Peatükk 5

Prognoosimisest

Lineaarsete mudelite üheks peamiseks kasutusalaks on prognoosimine — teades sõltumatute tunnuste (X-tunnuste) väärtuseid, üritatakse võimalikult täpselt välja pakkuda, ennustada, Y-tunnuse väärtust. Prognoosimiseks on meil kasutada ajaloolised vaatlused, mille põhjal hinnata mudeli parameetreid ($\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$). Prognoosida soovime uut Y-tunnuse väärtust Y_{uus}. Selleks, et suudaksime midagi öelda uue vaatluse kohta, peame midagi tema kohta teadma — peame oskama teda siduda vanade, ajalooliste väärtustega. Eeldame siin, et uus vaatlus genereeritakse, tekitatakse, sama meetodi ehk mudeli järgi kui vanad ajaloolised vaatlused:

$$Y_{uus} = \mathbf{x}_{uus}^T \boldsymbol{\beta} + \varepsilon_{uus}.$$

Paneme tähele: oleme eeldanud, et uue vaatluse tekitamisel osalevad parameetrite väärtused on needsamad mis ajalooliste vaatluste tekitamiseks kasutatud parameetrid — sama vektor $\boldsymbol{\beta}$ on mängus nii uute kui ajalooliste väärtuste loomisel. Lisaks eeldame, et uue vaatluse jääk on sõltumatu varasematest jääkidest, $\varepsilon_{uus} \perp \boldsymbol{\varepsilon}$, ja eeldame, et jääkide hajuvus ei muutu: $\mathrm{D}\varepsilon_{uus} = \mathrm{D}\boldsymbol{\varepsilon}_i = \sigma^2$.

5.1 Prognoosiintervall

Tundub olevat mõistlik kasutada uue juhusliku suuruse Y_{uus} prognoosimiseks suurust

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

Sellise prognoosi kasutamisel aset leidev prognoosiviga on

$$Y_{uus} - \hat{Y}_{uus} = \mathbf{x}_{uus}^T\boldsymbol{\beta} - \mathbf{x}_{uus}^T\hat{\boldsymbol{\beta}} + \varepsilon_{uus}.$$

Juhul, kui $\mathbf{y}$ ja Y_{uus} on normaaljaotusega, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}; \sigma^2\mathbf{I}); Y_{uus} \sim N(\mathbf{x}_{uus}^T\boldsymbol{\beta}; \sigma^2)$, siis on normaaljaotusega ka prognoosiviga (kahe sõltumatu normaaljaotusega juhusliku suuruse summa on ka normaaljaotusega):

$$\begin{aligned} Y_{uus} - \hat{Y}_{uus} &\sim N\left(0; \mathrm{D}(\mathbf{x}_{uus}^T\hat{\boldsymbol{\beta}}) + \mathrm{D}(\varepsilon_{uus})\right) \\ &\sim N\left(0; \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus}\sigma^2 + \sigma^2\right) \\ &\sim N\left(0; \sigma^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})\right) \end{aligned}$$

Seega standardiseeritud prognoosiviga on standardse normaaljaotusega juhuslik suurus:

$$\frac{Y_{uus} - \hat{Y}_{uus}}{\sqrt{\sigma^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}} \sim N(0; 1).$$

Saadud tulemusest oleks kasu prognoosi täpsuse kirjeldamisel siis, kui teaksime jääkide hajuvust (σ^2). Paraku me ei tea jääkide hajuvust täpselt (kuigi hinnata me teda juba oskame). Järgmisena üritame sellest tundmatust suurusest kuidagi lahti saada. Appi saame võtta varem tõestatud tulemuse, vaata valemit 3.1 leheküljelt 35:

$$\frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}}{\sigma^2} \sim \chi^2_{df=n-\mathrm{rank}(\mathbf{X})}.$$

Kui $X \sim N(0; 1)$ ja $Y \sim \chi^2_{df}$, siis

$$\frac{X}{\sqrt{Y/df}} \sim t_{df}.$$

Järelikult ka (kasutades tähistust $p := \mathrm{rank}(\mathbf{X})$):

$$\begin{aligned} \frac{Y_{uus} - \hat{Y}_{uus}}{\sqrt{\sigma^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}} \Big/ \sqrt{\frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}}{\sigma^2(n-p)}} &\sim t_{df=n-p} \\ \frac{Y_{uus} - \hat{Y}_{uus}}{\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}} &\sim t_{df=n-p} \end{aligned}$$

kus $\hat{\sigma}^2 = \mathrm{MSE} = \mathbf{y}^T(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-p)$.

