Tentamen – Bayesiansk inferens och maskininlärning (TIF385)

Tid och plats:	12 januari, 2024, fm, Johanneberg.
Hjälpmedel:	Physics Handbook, Beta Mathematics Handbook, typgodkänd kalkylator.
Examinator:	Christian Forssén.
Jourhavande lärare:	Christian Forssén (031–772 3261).

Tentamen består av sex uppgifter som kan ge maximalt 60 poäng totalt. För att bli godkänd med betyg 3 krävs 24 poäng, för betyg 4 krävs 36 poäng och för betyg 5 krävs 48 poäng.

Rättningsprinciper: Alla svar skall motiveras (om ej annat anges) och införda storheter skall förklaras. Lösningarna förväntas vara välstrukturerade och begripligt presenterade. Skriv och rita tydligt. Vid tentamensrättning gäller följande allmänna principer:

- För maximal (10) poäng krävs fullständigt korrekt och välmotiverad lösning.
- Mindre fel ger 1-3 poängs avdrag. Gäller även mindre brister i presentationen.
- Lösningar som inte går att följa (t.ex. avsaknad av relevant figur, ej definierade variabler, svårtydd, etc) renderar poängavdrag även om svaret verkar vara korrekt.
- Allvarliga principiella fel ger fullt poängavdrag.
- Allvarliga fel som leder till orimliga resultat kan ge lägre poängavdrag om orimligheten pekas ut.
- Även skisserade lösningar kan ge delpoäng.

Lycka till!

---

1. Betrakta ett neuralt nätverk med två insignaler, ett gömt lager bestående av tio noder och ett utlager med en enda nod. Samtliga noder använder samma aktiveringsfunktion $f(z)$ och inkluderar en biasvikt vid beräkning av z.

   (a) Hur många fria parametrar har det neurala nätverket?

   (b) Inför lämpliga beteckningar och ge ett uttryck för utsignalen.

   (c) Ge tre exempel på hyperparametrar som skulle kunna ändras inför träning av modellen. Notera att nätverkets arkitektur är fixerad.

   (3 poäng per deluppgift, plus 1 poäng om samtliga är korrekta.)

2. Prestandan för en maskininlärningsmodell uppskattas ofta med en felfunktion (error function) som vi allmänt betecknar $E(\hat{y}, y)$, där $\hat{y}$ är en modellförutsägelse och y motsvarande måldata.

   (a) Antag att det finns tillgång till träningsdata ($\mathcal{D}_{\text{train}}$) och valideringsdata ($\mathcal{D}_{\text{val}}$). Introducera (definiera) E_{train} och E_{val} och beskriv när under träningen som dessa mått beräknas.

   (b) Skissa hur E_{train} och E_{val} typiskt beror på antalet gradientstegsiterationer vid träning av ett neuralt nätverk. Skillnaden $E_{\text{train}} - E_{\text{val}}$ kallas för generaliseringsgapet. Beskriv hur generaliseringsgapet kan användas under träning av ett neuralt nätverk.

   (c) Ge tre förslag på hur generaliseringsgapet kan minskas. Du behöver inte ta hänsyn till hur förslagen påverkar modellens träffsäkerhet utan bara att de kan minska generaliseringsgapet.

   (3 poäng per deluppgift, plus 1 poäng om åtminstone en är helt korrekt.)

3. Betrakta en statistisk modell $\mathcal{D} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, som relaterar N_d observationer i kolumnvektorn $\mathcal{D}$ till en linjär modell som beskrivs av designmatrisen $\boldsymbol{X}$ och parametervektorn $\boldsymbol{\theta}$ samt slumpvariabler $\boldsymbol{\epsilon}$. De sistnämnda beskriver osäkerheter och vi kan anta att elementen är oberoende och identiskt normalfördelade $p(\epsilon_i|I) = \mathcal{N}(0, \sigma_\epsilon^2)$.

