2.2 Maatriksi $\mathbf{P_X}$ omadused

Selles peatükis tõestame mõned, antud kursuse raames vajalikud maatriksi $\mathbf{P_X} := \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ omadused.

Teoreem 2.1 *Mistahes reaalarvuliste elementidega maatriksi* $\mathbf{X}$ *korral kehtivad järgmised võrdused:*

$$\mathbf{X}^T\mathbf{P_X} = \mathbf{X}^T \tag{2.4}$$

$$\mathbf{P_X}\mathbf{X} = \mathbf{X} \tag{2.5}$$

Tõestus:
Alustuseks tõestame paar lemmat.

Lemma 2.1 $\mathbf{A} = \mathbf{0} \Leftrightarrow tr(\mathbf{A}\mathbf{A}^T) = 0$

Tõestus: märkame, et maatriksi $\mathbf{A}\mathbf{A}^T$ diagonaalil asuvate elementide summaks on maatriksi $\mathbf{A}$ kõigi elementide ruutude summa. Viimane on null vaid juhul, kui kõik maatriksi $\mathbf{A}$ elemendid on nullid, ehk kui $\mathbf{A} = \mathbf{0}$.
$\square$

Lemma 2.2 $\mathbf{B}\mathbf{A}^T = \mathbf{C}\mathbf{A}^T \Leftrightarrow \mathbf{B}\mathbf{A}^T\mathbf{A} = \mathbf{C}\mathbf{A}^T\mathbf{A}$

Tõestus.
1. $\mathbf{B}\mathbf{A}^T = \mathbf{C}\mathbf{A}^T \Rightarrow \mathbf{B}\mathbf{A}^T\mathbf{A} = \mathbf{C}\mathbf{A}^T\mathbf{A}$

Väite esimese osa näitamine on triviaalne - korrutame võrrandit lihtsalt paremalt maatriksiga $\mathbf{A}$ ja saamegi soovitud tulemuse.

2. $\mathbf{B}\mathbf{A}^T = \mathbf{C}\mathbf{A}^T \Leftarrow \mathbf{B}\mathbf{A}^T\mathbf{A} = \mathbf{C}\mathbf{A}^T\mathbf{A}$

Vaatame maatriksit $(\mathbf{B}\mathbf{A}^T - \mathbf{C}\mathbf{A}^T)(\mathbf{B}\mathbf{A}^T - \mathbf{C}\mathbf{A}^T)^T$:

$$\begin{aligned}(\mathbf{B}\mathbf{A}^T - \mathbf{C}\mathbf{A}^T)(\mathbf{B}\mathbf{A}^T - \mathbf{C}\mathbf{A}^T)^T &= (\mathbf{B}\mathbf{A}^T - \mathbf{C}\mathbf{A}^T)\mathbf{A}(\mathbf{B} - \mathbf{C})^T \\ &= (\mathbf{B}\mathbf{A}^T\mathbf{A} - \mathbf{C}\mathbf{A}^T\mathbf{A})(\mathbf{B} - \mathbf{C})^T \\ &= \mathbf{0}.\end{aligned}$$

Kasutades Lemmat 2.1 saame järeldada, et $\mathbf{B}\mathbf{A}^T - \mathbf{C}\mathbf{A}^T = \mathbf{0}$ mis on muidugi samaväärne väitega $\mathbf{B}\mathbf{A}^T = \mathbf{C}\mathbf{A}^T$.

