

Bayesiläisen mallintamisen perusteet

Slide 1

- Johdanto
- Yksiparametrisia malleja
- Moniparametrisia malleja
- Päätely suurten otosten tapauksessa ja bayesiläisen päättelyn frekvenssiominaisuudet
- Hierarkiset mallit
- Laskennallisia menetelmiä, Markov-ketju Monte Carlo
- Päätösanalyysi
- Mallien tarkistus, vertailu ja parannus
- Datankeruuprosessin mallintaminen

Johdanto (luku 1)

Slide 2

- Termit ja notaatio
- Bayesin kaava, summa- ja tulosääntö
- Predikttiivinen jakauma
- Todennäköisyys epävarmuuden mittana
- Subjektiiivisuus vs. objektiivisuus
- Yksinkertainen simulaatio
- *Inverse-cdf*-menetelmä

Aksiomaattiset perustelut todennäköisyydelle*

- Todennäköisyyskalkyyli voidaan perustella aksiomaattisesti
 - useita variaatioita, joissa samat perusedat, mutta hieman esitystavassa eroa
 - kaksi peruslinjaa
 - todennäköisyys ja hyöty erikseen (esim. Cox)
 - todennäköisyys ja hyöty erottamattomia (esim. Bernardo & Smith)
 - Bernardo & Smith:in esitys olisi mukava esittää, mutta pitkä

Slide 3

Päätelyn vaatimukset (Cox 1946)*

1. Uskottavuuden asteet esitetään reaalityyppisillä
2. Kvalitatiivisesti "järkevä"
 - a) jos $(H|I') > (H|I)$ niin $(\bar{H}|I') < (\bar{H}|I)$
 - b) jos lisäksi $(D|HI') = (D|HI)$ niin $(HD|I') \geq (HD|I)$
3. Johdonmukaisuus
 - a) sisäinen johdonmukaisuus
 - b) asianmukaisuus
 - c) yhdenmukaisuus

Slide 4

→ nämä ehdot toteuttava on ekvivalentti bayeislaisen kalkyylin kanssa

Yksiparametrisia malleja (luku 2)

- Termit ja notaatio
- Binomi-jakauma
 - minkäläiselle datalle, kaava ja parametrit, konjugaattipriori
- Normaalijakauma
 - minkäläiselle datalle, kaava ja parametrit, konjugaattipriorit
- (ei tarvitse tentissä osata johtaa posteriorijakaumia tms.)
- Posteriorijakauman esittäminen
 - odotusarvo, hajonta, kvantiilit, intervallit, ja hpd
- Konjugaattipriorin vs. ei-konjugaattipriori
 - kummankin edut ja ongelmat, esim. vaikutus laskentaan
- Informatiivinen priorin vs. ei-informatiivinen priorin
 - kummankin edut ja ongelmat

Slide 5

Moniparametrisia malleja (luku 3)

- Termit ja notaatio
- Marginaalijakauma ja marginalisaatio
- Normaalijakauma
 - ei-informatiivinen priorin ja sen ehdolliset jakaumat ja marginaalijakaumat
 - konjugaattipriorin
- (ei tarvitse tentissä osata johtaa posteriorijakaumia tms.)
- Semi-konjugaattisuus
- Multinomi-malli
 - milläiselle datalle, konjugaattipriorin

Slide 6

Päätely suurten otosten tapauksessa (luku 4)

Slide 7

- Normaalijakauma-approksimaatio
 - Taylor-sarjakehitelmä log-posteriorille
 - approksimaation parametrit
 - havaittu informaatio
 - kuinka voidaan laskea
 - edut ja ongelmat
 - asymptootisuuden merkitys
- Termit lyhyesti
 - ali-identifioituva malli ja ei-identifioituvat parametrit
 - valetointo
 - rajoittamaton likelihood
 - ei-aito posteriori
- Frekvenssiominaisuuksia ei kysytä tentissä

Hierarkiset mallit (luku 5)

Slide 8

- Hierarkinen malli
 - määrittely
 - millaiselle datalle
 - edut
 - laskenta ja poiminta yleisellä tasolla ja eri vaihtoehdot (ei yksityiskohtaisia kaavoja)
- Vaihtokelpoisuus
 - määrittelmä
 - esimerkkejä
 - suhde riippumattomuuteen
 - vaihtokelpoisuus kun yksiköistä on lisäinformaatiota
- Erillis- ja yhteismalli ja niiden yhteys hierarkiseen malliin

Laskennallisia menetelmiä (luku 10)

- Karkean estimoinnin merkitys
- Montako simulaationäytettä tarvitaan
 - simulaatioepävarmuus (*Monte carlo error*)
 - odotusarvolle
 - posterioritodennäköisyydelle

Slide 9

Laskennallisia menetelmiä (luku 11)

- Lyhyet kuvaukset seuraavista sekä edut ja ongelmat
 - suora simulointi
 - hilapointi
 - hylkäyspointi

Slide 10

Markov-ketju Monte Carlo (luku 11)

- Markov-ketju Monte Carlo
 - määritelmä
 - termit: siirtymäjakauma, alkupiste, stationaarinen jakauma, konvergenssi
 - edut ja ongelmat
- Gibbs-poiminta, Metropolis ja Metropolis-Hastings-algoritmit
 - kuvaus pseudokooditasolla
 - edut ja ongelmat
 - Gibbs-poiminta jos osa ehdollisista jakaumista ei suljetussa muodossa