   (a) Visa att trolighetsfunktionen blir

   $$p\left(\mathcal{D}|\boldsymbol{\theta}, I\right) = \left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{N_d/2} \exp\left[-\frac{1}{2}\frac{(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})^T(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})}{\sigma_\epsilon^2}\right].$$

   (b) Visa därefter att

   $$p\left(\mathcal{D}|\boldsymbol{\theta}, I\right) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T\Sigma_\theta^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\right],$$

   där $\boldsymbol{\theta}^* = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathcal{D}$ är lösningen till normalekvationen och $\Sigma_\theta^{-1} = \boldsymbol{X}^T\boldsymbol{X}/\sigma_\epsilon^2$.

   Ledtråd: Definiera $L(\boldsymbol{\theta}) = (\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})^T(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})/(2\sigma_\epsilon^2)$ och notera att gradientvektorn ges av $\boldsymbol{\nabla}_{\boldsymbol{\theta}} L = \boldsymbol{X}^T\left(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta}\right)/\sigma_\epsilon^2$ samt att Hessianmatrisen med andraderivator är $\boldsymbol{H} = \boldsymbol{X}^T\boldsymbol{X}/\sigma_\epsilon^2$.

   (10 poäng)

4. I Göteborg regnar det 60% av alla dagar. Metereologerna på SMHI försöker förutsäga om det kommer regna imorgon eller inte. Historiskt sett har de korrekt förutspått vädret 80% av alla regniga dagar, och 60% av alla dagar utan regn.

   Givet att prognosen säger regn imorgon, vad är sannolikheten att det faktiskt kommer regna imorgon? *(10 poäng)*

5. Betrakta övergångsmatrisen

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix}$$

   som genererar en icke-periodisk, positivt återkommande och icke-reducerbar Markovkedja med tre utfall (som betecknas $1, 2, 3$). Den första slumpvariabeln, X_0, i Markovkedjan beskrivs av sannolikhetsfördelningen $\boldsymbol{\pi}_0 = (0.5, 0.2, 0.3)$, dvs $\mathbb{P}(X_0 = 1) = 0.5$, osv.

   (a) Vad är sannolikheten $\mathbb{P}(X_1 = 1)$?

   (b) Vad är sannolikheten $\mathbb{P}\left(X_1 = 1 \,|\, X_0 = 1\right)$?

   (c) Vad är sannolikheten $\mathbb{P}(X_n = 1)$, där n är ett väldigt stort tal?

   (d) Vad är sannolikheten $\mathbb{P}\left(X_n = 1 \,|\, X_{n-1} = 1\right)$, där n är ett väldigt stort tal?

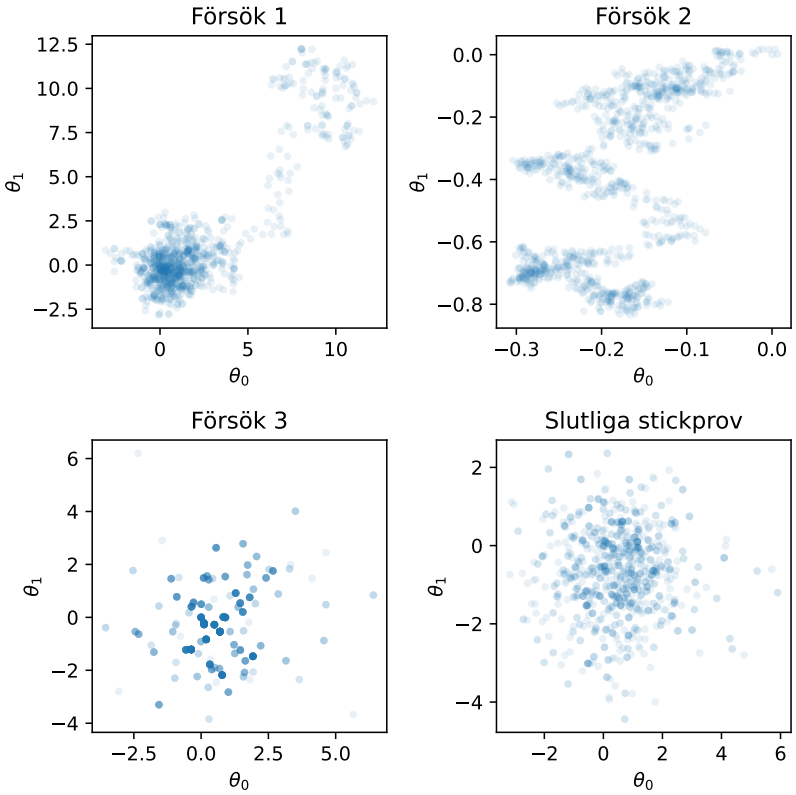
   (e) Är Markovkedjan tidsreversibel? (Motivera ditt svar.)

