vpelješ še matrike H)**

S pomočjo preslikave G^T preslikamo v ℓ-dimezionalen prostor tudi matrike S_W, S_B in S_M:

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

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 $sled(S_W)$ predstavlja, kako skupaj so si podatki v posameznem razredu, saj velja:

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

Podobno $sled(S_B)$ predstavlja ločenost med razredi, saj velja

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

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

$$sled(G^T S_B G)/sled(G^T S_W G),$$

ki pa ga zaradi lažjega računanja aproksimiramo kar s kriterijem

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

– **NEJASNOST!**

Kljub temu, da je ta optimizacijski kriterij lažje izračunljiv ima svoje pomanjkljivosti. Opazimo lahko, da kriterij lahko uporabimo le v primeru, ko je matrika S_W^ℓ nesingularna oziroma da kriterija ne moremo uporabiti, ko je matrika S_W^ℓ singularna (torej kadar je njena determinanta enaka 0). 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 je za matriko $A \in \mathbb{R}^{m \times n}$ velja $m > n$, saj je potem m-dimenzionalna matrika S_W, ki je sestavljena iz n vektorjev $a_j - c^{(i)},\ i = 1, \ldots, k,\ j \in N_j$. Iz n vektorjev pa lahko sestavimo le n dimenzionalen prostor, torej bo $m \times m$ matrika iz teh vektorjev očitno singularna. Njena determinanta bo tako 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. V prvem načinu se faza diskriminante 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 obravnavamo kriterij

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

kjer sta 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*

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

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

$$\Sigma_B = \left[\begin{array}{ccc|c} \beta_1 & & & \\ & \ddots & & 0_{q,m-q} \\ & & \beta_q & \\ \hline & 0_{n-q,q} & & 0_{n-q,m-q} \end{array}\right],$$

$\alpha_i \geq 0$ za $1 \leq i \leq m$ in $\beta_i \geq 0$ za $1 \leq i \leq q$.

Dokaz. Iz matrik K_A in K_B tvorimo združeno $(p+n) \times m$ matriko $K = \begin{pmatrix} K_A \\ K_B \end{pmatrix}$, 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

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

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

– **DOPOLNI! Tu je potrebno pisati celo razširjeno matriko kot produkt singularnega razcepa!!** V kolikor matriko Z_1 razdelimo na dve matriki: $Z_{11} \in \mathbb{R}^{m\times k}$, ki je sestavljena iz prvih k stolpcev matrike Z_1 in $Z_{12} \in \mathbb{R}^{m\times(m-k)}$, ki je sestavljena iz preostalih $m-k$ stolpcev matrike Z_1, lahko vidimo da velja:

$$Q^T K \begin{pmatrix} Z_{11} | Z_{12} \end{pmatrix} = \begin{bmatrix} \gamma_1 & & & & & \\ & \ddots & & & & \\ & & \gamma_k & & & \\ & & & \gamma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \gamma_m \\ \hline & & & 0_{p+n-m,m} & & \end{bmatrix}.$$

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

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

Tako iz zgornje enačbe (1) dobimo

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

iz česar sledi

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

in v kolikor še matriko Q razdelimo na podmatrike na naslednji način

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

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

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

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

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

.

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 dimzije $k \times k$. To enakost lahko pokažemo tako, da razpišemo spodnjo enačbo

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

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: $r = rang(K_B) = rang(K_{B_1})$. 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:

$$V^T K_{B_1} Z_2 = \left[\begin{array}{ccc:c} \beta_1 & & & \\ & \ddots & & 0_{t,k-t} \\ & & \beta_t & \\ \hdashline & 0_{n-t,t} & & 0_{n-t,k-t} \end{array}\right],$$

kjer je $t = min\{n, k\}$ in velja $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_r > \beta_{r+1} = \ldots = 0$.

Iz enačbe (2) sledi, da velja

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

Opazimo, da velja tudi:

$$V^T K_B Z_1 \begin{pmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{pmatrix} = V^T K_B \left(\; Z_{11} | Z_{12} \; \right) \begin{pmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{pmatrix} =$$
$$V^T K_B \begin{pmatrix} Z_{11} D^{-1} Z_2 & 0 \\ 0 & Z_{12} \end{pmatrix} = \begin{pmatrix} V^T K_B Z_{11} D^{-1} Z_2 & 0 \\ 0 & V^T K_B Z_{12} \end{pmatrix} =$$
$$\begin{pmatrix} V^T K_{B_1} Z_2 & 0 \\ 0 & 0_{m-k,m-k} \end{pmatrix}.$$

Če dodatno definirajmo še $\beta_{t+1} = \ldots = \beta_q = 0$, dobimo ravno

$$V^T K_B Z_1 \begin{pmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{pmatrix} = \begin{pmatrix} diag(\beta_1, \ldots, \beta_t) & 0 \\ 0 & diag(\beta_{t+1}, \ldots, \beta_q) \end{pmatrix}.$$

