

8001652 Introduction to Pattern Recognition
Tentti 15.12 2003

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

1. Olkoot $\omega_1, \dots, \omega_c$ c mahdollista tilaa (luokkaa) ja $\alpha_1, \dots, \alpha_a$ a mahdollista toimenpidettä. Määrittele termit:
 - (a) tappiofunktio (1 p)
 - (b) päätössääntö (1 p)
 - (c) päätössäännön kokonaisriski (2 p)
 - (d) Bayesin riski (2 p)

2. Ajatellaan c :n luokan ja yhden piirteen luokitteluongelmaa, jossa luokkien tiheysfunktiot ovat normaalijakautuneita samoilla variansseilla σ^2 , mutta eri keskiarvoilla μ_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 erotusfunktioita (discriminant functions) (4 p).

3. Ajatellaan kolmen kategorian k :n lähimmän naapurin luokittelijaa (k -nearest neighbor classifier) ja seuraavia kaksiulotteisia luokiteltuja prototyyppjä

ω_1	(4,4)	(14,7)	(18,6)	(7,4)	(14,9)	(16,2)	(2,16)
ω_2	(-7,8)	(3,11)	(-2,-1)	(4,3)	(3,4)	(4,-3)	(0,6)
ω_3	(3,2)	(8,1)	(3,3)	(6,3)	(5,10)	(3,9)	(1,7)

Luokittele piste (4,4) perustuen

- (a) lähimmän naapurin sääntöön
- (b) 3-lähimmän naapurin sääntöön
- (c) 7-lähimmän naapurin sääntöön

Käytä etäisyyksien mittaamiseen Euklidista etäisyyttä (siis se normaali etäisyys). Oikeat vastaukset ilman selitystä kuinka ne ollaan saatu tuotavat nolla pistettä, joten selitä miten laskit vastaukset. (6 p)

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.