MIT-3030 Mittausdatan analyysi 2

Tentti 27.11.2007

Toimintaohje ja arvosteluperusteet:

- vastaa kaikkiin viiteen tehtävään
- näistä viidestä huonoiten pisteitä tuottanut jätetään arvostelussa huomiotta
- max 5 pistettä per tehtävä (osioiden pisteytys ilmoitettu tehtävissä)
- siis tentissä mahdollisuus saada max 20 pistettä
- yhteispisteet kurssisuorituksesta: tentti (max 20) + projektityö (max 10) = 30

Tehtävä 1. Selitä seuraavat käsitteet (max 20 sanaa ja yksi matemaattinen lauseke per kohta; 1 p per kohta)

a) Maksimientropiajakauma
b) Normaalijakautuneen satunnaisvektorin pääkomponentit
c) Satunnaisvektorin X regressio satunnaisvektorin Y suhteen
d) Informaatio
e) Momentit generoiva funktio

Tehtävä 2. Systeemiä kuvaa lineaarinen tilamalli

$$
\begin{aligned}
x_{n+1} &= F_n x_n + G_n u_n + \varepsilon_{n+1} \\
\varepsilon_{n+1} &\sim N(0, \sigma_1^2 Id) \quad i.i.d \\
y_{n+1} &= C_{n+1} x_{n+1} + \nu_{n+1} \\
\nu_{n+1} &\sim N(0, \sigma_2^2 Id) \quad i.i.d
\end{aligned}
$$

Tilaestimoinnin tehtävänä on muodostaa mittaustulosten $y_1,\ldots y_{n+1}$ ja tunnettujen ulkoisten tekijöiden $u_1,\ldots u_n$ perusteella tilan x_{n+1} todennäköisyysjakauma. Tällainen estimaatti voidaan muodostaa rekursiivisesti Kalman-suodatukseksi kutsutulla menetelmällä.

a) Kuvaa Kalman-suodatuksen periaate. Matemaattisia lausekkeita ei tarvitse muodostaa. (4 p)
b) Miten periaate muuttuu, jos tila, ulkoinen tekijä ja mittaus ovatkin binäärejä skalaareja ja malli kuvattu muodossa (1 p)

$$
P(x_{n+1} \mid x_n, u_n = 0) = \begin{bmatrix} 1-p_0 & q_0 \\ p_0 & 1-q_0 \end{bmatrix}
$$

$$
P(x_{n+1} \mid x_n, u_n = 1) = \begin{bmatrix} 1-p_1 & q_1 \\ p_1 & 1-q_1 \end{bmatrix}
$$

$$
P(y_{n+1} \mid x_{n+1}) = \begin{bmatrix} 1-\tilde{p}_1 & \tilde{q}_1 \\ \tilde{p}_1 & 1-\tilde{q}_1 \end{bmatrix}
$$

Tehtävä 3. Olet laatimassa prosessioperaattorin päätöksenteon tueksi systeemimallia, jolla tietyn input-informaation perustella ennustetaan tuotteen keskeistä laatuominaisuutta. Käytettävissäsi on (esim) 40 input-muuttujaa ja näistä sekä ennustettavasta laatuominaisuudesta (esim) 1000 mittausarvon (per muuttuja) historiadata. Tehtäväsi on arvioida myös ennusteen epävarmuutta. Kuvaa epävarmuuden lähteet sekä ne periaatteet, joiden mukaan arvioit ja käsittelet niitä a) mallia identifioitaessa, b) mallia sovellettaessa.

Tehtävä 4. Erästä laitetta käytetään vaihtelevissa ympäristöolosuhteissa niin, että tietty laiteyksilö on koko elinikänsä samassa ympäristössä. Oletetaan, että ympäristön rasittavuutta voidaan luonnehtia suureella x. Oletetaan edelleen, että laitteen elinikä T on eksponenttijakautunut satunnaissuure:

$$f_{T|x}(t \mid x, \alpha_0, \alpha_1) = (\alpha_0 + \alpha_1 x) \cdot \exp(-(\alpha_0 + \alpha_1 x)t)$$

a) Olkoon meillä havaintojoukko rasittavuustekijästä ja laitteen eliniästä ja oletetaan, että nämä havaintoparit ovat toisistaan riippumattomia, $\{(x_i,t_i)\}$ i=1,..N. Johda maximum likelihood –yhtälöt parametrien α_0 ja α_1 estimaateille. (1,5 p) [HUOM! Yhtälöt eivät ratkea analyyttisesti; kohdissa b ja c ratkaisuista käytetään merkintää $\hat{\alpha}_0, \hat{\alpha}_1$].