□

Märkus: transponeerides lemma 2.2 väites võrduse mõlemaid pooli saame lugeda tõestatuks ka väite:

$$\mathbf{AB} = \mathbf{AC} \Leftrightarrow \mathbf{A}^T\mathbf{AB} = \mathbf{A}^T\mathbf{AC}. \tag{2.6}$$

Nüüd võime jätkata teoreemi tõestusega. Paneme tähele, et üldistatud pöördmaatriksi definitsioonist lähtuvalt kehtib järgmine väide:

$$\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{X}. \tag{2.7}$$

Rakendame võrdusele 2.7 lemmat 2.2 (ja tema väidet transponeeritud kujul, nagu kirjas samaväärsuses 2.6), võttes maatriksi $\mathbf{A}$ rolli $\mathbf{X}$, maatriksiks $\mathbf{B}$ maatriksi $\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-$ ja maatriksiks $\mathbf{C}$ ühikmaatriksi $\mathbf{I}$, ja saame:

$$(\mathbf{X}^T\mathbf{P}_\mathbf{X} =)\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T = \mathbf{X}^T$$

ja

$$(\mathbf{P}_\mathbf{X}\mathbf{X} =)\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X} = \mathbf{X}$$

□

Kasutades Teoreemi 2.1 on lihtne näha, et võrrandisüsteemi 2.1 lahendiks sobib $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$:

$$\begin{aligned}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}\\ &= \mathbf{X}^T\mathbf{P}_\mathbf{X}\mathbf{y}\\ &= \mathbf{X}^T\mathbf{y}.\end{aligned}$$

Sellega oleme vastanud ka küsimusele võrrandisüsteemi 2.1 lahenduvusest — vaadeldav süsteem on tõepoolest alati lahenduv.

Teoreem 2.2 *Maatriks* $\mathbf{P}_\mathbf{X}$ *ei sõltu üldistatud pöördmaatriksi* $(\mathbf{X}^T\mathbf{X})^-$ *valikust.*

Tõestus:

Olgu maatriksid $\mathbf{F}$ ja $\mathbf{G}$ kaks erinevat maatriksi $\mathbf{X}^T\mathbf{X}$ üldistatud pöördmaatriksit, st $\mathbf{X}^T\mathbf{X}\mathbf{F}\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{X}\mathbf{G}\mathbf{X}^T\mathbf{X}$ $(= \mathbf{X}^T\mathbf{X})$. Siis lemmast 2.2 ja samaväärsusest 2.6 järeldub, et ka

$$\mathbf{X}\mathbf{F}\mathbf{X}^T = \mathbf{X}\mathbf{G}\mathbf{X}^T.$$

□

Järeldus teoreemist 2.2: kuigi vähimruutude hinnang $\hat{\boldsymbol{\beta}}$ ei pruugi olla üheselt määratud, on vähimruutude hinnangut kasutades leitud prognoosid $(\hat{\mathbf{y}})$ alati üheselt määratud.

Teoreem 2.3 *Maatriks $\mathbf{P_X}$ on sümmeetriline.*

Tõestus:

Teoreemi väide oleks triviaalne, kui maatriks $\mathbf{X}^T\mathbf{X}$ oleks pööratav (sest sümmeetrilise maatriksi pöördmaatriks on alati sümmeetriline). Paraku võib sümmeetrilise maatriksi üldistatud pöördmaatriks olla ka mittesümmeetriline. Selgub aga, et iga sümmeetrilise maatriksi korral on vähemalt üks tema üldistatud pöördmaatriksitest sümmeetriline. Selliseks üldistatud pöördmaatriksiks on Moore-Penrose'i üldistatud pöördmaatriks.

Maatriks $\mathbf{A}^+$ on maatriksi $\mathbf{A}$ Moore-Penrose'i üldistatud pöördmaatriks, kui ta rahuldab järgmiseid nõudeid:

$$\begin{aligned}
\mathbf{A} &= \mathbf{A}\mathbf{A}^+\mathbf{A} \quad \text{(üldistatud pöördmaatriks)}\\
\mathbf{A}^+ &= \mathbf{A}^+\mathbf{A}\mathbf{A}^+\\
(\mathbf{A}\mathbf{A}^+)^T &= \mathbf{A}\mathbf{A}^+\\
(\mathbf{A}^+\mathbf{A})^T &= \mathbf{A}^+\mathbf{A}
\end{aligned}$$