   (2 poäng per deluppgift.)

6. En ingenjör har som uppgift att leverera 1.000 stickprov från en sannolikhetstäthetsfunktion $p(\theta_0, \theta_1)$ med Metropolissampling och normalfördelade stegförslag.

   Figuren nedan visar flera försök att genomföra samplingen. Respektive panel visar ett spridningsdiagram ('scatter plot') med 1.000 stickprov (mörkare punkter innebär att flera stickprov ligger ovanpå varandra).

   Ingenjören valde slutligen resultaten som visas i den sista panelen. Ge rimliga förslag till vad som kunde förbättras i de tre första försöken (det finns en huvudproblematik per försök; ge därför ett förslag per panel 1–3).



(3 poäng per panel [försök 1–3], plus 1 poäng om samtliga är korrekta.)

Tentamen – Bayesiansk inferens och maskininlärning (TIF385)

Tid och plats:	12 januari, 2024, fm, Johanneberg.
Hjälpmedel:	Physics Handbook, Beta Mathematics Handbook, typgodkänd kalkylator.
Lösningsskiss:	Christian Forssén.

Tentamen består av sex uppgifter som kan ge maximalt 60 poäng totalt. För att bli godkänd med betyg 3 krävs 24 poäng, för betyg 4 krävs 36 poäng och för betyg 5 krävs 48 poäng.

Rättningsprinciper: Alla svar skall motiveras (om ej annat anges) och införda storheter skall förklaras. Lösningarna förväntas vara välstrukturerade och begripligt presenterade. Skriv och rita tydligt. Vid tentamensrättning gäller följande allmänna principer:

- För maximal (10) poäng krävs fullständigt korrekt och välmotiverad lösning.
- Mindre fel ger 1-3 poängs avdrag. Gäller även mindre brister i presentationen.
- Lösningar som inte går att följa (t.ex. avsaknad av relevant figur, ej definierade variabler, svårtydd, etc) renderar poängavdrag även om svaret verkar vara korrekt.
- Allvarliga principiella fel ger fullt poängavdrag.
- Allvarliga fel som leder till orimliga resultat kan ge lägre poängavdrag om orimligheten pekas ut.
- Även skisserade lösningar kan ge delpoäng.

Detta är enbart en skiss av den fullständiga lösningen. Det kan innebära att vissa mellansteg i uträkningarna, som egentligen är nödvändiga för en komplett lösning, inte redovisas.

1. Betrakta ett neuralt nätverk med två insignaler, ett gömt lager bestående av tio noder och ett utlager med en enda nod. Samtliga noder använder samma aktiveringsfunktion $f(z)$ och inkluderar en biasvikt vid beräkning av z.

   (a) Hur många fria parametrar har det neurala nätverket?

   (b) Inför lämpliga beteckningar och ge ett uttryck för utsignalen.

   (c) Ge tre exempel på hyperparametrar som skulle kunna ändras inför träning av modellen. Notera att nätverkets arkitektur är fixerad.

   (3 poäng per deluppgift, plus 1 poäng om samtliga är korrekta.)

   Lösning:

(a) Två insignaler samt en bias till varje nod i det gömda lagret ger $3 \cdot 10 = 30$ vikter. Tio insignaler och en bias till den enda noden i utlagret ger 11 vikter. Totalt: 41 parametrar.

(b) Introducera matriserna $\boldsymbol{W}^{(l)}$ för vikterna i lager l (det gömda lagret motsvarar $l = 1$ och utlagret $l = 2$). Dessa matriser har element $W_{i,j}^{(l)}$ där varje nod representeras av en kolumn, dvs index j motsvarar nodnummer. Vi kan notera att viktmatrisen för det gömda lagret har storleken 2×2 och för utlagret 10×1.