b) Approksimoidaan parametrien α_0 ja α_1 jakaumaa normaalijakaumalla. Mikä on parametrien jakauman kovarianssimatriisi C_α lausuttuna havaintojen $\{x_i,t_i\}$ i=1,..N ja maximum likelihood –estimaattien $\hat{\alpha}_0, \hat{\alpha}_1$ avulla. Perustele lausekkeiden avulla, miksi kovarianssimatriisin arvo on keskimääräisesti tarkastellen kääntäen verrannollinen havaintojen lukumäärään N. (2 p)

c) Olemme asentamassa laitetta ympäristöön, jonka rasittavuusarvoksi olemme määrittäneet x_{as}. Ottaen huomioon parametrien epävarmuus, kirjoita lauseke laitteen eliniän todennäköisyysjakaumalle ko asennuksessa. Toisin sanoen: mikä on T_{as}:n jakauma ehdoilla, että a) T:n "todellinen" jakauma kuuluu eksponenttijakaumien perheeseen; b) että meillä on tiedossamme havaintojen historia $\{x_i,t_i\}$ i=1,..N, ja c) asennukseen liittyy rasittavuusarvo x_{as}:

$$f_{T|x_{as}}\left(t_{as} \mid M \in Exp(\alpha_0 + \alpha_1 x), \{x_i, t_i\}_{i=1}^{N}, x_{as}\right)$$

Ilmoita tämä todennäköisyysjakauma integraalimuodossa parametrijakauman avulla. Integraalia ei tarvitse laskea, tulos on kohdassa d. (1 p)

d) Kun tuon integraalin laskee analyyttisesti, tulos on

$$(\alpha_0 + \alpha_1 x - t\beta(x)/N) \cdot \exp[-(\alpha_0 + \alpha_1 x - t\beta(x)/2N) \cdot t]$$

$$\beta(x) = N \cdot [1 \quad x] C_\alpha \begin{bmatrix} 1 \\ x \end{bmatrix}$$

Tämä on kummallinen tulos, koska jakauma suurilla t:n arvoilla on saa negatiivisia arvoja. Miksi näin käy? Vinkki: α_0:n ja α_1:n priori? (0.5 p)

Tehtävä 5. Olemme historiatiedon perusteella osoittaneet, että paperinvalmistuksen sekoitussäiliön ulosvirtauksessa massan kuitusakeuden vaihtelu (poikkeamana keskiarvosta $x^{(mean)}$) noudattaa AR(2) –mallia:

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \varepsilon_n$$
$$\varepsilon_n \sim N(0, \sigma^2)$$

Kuten kurssilla osoitettiin, yleisesti tiedetään normaalijakaumalle pätevän, että jos

$$X \sim N_d(\mu, \Sigma)$$

(N_d viittaa d-dimensíoiseen normaalijakaumaan) ja Y = BX, missä B on d'*d –matriisi ja Y siis d'-dimensioinen satunnaisvektori, niin

$$Y \sim N_{d'}(B\mu, B\Sigma B^T)$$

a) Oletetaan, että x_0 ja x_{-1} tunnetaan. Osoita, että x:n arvoista välillä 1..n muodostettu satunnaisvektori noudattaa jakaumaa:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ .. \\ x_n \end{bmatrix} \sim N_n(m, C)$$

missä

$$m = A^{-1}\begin{bmatrix} a_1x_0 + a_2x_{-1} \\ a_2x_0 \\ 0 \\ .. \\ 0 \end{bmatrix}$$

$$C = \sigma^2 (A^{-1})^T A^{-1}$$

$$A = \begin{bmatrix} 1 & 0 & 0 & .. & 0 & 0 & 0 \\ -a_1 & 1 & 0 & .. & 0 & 0 & 0 \\ -a_2 & -a_1 & 1 & .. & 0 & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & 1 & 0 & 0 \\ 0 & 0 & 0 & .. & -a_1 & 1 & 0 \\ 0 & 0 & 0 & .. & -a_2 & -a_1 & 1 \end{bmatrix} \text{ (2,5 p)}$$

b) Seuraavan n:n aika-askeleen aikana ulosvirtauksen volumetrista määrää aiotaan ohjata sekvenssin f_1, .. f_n mukaisesti. Luonnollisesti kuitumäärä kullakin ajanhetkellä = volumetrinen virtaus * sakeus. Johda lausekkeet koko tarkastelujakson (summa n:n aika-askeleen yli) aikaisen kuitumäärän odotusarvolle ja hajonnalle. (Vinkki: tässäkin kohdassa on apua tehtävän alussa mainitusta normaalijakauman ominaisuudesta) (2 p)

c) Minkä rajojen sisässä kuitumäärä ko jaksolla on todennäköisyydellä 0.9974 (0.5 p)?