Moore-Penrose'i üldistatud pöördmaatriks on ühene ja eksisteerib alati. Neid väiteid me siinkohas tõestama ei hakka (teema sobib pigem mõnda maatriksite kursusesse). Küll aga näitame, et sümmeetrilise maatriksi $\mathbf{A}$ korral on $\mathbf{A}^+$ ka sümmeetriline. Selleks näitame, et $(\mathbf{A}^+)^T$ on ka maatriksi $\mathbf{A}$ Moore-Penrose'i üldistatud pöördmaatriks, ja seega $(\mathbf{A}^+)^T = \mathbf{A}^+$, sest Moore-Penrose'i üldistatud pöördmaatriks on ühene.

Näitame, et $(\mathbf{A}^+)^T$ rahuldab definitsiooni esimest tingimust:

$$\begin{aligned}
\mathbf{A} &= \mathbf{A}^T\\
\mathbf{A} &= (\mathbf{A}\mathbf{A}^+\mathbf{A})^T\\
\mathbf{A} &= \mathbf{A}(\mathbf{A}^+)^T\mathbf{A}.
\end{aligned}$$

Teine tingimus:

$$\begin{aligned}
\mathbf{A}^+ &= \mathbf{A}^+\mathbf{A}\mathbf{A}^+\\
(\mathbf{A}^+)^T &= (\mathbf{A}^+\mathbf{A}\mathbf{A}^+)^T\\
(\mathbf{A}^+)^T &= (\mathbf{A}^+)^T\mathbf{A}(\mathbf{A}^+)^T
\end{aligned}$$

Kolmanda tingimuse saame neljandast:

$$\begin{aligned}
(\mathbf{A}^+\mathbf{A})^T &= \mathbf{A}^+\mathbf{A} \\
\mathbf{A}(\mathbf{A}^+)^T &= ((\mathbf{A}^+\mathbf{A})^T)^T \\
\mathbf{A}(\mathbf{A}^+)^T &= (\mathbf{A}(\mathbf{A}^+)^T)^T
\end{aligned}$$

ja neljanda tingimuse saame samal viisil kolmandast:

$$\begin{aligned}
(\mathbf{A}\mathbf{A}^+)^T &= \mathbf{A}\mathbf{A}^+ \\
(\mathbf{A}^+)^T\mathbf{A} &= ((\mathbf{A}\mathbf{A}^+)^T)^T \\
(\mathbf{A}^+)^T\mathbf{A} &= ((\mathbf{A}^+)^T\mathbf{A})^T
\end{aligned}$$

Kuna sümmeetrilise maatriksi (näiteks $\mathbf{X}^T\mathbf{X}$) Moore-Penrose'i pöördmaatriks on sümmeetriline, siis on ka maatriks $\mathbf{X}(\mathbf{X}^T\mathbf{X})^+\mathbf{X}^T$ sümmeetriline. Kuna aga maatriks $\mathbf{P_X}$ ei sõltu üldistatud pöördmaatriksi valikust (Teoreem 2.2), siis järelikult peab $\mathbf{P_X}$ olema sümmeetriline mistahes pöördmaatriksi $(\mathbf{X}^T\mathbf{X})^-$ valiku korral.

□

Maatriksi $\mathbf{X}$ (dimensioonidega $n \times p$) veergude poolt moodustatav vektorruum (inglise k. *column space*) $\mathcal{C}(\mathbf{X})$ on defineeritud järgmiselt:

$$\mathcal{C}(\mathbf{X}) = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathcal{R}^p\}.$$

Teoreem 2.4 *Kui* $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{A})$, *siis* $\mathbf{P_X} = \mathbf{P_A}$.

Tõestus:

Kui $\mathcal{C}(\mathbf{X}) \subseteq \mathcal{C}(\mathbf{A})$, siis leidub selline maatriks $\mathbf{M}$, et $\mathbf{X} = \mathbf{AM}$ ($\mathcal{C}(\mathbf{X})$ baasivektorid saab esitada $\mathcal{C}(\mathbf{A})$ baasivektorite kaudu). Samuti $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\mathbf{X}) \Rightarrow \exists \mathbf{K}, \mathbf{A} = \mathbf{XK}$.