Vidare introduceras radvektorerna $\boldsymbol{b}^{(l)}$ som innehåller alla biasvikter för nod l.

Insignaler samlas i radvektorn $\boldsymbol{x} = (x_1, x_2)$. Utsignalerna från respektive nod betecknas av radvektorn $\boldsymbol{y}^{(l)}$ medan aktiveringarna för respektive nod betecknas av radvektorn $\boldsymbol{z}^{(l)}$. Notera att bägge dessa har längden 1 för utlagret med en enda nod och vi använder därför skalär notation för dessa.

Aktiveringsfunktionen betecknas $f(z)$ och ger en vektor om argumentet är en vektor. Då fås slutligen utsignalen $y^{(2)} = f(z^{(2)})$ där

$$
\begin{aligned}
z^{(2)} &= \boldsymbol{y}^{(1)} \cdot \boldsymbol{W}^{(2)} + b^{(2)} \\
\boldsymbol{y}^{(1)} &= f(\boldsymbol{z}^{(1)}) \\
\boldsymbol{z}^{(1)} &= \boldsymbol{x} \cdot \boldsymbol{W}^{(1)} + \boldsymbol{b}^{(1)}.
\end{aligned}
$$

(c) (1) Inlärningshastigheten η som används vid gradientstegsoptimeringen; (2) Antalet iterationer/epoker av gradientstegsoptimering; (3) Viktregularisering λ; (4) Mängden träningsdata och/eller uppdelningen i tränings-/valideringsdata. Även om det är diskutabelt att benämna (5) aktiveringsfunktionen som en "parameter" så kan man även tänka sig att byta ut denna. I vissa aktiveringsfunktioner finns det också hyperparametrar såsom (6) lutningen för $z < 0$ för läckande ReLUs. På samma sätt är det diskutabelt om man kan benämna (7) kostnadsfunktionen som en "parameter" men även denna skulle kunna bytas.

---

2. Prestandan för en maskininlärningsmodell uppskattas ofta med en felfunktion (error function) som vi allmänt betecknar $E(\hat{y}, y)$, där $\hat{y}$ är en modellförutsägelse och y motsvarande måldata.

(a) Antag att det finns tillgång till träningsdata ($\mathcal{D}_{\text{train}}$) och valideringsdata ($\mathcal{D}_{\text{val}}$). Introducera (definiera) E_{train} och E_{val} och beskriv när under träningen som dessa mått beräknas.

(b) Skissa hur E_{train} och E_{val} typiskt beror på antalet gradientstegsiterationer vid träning av ett neuralt nätverk. Skillnaden $E_{\text{train}} - E_{\text{val}}$ kallas för generaliseringsgapet. Beskriv hur generaliseringsgapet kan användas under träning av ett neuralt nätverk.

(c) Ge tre förslag på hur generaliseringsgapet kan minskas. Du behöver inte ta hänsyn till hur förslagen påverkar modellens träffsäkerhet utan bara att de kan minska generaliseringsgapet.

(3 poäng per deluppgift, plus 1 poäng om åtminstone en är helt korrekt.)

Lösning:

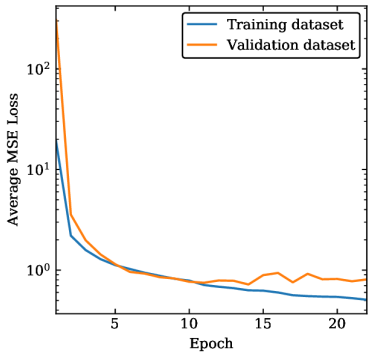
(a) Vi introducerar felfunktionen för tränings- respektive valideringsdata

$$E_{\text{train}} = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} E(\hat{y}(x_i), y_i), \quad \text{för } (x_i, y_i) \in \mathcal{D}_{\text{train}},$$

$$E_{\text{val}} = \frac{1}{N_{\text{val}}} \sum_{i=1}^{N_{\text{val}}} E(\hat{y}(x_i), y_i), \quad \text{för } (x_i, y_i) \in \mathcal{D}_{\text{val}}.$$

Vid slutet av varje träningsiteration (eller varje epok) beräknas dessa mått för den aktuella versionen av modellen. Notera alltså att de inte används för gradientstegsoptimering (det är kostnadsfunktionen som används till det).