Jääkide tegelik dispersioon σ^2 mis meid varem segas on lõpuks kadunud. Saadud tulemust võib kasutada mitmeti. Näiteks võime testida, ega andmeid genereeriv mehhanism pole vahepeal muutunud (kas uus vaatlus on pärit samast populatsioonist kust pärinevad varasemad vaatlused). Sellise hüpoteesi kontrollimiseks võime kasutada teststatistikut

$$t = \frac{Y_{uus} - \hat{Y}_{uus}}{\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}}.$$

Nullhüpoteesi kehtides (andmeid genereeriv mehhanism on jäänud samaks) on $t \sim t_{df=n-p}$. Seda tüüpi testi saab kasutada näiteks kontrollimaks, ega tööstusseadmete seadistus pole töö käigus paigast ära läinud.

Saadud tulemust saab kasutada aga ka prognoosiintervalli leidmiseks. $(1-\alpha)$-prognoosiintervall on vahemik, kuhu järgmise juhusliku suuruse väärtus jääb tõenäosusega $(1-\alpha)$:

$$\begin{aligned} P\left(t_{\alpha/2;df=n-p} \leq \frac{Y_{uus} - \hat{Y}_{uus}}{\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}} \leq t_{1-\alpha/2;df=n-p}\right) &= 1-\alpha \\ P\Bigg(\hat{Y}_{uus} + t_{\alpha/2;df=n-p}\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})} \leq Y_{uus} \leq \\ \leq \hat{Y}_{uus} + t_{1-\alpha/2;df=n-p}\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}\Bigg) &= 1-\alpha \end{aligned}$$

Ehk, arvestades et $t_\alpha = -t_{1-\alpha}$, oleme saanud $(1-\alpha)$-prognoosiintervalli leidmiseks valemi:

$$\hat{Y}_{uus} \pm t_{\alpha/2;df=n-p}\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{x}_{uus})}.$$

Antud prognoosiintervalli leidmiseks oleme kasutanud normaaljaotuse eeldust. Oleme normaaljaotuse eeldust kasutanud ka varem — t-testi, F-testi või usaldusintervallide konstrueerimisel. Kõik eeldused pole aga paraku sündinud võrdsetena. Kui meie andmed pole normaaljaotusega, siis normaaljaotuse eeldusest lähtuv usaldusintervall või F-testi tulemus on tavaliselt siiski täiesti aktsepteeritav (võimalik viga on enamasti, vähemalt suuremate andmestike korral, tühine). Ülaltoodud valemi abil leitud prognoosiintervall võib aga osutuda naeruväärseks isegi suure valimi korral. Uuritava tunnuse jaotus ei pruugi isegi kuigivõrd erineda normaaljaotusest, aga juba on leitud prognoosiintervall küsitav ja kahtlane.

Mida teha kui andmed pole normaaljaotusega? Lahendusi antud probleemile on välja pakutud mitmeid. Üheks lihtsasti kasutatavaks lahenduseks

võiks olla näiteks D.J.Olive (Olive, 2006) poolt pakutud prognoosiintervallide kasutamine:

$$[\hat{Y}_{uus} + a_n\hat{q}_{\alpha/2}; \hat{Y}_{uus} + a_n\hat{q}_{1-\alpha/2}],$$

kus

$$a_n = \left(1 + \frac{15}{n}\right)\sqrt{\frac{n}{n-p}} \cdot \sqrt{1 + \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{x}_{uus}}$$

ja $\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}$ on hinnangud regressioonmudeli jääkide $(\alpha/2)$ ja $(1-\alpha/2)$-kvantiilile.

5.2 Parim Prognoos

Prognoosisime uut juhuslikku suurust Y_{uus} valemiga $\hat{Y}_{uus} = \mathbf{x}_{uus}^T\hat{\boldsymbol{\beta}}$. Kas see on ikka tark ja mõistlik prognoos? Milline oleks olemasoleva informatsiooni $(\mathbf{x}_{uus}, \mathbf{y}, \mathbf{X})$ valguses tark ja mõistlik prognoos?

Definitsioon 5.1 *Juhusliku suuruse Y_{uus} parimaks prognoosiks (Best Predictor) kutsutakse sellist juhuslikku suurust $\hat{Y}_{uus} = \hat{Y}_{uus}(\mathbf{x}_{uus}, \mathbf{y}, \mathbf{X})$, mille puhul keskmine prognoosi ruutviga on minimaalne:*

$$E(Y_{uus} - \hat{Y}_{uus})^2 \leq E(Y_{uus} - \tilde{Y}_{uus})^2$$

mistahes teise juhusliku suuruse Y_{uus} prognoosi $\tilde{Y}_{uus}$ korral.