Nendest kahest järeldusest saame, et

$$\mathbf{X} = \mathbf{XKM}. \tag{2.8}$$

$$\begin{aligned}
\mathbf{P_A} &= \mathbf{A}(\mathbf{A}^T\mathbf{A})^-\mathbf{A}^T \\
&= \mathbf{XK}(\mathbf{K}^T\mathbf{X}^T\mathbf{XK})^-\mathbf{K}^T\mathbf{X}^T
\end{aligned}$$

Veendume, et maatriks $\mathbf{K}(\mathbf{K}^T\mathbf{X}^T\mathbf{X}\mathbf{K})^-\mathbf{K}^T$ on maatriksi $\mathbf{X}^T\mathbf{X}$ üheks üldistatud pöördmaatriksiks:

$$\begin{aligned}\mathbf{X}^T\mathbf{X}\mathbf{K}(\mathbf{K}^T\mathbf{X}^T\mathbf{X}\mathbf{K})^-\mathbf{K}^T\mathbf{X}^T\mathbf{X} &= \mathbf{X}^T\mathbf{P}_{\mathbf{XK}}\mathbf{X}\\ &\overset{(2.8)}{=} \mathbf{X}^T\mathbf{P}_{\mathbf{XK}}\mathbf{X}\mathbf{K}\mathbf{M}\\ &\overset{(\text{teoreem } 2.1)}{=} \mathbf{X}^T\mathbf{X}\mathbf{K}\mathbf{M}\\ &\overset{(2.8)}{=} \mathbf{X}^T\mathbf{X}.\end{aligned}$$

Seega oleme näidanud, et $\mathbf{P_A} = \mathbf{P_X}$.
□

Viimatitõestatud teoreemi tähendus väärib veidi lahtimõtestamist. Nimelt võime lisada oma mudelile (mudeli maatriksiga $\mathbf{X}$) uusi tunnuseid ja saada uue mudeli maatriksi $\mathbf{A}$. Kui lisatud tunnused on varem mudelis olnud tunnuste lineaarkombinatsioonid, siis selline „mudeli täiendamine“ ei muuda vaatlusvektorile $\mathbf{y}$ mudeli poolt antud prognoose $\hat{\mathbf{y}}$, kuna $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{A})$. Juhul, kui mudeli maatriks sisaldab teistest veergudest lineaarselt sõltuvat veergu, siis võime selle veeru ja vastava parameetri mudelist eemaldada, ilma et mudeli prognoosid $\hat{\mathbf{y}}$ või mudeli poolt tehtavad prognoosivead $\hat{\boldsymbol{\varepsilon}}$ sellest muutuksid.

Ülesanded

Defineerime maatriksi $\mathbf{P}_{\mathbf{X},\mathbf{W}}$ järgmiselt: $\mathbf{P}_{\mathbf{X},\mathbf{W}} := \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^-\mathbf{X}^T\mathbf{W}$, kus $\mathbf{W}$ on mingi sümmeetriline positiivselt määratud maatriks. Tõesta, et

1. $\mathbf{P}_{\mathbf{X},\mathbf{W}} \cdot \mathbf{X} = \mathbf{X}$;

2. $\mathbf{P}_{\mathbf{X},\mathbf{W}}$ ei sõltu üldistatud pöördmaatriksi $(\mathbf{X}^T\mathbf{W}\mathbf{X})^-$ valikust.

Ülesande lahendamisel võid kasutada järgmist tulemust:
Sümmeetrilise positiivselt määratud maatriksi $\mathbf{W}$ korral leidub selline pööratav sümmeetriline maatriks $\mathbf{W}^{1/2}$, nii et $\mathbf{W}^{1/2}\mathbf{W}^{1/2} = \mathbf{W}$.