(b) Se figur nedan. Skillnaden mellan de heldragna kurvorna är generaliseringsgapet.



Om felfunktionen är densamma som kostnadsfunktionen så kommer E_{train} att minska som en funktion av antalet iterationer. Även E_{val} kan förväntas att minska till en början, men när man närmar sig överoptimering så avstannar minskningen. Detta kan användas som en signal att upphöra med träningen.

(c) (1) Minska modellkomplexiteten (vilket ger mindre risk för överanpassning); (2) Viktregularisering; (3) Avbryt träningen när E_{val} inte längre minskar; (4) Använd mer träningsdata.

---

3. Betrakta en statistisk modell $\mathcal{D} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$, som relaterar N_d observationer i kolumnvektorn $\mathcal{D}$ till en linjär modell som beskrivs av designmatrisen $\boldsymbol{X}$ och parametervektorn $\boldsymbol{\theta}$ samt slumpvariabler $\boldsymbol{\epsilon}$. De sistnämnda beskriver osäkerheter och vi kan anta att elementen är oberoende och identiskt normalfördelade $p(\epsilon_i|I) = \mathcal{N}(0, \sigma_\epsilon^2)$.

(a) Visa att trolighetsfunktionen blir

$$p\left(\mathcal{D}|\boldsymbol{\theta}, I\right) = \left(\frac{1}{2\pi\sigma_\epsilon^2}\right)^{N_d/2} \exp\left[-\frac{1}{2}\frac{(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})^T(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})}{\sigma_\epsilon^2}\right].$$

(b) Visa därefter att

$$p\left(\mathcal{D}|\boldsymbol{\theta}, I\right) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^T\Sigma_\theta^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\right],$$

där $\boldsymbol{\theta}^* = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\mathcal{D}$ är lösningen till normalekvationen och $\Sigma_\theta^{-1} = \boldsymbol{X}^T\boldsymbol{X}/\sigma_\epsilon^2$.

Ledtråd: Definiera $L(\boldsymbol{\theta}) = (\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})^T(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta})/(2\sigma_\epsilon^2)$ och notera att gradientvektorn ges av $\boldsymbol{\nabla}_{\boldsymbol{\theta}} L = \boldsymbol{X}^T\left(\mathcal{D} - \boldsymbol{X}\boldsymbol{\theta}\right)/\sigma_\epsilon^2$ samt att Hessianmatrisen med andraderivator är $\boldsymbol{H} = \boldsymbol{X}^T\boldsymbol{X}/\sigma_\epsilon^2$.

(10 poäng)

Lösning:

Detta är bevis 3 från bevislistan (se kurshemsida och hänvisning till kompendiet).

---

4. I Göteborg regnar det 60% av alla dagar. Metereologerna på SMHI försöker förutsäga om det kommer regna imorgon eller inte. Historiskt

sett har de korrekt förutspått vädret 80% av alla regniga dagar, och 60% av alla dagar utan regn.

Givet att prognosen säger regn imorgon, vad är sannolikheten att det faktiskt kommer regna imorgon? *(10 poäng)*

Lösning:

- Introducera propositionen R=det regnar imorgon; P=prognosen säger att det skall regna imorgon.
- Enligt uppgiften är det känt att $\mathbb{P}(R \,|\, I) = 6/10$ samt att $\mathbb{P}\left(\bar{P} \,|\, \bar{R}, I\right) = 6/10$ och $\mathbb{P}(P \,|\, R, I) = 8/10$.
- Vi söker $\mathbb{P}(R \,|\, P, I)$ vilken vi tecknar med Bayes formel

$$\mathbb{P}(R \,|\, P, I) = \frac{\mathbb{P}(P \,|\, R, I)\,\mathbb{P}(R \,|\, I)}{\mathbb{P}(P \,|\, I)}.$$