Keskväärtus antud definitsioonis on võetud nii üle kõikmõeldavate Y_{uus} väärtuste kui ka üle kõikmõeldavate valimite $\mathbf{y}$.

Teoreem 5.1 *Juhusliku suuruse Y_{uus} parimaks prognoosiks on tinglik keskväärtus $Y_0 := E(Y_{uus}|\mathbf{y})$.*

Tõestus:

Vaatame keskmist ruutviga mingi suvalise prognoosi $\tilde{Y}_{uus} = \tilde{Y}_{uus}(\mathbf{y}, \mathbf{x}_{uus})$ korral:

$$\begin{aligned} \mathrm{E}(Y_{uus} - \tilde{Y}_{uus})^2 &= \mathrm{E}(Y_{uus} - Y_0 + Y_0 - \tilde{Y}_{uus})^2 \\ &= \mathrm{E}(Y_{uus} - Y_0)^2 + \mathrm{E}(Y_0 - \tilde{Y}_{uus})^2 + 2\mathrm{E}[(Y_{uus} - Y_0)(Y_0 - \tilde{Y}_{uus})] \end{aligned}$$

Vaatame viimast liidetavat ja peame meeles, et $\mathrm{E}X = \mathrm{E}[\mathrm{E}(X|Y)]$:

$$\begin{aligned}
\mathrm{E}[(Y_{uus}-Y_0)(Y_0-\tilde{Y}_{uus})] &= \mathrm{E}_{\mathbf{y}}\left[\mathrm{E}_{Y_{uus}|\mathbf{y}}\left((Y_{uus}-Y_0)(Y_0-\tilde{Y}_{uus})\right)\right] \\
&= \mathrm{E}_{\mathbf{y}}\left[(Y_0-\tilde{Y}_{uus})\cdot \mathrm{E}_{Y_{uus}|\mathbf{y}}\left(Y_{uus}-Y_0\right)\right] \\
&= \mathrm{E}_{\mathbf{y}}[(Y_0-\tilde{Y}_{uus})\cdot 0] \\
&= 0
\end{aligned}$$

Sest antud valimi korral ehk fikseeritud $\mathbf{y}$ väärtuse korral on $Y_0-\tilde{Y}_{uus}$ konstant ja $\mathrm{E}_{Y_{uus}|\mathbf{y}}Y_{uus}=Y_0$.

Oleme saanud, et

$$\mathrm{E}(Y_{uus}-\tilde{Y}_{uus})^2=\mathrm{E}(Y_{uus}-Y_0)^2+\mathrm{E}(Y_0-\tilde{Y}_{uus})^2$$

kust võime välja lugeda, et prognoosi ruutvea minimiseerimiseks peame valima $\tilde{Y}_{uus}=Y_0(=\mathrm{E}(Y_{uus}|\mathbf{y}))$.

Seega on parimaks prognoosiks tinglik keskväärtus. Meid huvitavas kontekstis

$$\mathrm{E}(Y_{uus}|\mathbf{y})=\mathrm{E}Y_{uus}=\mathbf{x}_{uus}^T\boldsymbol{\beta}$$

sest Y_{uus} ja $\mathbf{y}$ on tehtud eelduse kohaselt sõltumatud ($\varepsilon_{uus}\perp\boldsymbol{\varepsilon}$).

Paraku näitab saadud tulemus, et parim prognoos on tavaliselt kättesaamatu luksus — parameetrite $\boldsymbol{\beta}$ tegelikud väärtused pole meile ju teada (me saame neid vaid hinnata, aga tegelik väärtus on ju ikkagi veidi midagi muud kui hinnang).

5.3 Parim Lineaarne Nihketa Prognoos

Eelnevalt nägime, et parim prognoos on luksus, mida me harva endale lubada saame (populatsiooni parameetrite $\boldsymbol{\beta}$ täpne mõõtmine on enamasti väga kallis — kõiksed uuringud on tavaliselt palju kordi kallimad kui valikuuuringud). Antud peatükis üritame otsida parimat prognoosi veidi kitsamast võimalike prognooside klassist. Me ei otsi enam parimat prognoosi kõikmõeldavate prognooside seast, vaid otsime parimat prognoosi nihketa lineaarsete prognooside sest.

