

8001652 Introduction to Pattern Recognition  
Tentti 5.4 2004

Huom! Jos et ole varma termien suomennoksista katso englanninkielinen versio paperin toiselta puolelta.

1. Selitä seuraavat kolme virhelähdettä kun käytetään maksimioskottavuus lähestymistapaa ohjattuun luokitteluun ja selitä myös kuinka näitä virheitä voitaisiin periaatteessa pienentää:
  - (a) Bayesin virhe (2 p)
  - (b) Mallivirhe (2 p)
  - (c) Estimointivirhe (2 p)
2. Ajatellaan  $c$ :n luokan ja yhden piirteen luokitteluongelmaa, jossa luokkien tiheysfunktiot ovat normaalijakautuneita samoilla variansseilla  $\sigma^2$ , mutta eri keskiarvoilla  $\mu_i$ , t.s.  $p(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2}(\frac{x-\mu_i}{\sigma^2})^2]$ .
  - (a) Esitä minimivirhepäätössääntö, joka siis minimoi luokitteluvirheen. (2 p)
  - (b) Johda minimivirhepäätössäännön kanssa ekvivalentti päätössääntö, joka hyödyntää vain lineaarisia erotinfunktioita (discriminant functions) (4 p).
3. Ajatellaan lineaarista konetta diskriminanttifunktioilla  $g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$ ,  $i = 1, \dots, c$ . Näytä, että päätösalueet ovat konvekseja näyttämällä että jos  $\mathbf{x}_1 \in \mathcal{R}_i$  ja  $\mathbf{x}_2 \in \mathcal{R}_i$ , niin  $\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 \in \mathcal{R}_i$  kaikille  $0 \leq \lambda \leq 1$ .
4. K-means klusterointi. Vastauksestasi tulisi käydä ilmi mihin K-means klusterointia käytetään, mitä kriteerifunktiota käytetään, ja mikä on itse algoritmi. Lyhyt kuvaus algoritmin/kriteerifunktion huonoista puolista vaaditaan myös kuuteen pisteeseen.