- Vi behöver sannolikheten i nämnaren och utnyttjar först marginalisering och därefter produktregeln

$$\begin{aligned}\mathbb{P}(P \,|\, I) &= \mathbb{P}(P, R \,|\, I) + \mathbb{P}\left(P, \bar{R} \,|\, I\right) \\ &= \mathbb{P}(P \,|\, R, I)\,\mathbb{P}(R \,|\, I) + \mathbb{P}\left(P \,|\, \bar{R}, I\right)\mathbb{P}\left(\bar{R} \,|\, I\right).\end{aligned}$$

  Normaliseringen ger att $\mathbb{P}\left(\bar{R} \,|\, I\right) = 1 - \mathbb{P}(R \,|\, I) = 4/10$ samt $\mathbb{P}\left(P \,|\, \bar{R}, I\right) = 1 - \mathbb{P}\left(\bar{P} \,|\, \bar{R}, I\right) = 4/10$.
- Insättning av alla betingade sannolikheter ger

$$\mathbb{P}(R \,|\, P, I) = \frac{48/100}{64/100} = \frac{3}{4} = 0.75$$

---

5. Betrakta övergångsmatrisen

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix}$$

som genererar en icke-periodisk, positivt återkommande och icke-reducerbar Markovkedja med tre utfall (som betecknas $1, 2, 3$). Den första slumpvariabeln, X_0, i Markovkedjan beskrivs av sannolikhetsfördelningen $\boldsymbol{\pi}_0 = (0.5, 0.2, 0.3)$, dvs $\mathbb{P}(X_0 = 1) = 0.5$, osv.

(a) Vad är sannolikheten $\mathbb{P}(X_1 = 1)$?

(b) Vad är sannolikheten $\mathbb{P}(X_1 = 1 \,|\, X_0 = 1)$?

(c) Vad är sannolikheten $\mathbb{P}(X_n = 1)$, där n är ett väldigt stort tal?

(d) Vad är sannolikheten $\mathbb{P}(X_n = 1 \,|\, X_{n-1} = 1)$, där n är ett väldigt stort tal?

(e) Är Markovkedjan tidsreversibel? (Motivera ditt svar.)

(2 poäng per deluppgift.)

Lösning:

(a) Sannolikhetsfördelningen $\boldsymbol{\pi}_1$ för X_1 fås från $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_0 T$ vilket ger $\boldsymbol{\pi}_1 = (0.18, 0.64, 0.18)$. Svaret är därför $\mathbb{P}(X_1 = 1) = 0.18$.

(b) $\mathbb{P}(X_1 = 1 \,|\, X_0 = 1) = 0$, fås direkt från det översta vänstra elementet i övergångsmatrisen.

(c) Eftersom Markovkedjan är icke-periodisk, positivt återkommande och icke-reducerbar så har den en gränsdistribution, $\boldsymbol{\pi}$, vilken också är den stationära fördelningen. Vi behöver alltså lösa vänstereegenvärdesekvationen $\boldsymbol{\pi} = \boldsymbol{\pi} T$ med normaliseringsvillkoret $\sum_i \pi_i = 1$. Detta kan tecknas som ett ekvationssystem, eller genom att transponera till den "vanliga" höger-ekvationen $T^t \boldsymbol{\pi}^t = \boldsymbol{\pi}^t$. Man finner att $\boldsymbol{\pi} = \frac{1}{122}(27, 50, 45)$. Svaret är därför $\mathbb{P}(X_n = 1) = 27/122 \approx 0.22$.

(d) Kedjan är stationär eftersom den beskrivs av en övergångsmatris. Detta betyder att övergångssannolikheterna är samma vid alla tidssteg. Alltså, $\mathbb{P}(X_n = 1 \,|\, X_{n-1} = 1) = 0$, fås direkt från det översta vänstra elementet i övergångsmatrisen.

