

Luonnollisen kielen tilastollinen käsittely, T-61.281

Tenttiinlukuohje

Yleisesti ottaen tentissä edellytetään kurssikirjan sisällön tuntemista siinä laajuudessa kuin mitä luennoilla on esitetty ja soveltamisosaamista sillä tasolla kuin mitä laskuharjoituksissa on käyty läpi. Kurssikirjaa lukemalla voi parantaa yleisnäkemystään. Tentissä edellytetään joidenkin matemaattisten perusteiden muistamista varsinkin luvussa 2 esitettyjä asioita sekä joitakin seuraavassa erikseen mainittuja kuten Bayesin kaava sekä saannin ja tarkkuuden määritelmät.

Kirjan alkuosa on taustaa ja menetelmäjohtantoa. Perusasiat olisi hyvä osata, menetelmällisenä taustana varsinkin luku 2.

Luku 5. Vähänkään spesifimmät kaavat annetaan tentissä. Esimerkiksi keskiarvo ja varianssi tulee toki osata. Esimerkiksi t-testiä jne. pitää osata soveltaa, mutta kaavat annetaan.

Luku 6. Ehdollisen todennäköisyyden käsite ja ML-estimaatti osattava ulkoa. Esimerkiksi luvussa 6.2.2. esitettyjä kaavoja ei tarvitse osata ulkoa. Käsitteellisesti on hyvä ymmärtää testidatan käyttö ja ristiinvalidointi.

Luku 7. Perusasiat: Bayesin kaava osattava ulkoa ja osattava soveltaa. Luvussa 7.4 EM-algoritmin ymmärtäminen on keskeistä, mutta kaavoja ei tarvitse muistaa ulkoa.

Luku 8. Tarkkuus ja saanti tärkeät käsitteet laskukaavoineen. Muita esitettyjä asioita ei tarvitse muistaa kaavatasolla. Soveltamisosaaminen olennaista.

Luku 9. Kielioppimallit: Markov-mallien perusoletukset ja perustn:iien laskenta on keskeistä, HMM:iien peruskäsitteet. Ymmärrys algoritmien soveltamisesta (Viterbi, Backward-Forward) on tärkeää mutta kaavoja ei tarvitse muistaa ulkoa.

Luku 10. Sanaluokkien taggaus soveltamistasolla.

Luku 11. Yhteysvapaat kieliopit ja niiden tn-versio, kaavatasolla on osattava, miten lasketaan todennäköisyys merkkijonolle, muuten sovellettavat menetelmät ja kaavat annetaan. Osattava soveltaa samalla tasolla kuin mitä laskuharjoituksissa on käsitelty

Luku 12. Probabilistinen jäsentäminen, samalla tasolla kuin luku 11.

Luku 13. Käsitteistö luentojen tasolla, soveltamisosaaminen.

Luku 14. Menetelmät: perusmenetelmien ominaisuudet (käsitteistö) ja soveltaminen (esim. menetelmien keskeisyys), tärkeä: k-means.

Luku 15 Tiedonhaku (jälleen tarkkuus, saanti), vektoriavaruusmalli: perusasiat (kosinimitta [osattava kaavatasolla], termien painotus), tärkeä: TF/(I)DF tunnettava.