Datasta Tietoon, syksy 2011

HARJOITUSTEHTÄVÄT 2 [pe 11.11.2011, ma 14.11.2011]

H2 / 1. (Pääkomponenttianalyysi)

On annettuna seuraava datamatriisi $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 2 & 5 & 6 & 7 \\ 1 & 3 & 5 & 7 \end{bmatrix}$$

a) Piirrä $\mathbf{X}$:n sarakkeet (x_1, x_2) - koordinaatistoon
b) Keskiarvoista $\mathbf{X}$ vähentämällä sarakkeista niiden keskiarvovektori
c) Muodosta kovarianssimatriisi $\mathbf{C}$ ja laske sen suurinta ominaisarvoa vastaava ominaisvektori. Piirrä sen suunta kohdan a) kuvaan. Miten tuloksia voi tulkita pääkomponenttianalyysin mukaisesti?

H2 / 2. (Pääkomponenttianalyysi)

Olkoon $\mathbf{x}$ nollakeskiarvoinen satunnaisvektori, josta on olemassa otos $\mathbf{x}(1), ..., \mathbf{x}(n)$. Olkoon $\mathbf{w}$ yksikkövektori (siis $\|\mathbf{w}\| = 1$) ja $y = \mathbf{w}^T\mathbf{x}$. Halutaan maksimoida y:n varianssi $E\{y^2\} = E\{(\mathbf{w}^T\mathbf{x})^2\}$. Osoita että se maksimoituu, kun $\mathbf{w}$ on matriisin $E\{\mathbf{x}\mathbf{x}^T\}$ suurinta ominaisarvoa vastaava ominaisvektori.

H2 / 3. (ML-estimointi)

Laske suurimman uskottavuuden estimaatti eksponentiaalijakauman

$$p(x|\lambda) = \lambda e^{-\lambda x}$$

parametrille λ kun suureesta x on olemassa otos $x(1), ..., x(n)$.

H2 / 4. (Bayes-estimointi)

On annettu otos $x(1), ..., x(n)$ suureesta, jonka tiedetään olevan normaalijakautunut

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

On syytä olettaa että keskiarvo μ on lähellä nollaa. Koodataan tämä olettamus priorijakaumaan

$$p(\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2}.$$

Laske Bayes-MAP-estimaatti odotusarvolle μ ja tulkitse sitä kun varianssi σ^2 vaihtelee pienestä suureen.

PISTETEHTÄVÄ P2 [DL ma 21.11.2011 klo 14.00]

Palvelukeskukseen saapuu keskimäärin λ puhelua minuutissa satunnaisina hetkinä. Voidaan osoittaa että tällöin todennäköisyys sille, että puheluja tulee minuutissa k kpl, noudattaa Poisson-jakaumaa:

$$P(\text{puheluja } k \text{ kpl}) = p(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1}$$

Mitataan puhelujen määrää n:n minuutin mittaisen ajanjakson aikana ja saapuvat määrät ovat $k_1, k_2, ..., k_n$. Laske suurimman uskottavuuden estimaatti odotusarvolle λ.