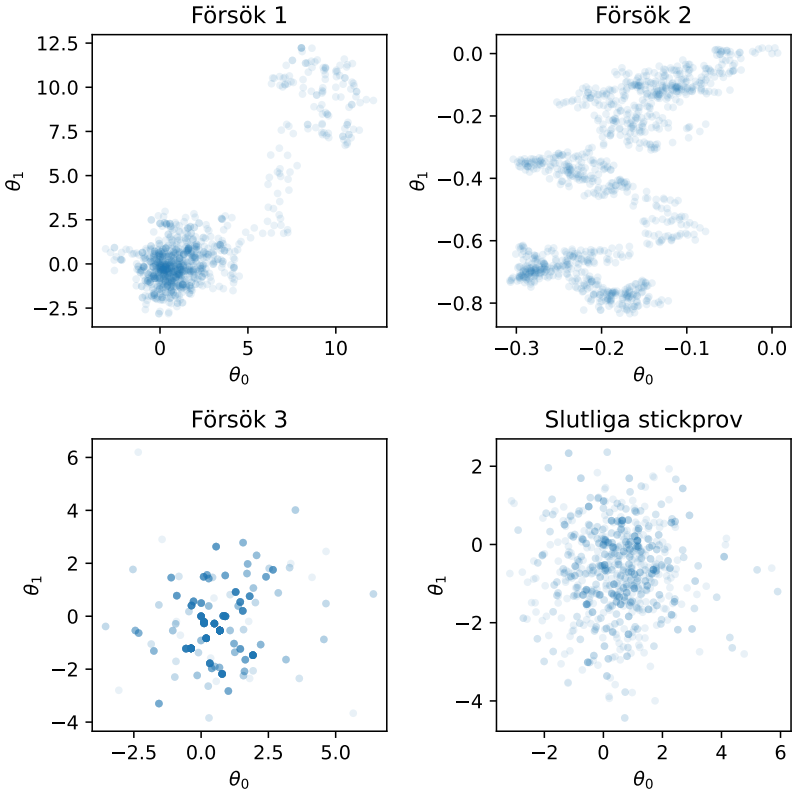
(e) Nej, denna Markovkedja är inte tidsreversibel eftersom den inte uppfyller detaljerad balans. Enkelt att visa explicit för t.ex. $i = 1, j = 2$ att $\pi_i T(i,j) \neq \pi_j T(j,i)$.

---

6. En ingenjör har som uppgift att leverera 1.000 stickprov från en sannolikhetstäthetsfunktion $p(\theta_0, \theta_1)$ med Metropolissampling och normalfördelade stegförslag.

   Figuren nedan visar flera försök att genomföra samplingen. Respektive panel visar ett spridningsdiagram ('scatter plot') med 1.000 stickprov (mörkare punkter innebär att flera stickprov ligger ovanpå varandra).

Ingenjören valde slutligen resultaten som visas i den sista panelen. Ge rimliga förslag till vad som kunde förbättras i de tre första försöken (det finns en huvudproblematik per försök; ge därför ett förslag per panel 1–3).



(3 poäng per panel [försök 1–3], plus 1 poäng om samtliga är korrekta.)

Lösning:

Försök 1 Panelen uppvisar en lång sekvens av stickprov som sträcker sig från det övre högra hörnet ner till det intressanta området (som här ligger i nedre vänstra hörnet). Denna transportsträcka visar att kedjan tagit tid på sig att konvergera till sin gränsfördelning. Eftersom uppgiften var att samla 1.000 stickprov från gränsfördelningen så kan det vara bättre att kasta bort resultaten från de första iterationerna och börja spara stickprov först efter denna 'burn-in' period.

Försök 2 Stickproven ligger väldigt nära varandra i ett litet område. Nästan

alla föreslagna positioner verkar ha accepterats (det är väldigt få repeterade punkter). Detta tyder på att man har använt en för liten steglängd i stegförslagsfunktionen. En jämförelse med förslag 4 avslöjar vi att vi inte har utforskat hela det relevanta området. Steglängden bör ökas. Ett alternativ kan vara att köra kedjan betydligt längre och bara spara stickprov från iterationer med långa mellanrum.

Försök 3 Här ser vi tvärtom väldigt få accepterade stegförslag. Det är snarare några få punkter som har repeterats många gånger. Detta tyder på att vi har en för stor steglängd och att vi ofta kliver bort från det intressanta området med resultatet att den nya punkten inte accepteras. En jämförelse med förslag 4 avslöjar vi att vi rör oss över det relevanta området, men samplingen är inte effektiv. Steglängden bör minskas.