Slide 11

Markov-ketju Monte Carlo (luku 11)

- Sisäänajo
- Peräkkäisten näytteiden riippuvuus
 - termi efektiivinen näytteiden määrä
- Konvergenssidiagnostiikka
 - määritelmä
 - useat ketjut
 - PSRF (sanallisesti kuvattuna riittää)

Slide 12

Päätösanalyysi (luku 22)

- Päätösanalyysin termit ja vaiheet
 - päätös
 - lopputulos ja sen jakauma annettuna päätös
 - hyöty- tai kustannusfunktio
 - hyödyn jakauma
 - odotettu hyöty

Slide 13

- Päätösanalyysilasku
 - tasoltaan vastaava kuin luku 22.3, tehtävä 22.1 tai luentoexamplesimerkki (testaa samalla Bayesin kaavan käytön)

Mallien tarkistus, vertailu ja parannus (luku 6)

- Onko mallin tuloksissa järkeä?
- Ulkoinen validointi
- Posterioriprediktiivinen tarkistus
 - posterioriorediiktiviset replikaatit ja p -arvot
 - edut ja ongelmat

Slide 14

- Herkkyysanalyysi
- DIC
 - määritelmä ja kuinka lasketaan
 - efektiivinen parametrien määrä
 - edut ja ongelmat
- Bayes factor
 - määritelmä ja ongelmat

Datankeruuoprosessin mallintaminen (luku 7)

Slide 15

- Perustermien määrittelyt lyhyesti
 - havaittu, puuttuva, koko data
 - stabiilisuus
 - super- ja äärellisen populaation päättely
 - *ignorability*
 - satunnaistus, deterministen koesuunnitelma
 - otantatutkimus, staunnaistettu otanta, ositettu otanta, ryväsotanta
 - suunnitellut kokeet, täysin satunnaistetut, deterministiset
 - koesuunnitelmaan herkkyys
 - havaintotutkimukset
 - sensurointi, katkaisu

Tentti

Slide 16

- Termit, menetelmät ja algoritmit kuvauksia
 - lyhyet tai pidemmät kuvaukset
 - aina mukaan edut ja ongelmat
 - eri termien tai menetelmien välisiä suhteita ja eroja
 - algoritmit pseudokoodilla
 - tärkeimmistä asioista myös kaavat
 - kysymykset esim. muotoa
 - Kuvaile lyhyesti vaihtoehtoisia tapoja <tehdä jotakin>
 - Kerro lyhyesti mitä seuraavat termit tarkoittavat ja mihin niitä käytetään
 - Kuvaa lyhyesti <z>-algoritmi/menetelmä ja kerro sen edut ja ongelmat
 - Kuvaa lyhyesti algoritmit/menetelmät/termit <x> ja <y> ja vertaile niiden ominaisuuksia
 - Mitä on <x> ja mihin sitä käytetään?
- Yksi laskutehtävä

Mahdollisia jatkokurssin aiheita

- Priorit
 - ei-informatiiviset priorit
 - referenssipriorit
 - maksimientropiapriorit
 - herkkyysanalyysi
 - "prior elicitation" eli priori-informaation hankkiminen asiantuntijoilta

Slide 17

Mahdollisia jatkokurssin aiheita

- Monte Carlo
 - teoriaa
 - perfect sampling
 - hybrid Monte Carlo
 - population Monte Carlo
 - slice sampling
 - adaptiiviset
 - mahdollisuus vähentää satunnaiskävelyä
 - importance sampling
 - tempering-menetelmät
 - trans-dimensionaaliset
 - konevergenssidiagnostikat
- Variaatiolaskennan menetelmät

Slide 18

Mahdollisia jatkokurssin aiheita

- Päätösanalyysi
- Mallin valinta
 - Bayes-factor ja evidenssitermin laskenta
 - trans-dimensionaalinen MCMC
 - ristiinvalidointi
 - DIC ja muut informaatiokriteerit

Slide 19

Mahdollisia jatkokurssin aiheita

- Regressiomallit
 - lineaarimalli
 - hierarkinen lineaarimalli
 - yleistetty lineaarimalli
 - robustit mallit
 - sekamallit (mixture models)
 - puuttuvan datan mallit
 - faktorianalyysi
 - *large p, small n*
 - epälineaariset mallit
 - MLP-neuroverkot
 - gaussiset prosessit
 - Bayesian support vector machine
 - jne.

Slide 20

Jatkokurssin aiheita

- Dynaamiset mallit
 - aikasarjamallinnus
 - *hidden Markov models*
 - Bayesian tracking and filtering
 - Kalman filter
 - extended Kalman filter
 - unscented Kalman filter
 - sequential Monte Carlo
 - particle filter
 - jne.

Slide 21

Jatkokurssin aiheita

- Tiheystestimointi
- Bayes-verkot
 - binaariset
 - diskreetit
 - gaussiset
 - dynaamiset
- Spatiaalimallit
 - *conditional auto regressive* (CAR)
 - *Markov random field* (MRF)
- *Survival models*

Slide 22

Jatkokurssin aiheita

- Sovellukset
 - kuva-analyysi, konenäkö
 - bioinformatiikka, lääketiede, dna-analyysi
 - *econometrics, econophysics*
 - fysiikkaa
 - jne.

Slide 23