Datasta Tietoon, syksy 2007

HARJOITUSTEHTÄVÄT 3 [pe 23.11.2007, ke 28.11.2007]

H3 / 1. (MLE-regressio)

On annettu n mittausparia $(y(i), x(i)),\ i = 1, .., n$ joistakin muuttujista x, y joiden välillä arvellaan olevan lineaarinen yhteys: $y = \theta x$. Mittauksiin sisältyy kuitenkin virhettä: $y(i) = \theta x(i) + \epsilon(i)$ missä $\epsilon(i)$ on mittausvirhe ("kohina") i:nnessä pisteessä. Oletetaan että mittausvirhe $\epsilon(i)$ on normaalijakautunut keskiarvolla 0 ja keskihajonnalla σ.

Ratkaise kulmakerroin θ suurimman uskottavuuden estimoinnilla.

H3 / 2. (Bayes-regressio)

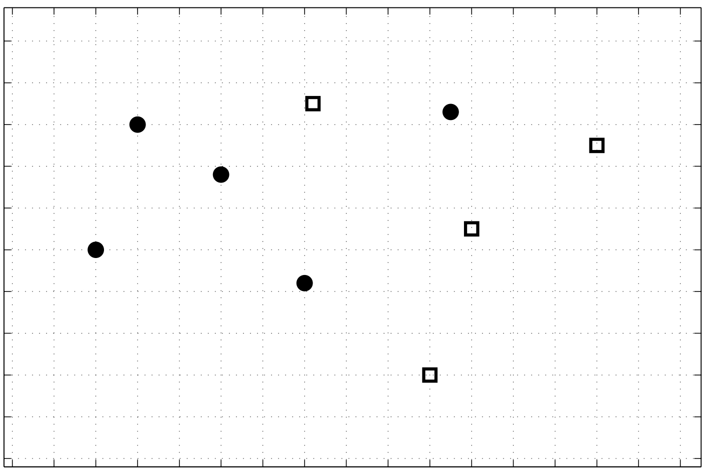
Lisätään edelliseen tehtävään etukäteistietoa:

1. Arvellaan, että kulmakerroin θ on suunnilleen 1. Mallitetaan tähän liittyvä epävarmuus olettamalla normaalinen priorijakauma jonka keskiarvo on 1 ja keskihajonta 0.5.

2. Arvellaan, että regressiosuoran ei ehkä kuitenkaan pitäisi kulkea origon kautta, jolloin se onkin muotoa $y = \alpha + \theta x$. Mittausten välinen yhteys on silloin $y(i) = \alpha + \theta x(i) + \epsilon(i)$. Mallitetaan uuteen parametriin α liittyvä epävarmuus olettamalla että sillä on normaalinen priorijakauma jonka keskiarvo on 0 ja keskihajonta 0.1. Laske Bayes-estimaatit parametreille α, θ.

H3 / 3. (Lähimmän naapurin luokitin)

Oheisessa kuvassa on 2 dimensiossa 2 luokkaa (ympyrät ja ruudut). Piirrä kuvaan Lähimmän naapurin luokittimen (1-NN -luokittimen) rajapinta luokkien välille.



H3 / 4. (Bayes-luokitin)

Oletetaan kaksi luokkaa skalaarimuuttujalle x. Luokkien tiheysfunktiot $p(x|\omega_1), p(x|\omega_2)$ ovat normaalijakautuneita siten että molempien keskiarvo on 0 mutta hajonnat σ_1, σ_2 ovat erisuuret. Prioritodennäköisyydet ovat $P(\omega_1), P(\omega_2)$. Piirrä tiheysfunktiot. Mihin laittaisit luokkarajat? Johda Bayes-luokittimen luokkarajat.

H3 / 5. (Bayes-luokitin binääridatalle)

Kuten luennolla esitettiin, tehdään Bayes-luokitin binäärivektoreille $\mathbf{x} = (x_1, ...x_d)^T$ missä kukin x_i on joko 0 tai 1. Meillä on kaksi luokkaa, joista luokassa ω_1 on ykkösen todennäköisyys p ja luokassa ω_2 se on $q < p$. Oletetaan vektorin alkiot toisistaan riippumattomiksi. Prioritodennäköisyydet ovat $P(\omega_1), P(\omega_2)$.

a) Merkitään vektorissa $\mathbf{x}$ olevien ykkösten lukumäärää N:llä. Lausu vektorin $\mathbf{x}$ todennäköisyys luokassa 1 ja luokassa 2 tämän avulla.

b) Muodosta Bayes-luokitin joka antaa päätösalueet suureelle N.