Definitsioon 5.2 *Juhusliku suuruse Y_{uus} parimaks lineaarseks nihketa prognoosiks (Best Linear Unbiased Predictor) kutsutakse sellist prognoosi $\hat{Y}_{uus}=\hat{Y}_{uus}(\mathbf{x}_{uus},\mathbf{y},\mathbf{X})$, mis rahuldab järgmiseid nõudeid:*

1. *Lineaarsus*

$$\hat{Y}_{uus}=\mathbf{A}\mathbf{y}+c;$$

2. *Nihketus*
$$E\hat{Y}_{uus} = EY_{uus} \quad (\forall \boldsymbol{\beta});$$

3. *Parim*
$$E(Y_{uus} - \hat{Y}_{uus})^2 \leq E(Y_{uus} - \tilde{Y}_{uus})^2$$
mistahes teise lineaarse nihketa prognoosi $\tilde{Y}_{uus}$ *korral.*

Kui prognoos on lineaarne ja nihketa, siis

$$\begin{aligned} \mathrm{E}\hat{Y}_{uus} &= \mathrm{E}Y_{uus} \quad (\forall \boldsymbol{\beta}) \\ \mathrm{E}(\mathbf{A}\mathbf{y} + c) &= \mathbf{x}_{uus}^T \boldsymbol{\beta} \quad (\forall \boldsymbol{\beta}) \\ \mathbf{A}\mathbf{X}\boldsymbol{\beta} + c &= \mathbf{x}_{uus}^T \boldsymbol{\beta} \quad (\forall \boldsymbol{\beta}). \end{aligned}$$

Viimasest võrdusest näeme, et $c = 0$ (ülaltoodud tingimus peab kehtima ka $\boldsymbol{\beta} = 0$ korral) ja saame, et
$$\mathbf{A}\mathbf{X} = \mathbf{x}_{uus}^T. \tag{5.1}$$

Teoreem 5.2 *Parim lineaarne nihketa hinnang juhuslikule suurusele* Y_{uus} *on kujul*

$$\begin{aligned} \hat{Y}_{uus} &= \mathbf{x}_{uus}^T \hat{\boldsymbol{\beta}} \\ &= \mathbf{x}_{uus}^T (\mathbf{X}^T\mathbf{X})^- \mathbf{X}^T \mathbf{y} \\ &= \mathbf{A}\mathbf{y}, \end{aligned}$$

kus $\mathbf{A} := \mathbf{x}_{uus}^T (\mathbf{X}^T\mathbf{X})^- \mathbf{X}^T$.

Tõestus.
Olgu antud kaks lineaarset nihketa hinnangut, $\hat{Y}_{uus} = \mathbf{A}\mathbf{y}$ ja $\tilde{Y}_{uus} = \mathbf{B}\mathbf{y}$.

$$\begin{aligned} \mathrm{E}(Y_{uus} - \mathbf{B}\mathbf{y})^2 &= \mathrm{E}(Y_{uus} - \mathbf{A}\mathbf{y} + \mathbf{A}\mathbf{y} - \mathbf{B}\mathbf{y})^2 \\ &= \mathrm{E}(Y_{uus} - \mathbf{A}\mathbf{y})^2 + \mathrm{E}(\mathbf{A}\mathbf{y} - \mathbf{B}\mathbf{y})^2 + \\ &\quad +2\mathrm{E}[(Y_{uus} - \mathbf{A}\mathbf{y})(\mathbf{A}\mathbf{y} - \mathbf{B}\mathbf{y})]. \end{aligned}$$

Kui suudaksime näidata, et valiku $\mathbf{A} := \mathbf{x}_{uus}^T (\mathbf{X}^T\mathbf{X})^- \mathbf{X}^T$ puhul oleks viimane liidetav 0, oleksimegi teoreemi tõestanud (sest siis on mistahes teise lineaarse nihketa prognoosi keskmine ruutviga kirja pandav kui prognoosi $\mathbf{A}\mathbf{y}$ keskmine ruutviga pluss midagi mittenegatiivset, seega iga teise mõeldava prognoosi keskmine ruutviga saab tulla vaid suurem kui prognoosi $\mathbf{A}\mathbf{y}$ keskmine ruutviga.