Podobno bomo pokazali tudi za matriko K_A. Za G definirajmo matriko, ki jo dobimo s preoblikovanjem enačbe (**??**):

$$G := K_{B_1} Z_2 = V diag(\beta_1, \ldots, \beta_p).$$

Opazimo, da so stolpci matrike $A_1 Z_2$ medsebojno ortogonalni, ker velja:

$$\begin{gathered}(K_{A_1} Z_2)^T (K_{A_1} Z_2) = Z_2^T K_{A_1}^T K_{A_1} Z_2 \\ = Z_2^T (I_k - K_{B_1}^T K_{B_1}) Z_2 = Z_2^T Z_2 - Z_2^T K_{B_1}^T K_{B_1} Z_2 \\ = I_k - G^T G = I_k - diag(\beta_1, \ldots, \beta_p) V^T V diag(\beta_1, \ldots, \beta_p) \\ = diag(1 - \beta_1^2, \ldots, 1 - \beta_p^2).\end{gathered}$$

Tako se v diagonalne vrednosti preslikajo dolžine posameznih stolpcev (skalarni produkti enakih stolpcev), vsi ostali skalarni produkti pa so enaki 0. Če za matriko A dodatno definiramo $\alpha_i = 0 za i = k+1, \ldots, n$ lahko vidimo, da iz $K_{A_1} = K_A Z_{11} D^{-1}$ sledi:

$$U^T K_{A_1} Z_2 = U^T K_A Z_{11} D^{-1} Z_2 = diag(\alpha_1, \ldots, \alpha_k).$$

Podobno, kot za matriko B tu lahko pokažemo:

$$\begin{gathered}U^T K_A Z_1 \begin{pmatrix} D^{-1} Z_2 & 0 \\ 0 & I_{m-k} \end{pmatrix} \\ = \begin{pmatrix} U^T K_A Z_{11} D^{-1} Z_2 & 0 \\ 0 & U^T K_A Z_{12} I_{m-k} \end{pmatrix} = \begin{pmatrix} diag(\alpha_1, \ldots, \alpha_k) & 0 \\ 0 & diag(\alpha_{k+1}, \ldots, \alpha_n) \end{pmatrix},\end{gathered}$$

kjer smo pri prehodu v zadnjo vrstico ponovno uporabili enačbo (2). Sedaj definiramo matriko X na sledeči način:

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

– **NEJASNOST (oni tu vzamejo brez matrike Z_1)!** □

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 $K = \begin{pmatrix} K_A \\ K_B \end{pmatrix}$ in $t = rang(K)$ obstajajo ortogonalne matrike $U \in \mathbb{R}^{p \times p}$, $V \in \mathbb{R}^{n \times n}$, $W \in \mathbb{R}^{t \times t}$ in $Q \in \mathbb{R}^{m \times m}$, da velja:*

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

kjer je

$$\Sigma_A = \begin{pmatrix} I_A & & \\ & D_A & \\ & & 0_A \end{pmatrix} \quad \textit{in} \quad \Sigma_B = \begin{pmatrix} 0_B & & \\ & D_B & \\ & & I_B \end{pmatrix}.$$

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

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

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

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

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

Dokaz. Dovolj je, če ta izrek dokažemo za vsa kompleksna števila. Iz dejstva, da je množica realnih števil ($\mathbb{R}$) podmnožica množice kompleksnih števil ($\mathbb{C}$), sledi, da potem ta izrek velja tudi za vsa realna števila. Definirajmo matriko K, ki je sestavljena kot matrika sestavljena iz K_A in K_B, torej

$$K := \begin{pmatrix} K_A \\ K_B \end{pmatrix}.$$

Na zgoraj definirani matriki K lahko sedaj naredimo singularni razcep. Tako vemo, da za matriko K obstajata unitarni matriki $P \in \mathbb{C}^{(m+p)\times(m+p)}$ in $Q \in \mathbb{C}^{n\times n}$, da velja

$$P^H K Q = \begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix}$$

, kjer ima matrika R enak rang kot matrika K. Matriki Q in P sedaj ločimo na sledeče podmatrike:

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

kjer je matrika $Q_1 \in \mathbb{C}^{m\times t}$ sestavljena iz prvih k stolpcev matrike Q, matrika $P_1 \in \mathbb{C}^{(p+n)\times t}$ pa izprvih t stolpcev matrike P in njena podmatrika $P_{11} \in \mathbb{C}^{p\times t}$ pa iz prvih m vrstic matrike P_1. Vemo, da ker je matrika P unitarna matrika, velja $\|P\|_2 \leq 1$ in posledično velja še $\|P_{11}\|_2 \leq \|P_1\|_2 \leq \|P\|_2 \leq 1$. Iz izreka iz numeričnih metod velja, da posledično nobena lastna vrednost matrike P_{11} ni večja od 1.

Singularni razcep podobno kot na matriki K naredimo tudi na matriki P_{11}. Tako dobimo takšni matriki $U \in \mathbb{C}^{p\times p}$ in $W \in \mathbb{C}^{\times t}$, da velja

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

kjer je

$$\Sigma_A = \begin{pmatrix} 0_B & & \\ & D_B & \\ & & I_B \end{pmatrix},$$

kjer je matrika D_B diagonalna matrika z diagonalnimi vrednostmi $\alpha_{r+1}, ..., \alpha_{r+s}$, za katere velja $1 > \alpha_{r+1} \geq \ldots \geq \alpha_{r+s} > 0$.

