Oulun yliopisto
Matemaattisten tieteiden yksikkö
Tilastollinen hahmontunnistus
Harjoitus 5

1. Tarkastellaan yksiulotteista kahden luokan tapausta, jossa $P_1 = P_2 = 0.5$ ja

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{ja} \qquad f_2(x) = \frac{1}{\sqrt{\pi}} e^{-(x-1)^2}.$$

Johda Bayesin luokitin ja laske sen virhetodennäköisyys.

2. Tarkastellaan edelleen yksiulotteista kahden luokan tapausta ja olkoon nyt

$$f_j(x) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a_j}{b}\right)^2}, \qquad j = 1, 2,$$

missä $a_1, a_2, b \in \mathbb{R}$, $a_1 \neq a_2$ ja $b > 0$.

(a) Osoita, että funktiot f_j todella ovat tiheysfunktoita.

(b) Osoita, että jos $P_1 = P_2 = 0.5$, niin Bayesin luokittelijan päätöspinta (itseasiassa piste) sijaitsee funktioiden f_1 ja f_2 huippujen puolivälissä, riippumatta b:n arvosta.

(c) Hahmottele posterioritodennäköisyyden $P(1|x)$ kuvaaja kun $a_1 = 3$, $a_2 = 5$ ja $b = 1$.

(d) Kuinka posterioritodennäköisyydet $P(1|x)$ ja $P(2|x)$ käyttäytyvät, kun $x \to \infty$ tai $x \to -\infty$? Miten tulkitset tilanteen?

3. Tarkastellaan edellisen tehtävän tilannetta ja oletetaan, että $P_1 = P_2 = 0.5$.

(a) Osoita, että Bayesin virhetodennäköisyys on

$$\frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left| \frac{a_2 - a_1}{2b} \right|.$$

(b) Hahmottele Bayesin virhetodennäköisyyden kuvaaja suureen $|a_2 - a_1|/b$ funktiona. Tulkitse havaintosi.

(c) Mikä on Bayesin virheen on suurin arvo ja millaisessa tilanteessa se esiintyy?

4. Olkoon X d-ulotteinen jatkuva satunnaisvektori ja tarkastellaan pariin (X, J) liittyvää c:n luokan hahmontunnistusongelmaa. Olkoon $g_X^* : \mathbb{R}^d \to \{1, \ldots, c\}$ tähän ongelmaan liittyvä Bayesin luokitin. Olkoon sitten $h : \mathbb{R}^d \to \mathbb{R}^d$ bijektio ja

$Y = h(X)$ (eli siis $Y = h \circ X$). Oletetaan, että h on niin säännöllinen, että Y on myös jatkuva satunnaisvektori ja että parille (Y, J) voidaan määritellä Bayesin luokitin $g_Y^* : \mathbb{R}^d \to \{1, \ldots, c\}$ luennoilla esitettyyn tapaan (tarkkaan ottaen pitää lisäksi olettaa, että h^{-1} on Borelin funktio). Osoita, että näiden kahden Bayesin luokittimen virhetodennäköisyydet ovat täsmälleen samat. Esimerkkinä tällaisesta tilanteesta olisi vaikka säännöllinen lineaarikuvaus $h(X) = AX$, jota käyttämällä ei siis voida parantaa (eikä huonontaa) pienintä mahdollista virhetodennäköisyyttä.

5. Hahmovektorin $X = [X^{(1)}, \ldots, X^{(d)}]^T$ luokkatiheysfunktiot ovat

$$f_j(x) = \prod_{k=1}^{d} 1_{]0,\infty[}(x^{(k)}) \frac{1}{\lambda_j} \exp\left(-\frac{x^{(k)}}{\lambda_j}\right), \quad j = 1, 2,$$

eli kun $J = j$, X:n komponentit ovat riippumattomia eksponenttijakautuneita satunnaismuuttujia odotusarvolla λ_j. Tarkastellaan luennoilla määriteltyä satunnaismuuttujaa

$$S(X) = -\log\left[\frac{f_1(X)}{f_2(X)}\right].$$

Johda $S(X)$:n ehdolliset tiheysfunktiot g_1 ja g_2, eli siis tiheysfunktiot, joille pätee

$$P(S(X) \in B | J = j) = \int_B g_j(s)\, ds, \quad \text{kun } B \subset \mathbb{R}^d \text{ (Borelin joukko).}$$

(Vihje: tässä auttaa tulos, jonka mukaan riippumattomien eksponettijakautuneiden satunnaismuutujien summalla on gammajakauma; vrt. tn-laskennan jatkokurssi.)

Esitä Bayesin virhetodennäköisyys lausekkeena, jossa esiintyy funktioiden g_1 ja g_2 määrättyjä integraaleja sekä prioritodennäköisyydet P_1 ja P_2.