Näitame, et $\mathrm{E}[(Y_{uus} - \mathbf{Ay})(\mathbf{Ay} - \mathbf{By})] = 0$:

$$\begin{aligned} \mathrm{E}[(Y_{uus} - \mathbf{Ay})(\mathbf{Ay} - \mathbf{By})] &= \mathrm{E}[(Y_{uus} - \mathbf{Ay})\mathbf{y}^T(\mathbf{A}^T - \mathbf{B}^T)] \\ &= \mathrm{E}[Y_{uus}\mathbf{y}^T(\mathbf{A}^T - \mathbf{B}^T)] - \mathrm{E}[\mathbf{Ayy}^T(\mathbf{A}^T - \mathbf{B}^T)] \end{aligned} \tag{5.2}$$

Vaatame esmalt esimest liidetavatest:

$$\begin{aligned} \mathrm{E}[Y_{uus}\mathbf{y}^T(\mathbf{A}^T - \mathbf{B}^T)] &= \mathrm{E}_{Y_{uus}}\left\{\mathrm{E}_{Y|Y_{uus}}\left[Y_{uus}\mathbf{y}^T(\mathbf{A}^T - \mathbf{B}^T)\right]\right\} \\ &= \mathrm{E}_{Y_{uus}}\left\{Y_{uus}\mathrm{E}_{Y|Y_{uus}}\left(\mathbf{y}\right)^T(\mathbf{A}^T - \mathbf{B}^T)\right\} \\ &= \mathrm{E}_{Y_{uus}}\left\{Y_{uus}\boldsymbol{\beta}^T\mathbf{X}^T(\mathbf{A}^T - \mathbf{B}^T)\right\} \\ &= \mathrm{E}_{Y_{uus}}\left\{Y_{uus}\boldsymbol{\beta}^T \cdot 0\right\} \\ &= 0 \end{aligned}$$

sest $\mathbf{AX} = \mathbf{BX} = \mathbf{x}_{uus}^T$ (vaata valemit 5.1) ja järelikult $(\mathbf{A} - \mathbf{B})\mathbf{X} = 0$.

Pöördume tagasi valemi (5.2) juurde ja asume uurima teise liidetava väärtust.

$$\begin{aligned} \mathrm{E}\left[\mathbf{Ayy}^T(\mathbf{A}^T - \mathbf{B}^T)\right] &= \mathbf{A}\mathrm{E}\left(\mathbf{yy}^T\right)(\mathbf{A}^T - \mathbf{B}^T) \\ &= \mathbf{A}\left(\sigma^2 \cdot \mathbf{I} + \mathrm{E}\mathbf{y}(\mathrm{E}\mathbf{y})^T\right)(\mathbf{A}^T - \mathbf{B}^T) \\ &= \sigma^2\mathbf{A}(\mathbf{A}^T - \mathbf{B}^T) + \mathbf{AX}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{X}^T(\mathbf{A}^T - \mathbf{B}^T) \\ &= 0 + 0 \end{aligned}$$

sest $\mathbf{X}^T(\mathbf{A}^T - \mathbf{B}^T) = 0$ ja valiku $\mathbf{A} = \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ korral ka

$$\mathbf{A}(\mathbf{A}^T - \mathbf{B}^T) = \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T(\mathbf{A}^T - \mathbf{B}^T) = \mathbf{x}_{uus}^T(\mathbf{X}^T\mathbf{X})^- \cdot 0 = 0.$$

Oleme seega saanud, et

$$\mathrm{E}(Y_{uus} - \mathbf{By})^2 = \mathrm{E}(Y_{uus} - \mathbf{Ay})^2 + \mathrm{E}(\mathbf{Ay} - \mathbf{By})^2$$

ja järelikult ei saa ükski teine lineaarne nihketa prognoos omada väiksemat keskmist prognoosiviga kui prognoos $\mathbf{Ay} = \mathbf{x}_{uus}^T\hat{\boldsymbol{\beta}}$.

□

Ülesanne

Kasutades 30 ajaloolist mõõtmist hinnati lineaarse mudeli

$$y = c_0 + c_1 \cdot x + c_2 \cdot x^2 + \varepsilon$$

parameetreid. Hinnatud mudel nägi välja järgmine:

$$y = 2 + 0x + 2x^2 + \varepsilon$$

Teame parameetervektori $\boldsymbol{\beta} = (c_0, c_1, c_2)^T$ hinnangu (hinnatud) dispersioonimaatriksit:

$$\begin{pmatrix} 1 & 0 & -0{,}2 \\ 0 & 1 & -0{,}3 \\ -0{,}2 & -0{,}3 & 1 \end{pmatrix}$$

ja oleme leidnud ka jääkide dispersiooni hinnangu, $MSE = 68$.

Kavatseme teha kaks (sõltumatut) uut mõõtmist väärtustel $x = 1$ ja $x = 3$.

- Leia 95%-prognoosiintervall esimesel täiendaval mõõtmisel ($x = 1$) nähtava y-tunnuse väärtusele.
- Leia 95%- prognoosiintervalli nende kahe vaatluse ($x = 1$; $x = 3$) y-tunnuste aritmeetilisele keskmisele.