(Spodnji del je vprašljiv? – **DOPOLNI!**)

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

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

kjer je matrika L_1 spodnjetrikotna z diagonalnimi elementi večjimi od 0. Opazimo lahko, da velja spodnja enakost

$$\begin{pmatrix} U^T & 0 \\ 0 & V^T \end{pmatrix} \begin{bmatrix} P_{11} \\ P_{21} \end{bmatrix} W = \begin{bmatrix} U^H P_{11} W \\ V^H P_{21} W \end{bmatrix} = \begin{pmatrix} \Sigma_A \\ L \end{pmatrix}.$$

Zgornja matrika $\begin{pmatrix} \Sigma_A \\ L \end{pmatrix}$ je unitarna, saj je produkt unitarnih matrik. Posledično so njeni stolpci ortonormirani, kar implicira to, da za matrik L veljajo vse lastnosti, ki morajo veljati za matriko Σ_B.

S preoblikovanjem zgornje formule in uporabi (zgornjega) singularnega razcepa na sestavljeni matriki $C = \begin{pmatrix} A \\ B \end{pmatrix}$ se izkaže, da velja:

$$\begin{pmatrix} A \\ B \end{pmatrix} Q = \begin{pmatrix} P_1 R, & 0 \end{pmatrix} = \begin{pmatrix} U\Sigma_A W^H R & 0 \\ V\Sigma_B W^H R & 0 \end{pmatrix}$$

. $\square$

Iz posplošenega singularnega razcepa, ki sta ga definirala Paige in Saunders neposredno sledi Van Loanova posplošitev singularnega razcepa. S preoblikovanjem enačbe (**??**) –**tu bi moralo pisati enačba 3**, dobimo:

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

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

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

Iz enačbe (**??**) tako tu očitno sledi:

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

Matriko X definiramo kot:
$$Q\begin{pmatrix} R^{-1}W & 0 \\ 0 & I \end{pmatrix}$$
in dobimo ravno Van Loanovo posplošitev razcepa:
$$U^T K_A X = \begin{pmatrix} \Sigma_A, & 0 \end{pmatrix}.$$
Podobno lahko pokažemo tudi za matriko K_B, za katero dobimo enačbo:
$$U^T K_B X = \begin{pmatrix} \Sigma_B, & 0 \end{pmatrix}.$$
Tako je definirana matrika X ravno iskana matrika iz prve posplošitve singularnega razcepa in matriki $\begin{pmatrix} \Sigma_A, & 0 \end{pmatrix}$ in $U^T K_B X = \begin{pmatrix} \Sigma_B, & 0 \end{pmatrix}$ ravno iskani diagonalni matriki.

Za nadaljnje delo definirajmo matrike
$$H_W := [A_1 - c^{(1)}e^{(1)^T}, \ldots, A_k - c^{(k)}e^{(k)^T}],$$
$$H_B := [(c^{(1)} - c)e^{(1)^T}, \ldots, (c^{(k)} - c)e^{(k)^T}]$$
in
$$H_M := [a_1 - c, ..., a_n - c] = A - ce^T = H_W + H_B,$$
kjer velja: $e^{(i)} = (1, \ldots, 1)^T \in \mathbb{R}^{n_i \times 1}$ in $e = (1, \ldots, 1)^T \in \mathbb{R}^{n \times 1}$.

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:
$$S_B = H_B H_B^T,$$
prav tako pa tudi matrika 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 ($rang(AB) \leq \min(rang(A), rang(B))$, 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 $J_1 = sled(S_2^{-1}S_1)$ za nesingularno matriko S_2

Tu izhajamo iz optimizacije optimizacijskega kriterija

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

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 spodnjetrikotna matrika V s pozitivnimi elementi na diagonali, da velja:

$$S_2 = VV^T.$$

3.2 Posplošitev maksimizacijskega kriterija $sled((S_W^Y)^{-1}S_B^Y)$ 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

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

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

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

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

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

$$(AV_2)^T(AV_2) = V_2^T A^T A V_2 = V_2^T (0, \ldots, 0) = 0$$

sledi $AV_2 = 0$. Sedaj definiramo $u_i := \frac{1}{\sigma_i} Av_i$ za $i = 1, \ldots, r$. Vekorji $u_1, \ldots, u_r$ so ortonormirani, saj je

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

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

$$U^T AV = \begin{pmatrix} U_1^T AV_1 & U_1^T AV_2 \\ U_2^T AV_1 & U_2^T AV_2 \end{pmatrix}.$$

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

$$u_k^T A v_i = \sigma_i u_k^T u_i = \sigma_i \delta ik,$$

torej $U_2^T AV_1 = 0$ in $U_1^T AV_1 = diag(\sigma_1, \ldots, \sigma_r)$. Dobimo singularni razcep $A = U \Sigma V^T$, kjer je $S = diag(\sigma_1, \ldots, \sigma_r)$ in

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

□

7 Viri