

Oulun yliopisto
 Matemaattisten tieteiden laitos
 Tilastollinen hahmontunnistus
 13. harjoitus, viikko 16, 2010

1. Olkoon $g : \mathbb{R}^d \rightarrow \{1, \dots, c\}$ luokitin. Määritellään g :n *ehdollinen riski ehdolla* $X = x$ kaavalla

$$e(g, x) = 1 - P(g(x)|x). \quad (1)$$

Tässä siis $P(g(x)|x)$ on luokan $j = g(x)$ posterioritodennäköisyys pisteessä $x \in \mathbb{R}^d$. Koska toisaalta pätee (miksi?)

$$1 - P(g(x)|x) = \sum_{j \neq g(x)} P(j|x),$$

on $e(g, x)$ itseasiassa (miksi?) todennäköisyys, että g luokittelee väärin pisteessä x eli

$$e(g, x) = P(g(X) \neq J|X = x).$$

Osoita kaavan (1) avulla, että g :n virhetodennäköisyydelle $e(g)$ pätee

$$e(g) = \mathbb{E}e(g, X). \quad (2)$$

(Vihje: Lauseen 3.3 todistuksessa johdetusta luokitteluvirheen kaavasta on hyötyä).

2.

Olkoon $c = 2$ ja tarkastellaan Bayesin luokitinta g^* ja sen luokitteluvirhettä $e(g^*)$. Tehtävänä on johtaa $e(g^*)$:lle ns. *Chernoffin yläraja*

$$e(g^*) \leq P_1^u P_2^{1-u} e^{-J_C(u)},$$

missä $u \in]0, 1[$ on mielivaltainen ja

$$J_C(u) = -\log \int_{\mathbb{R}^d} f_1(x)^u f_2(x)^{1-u} dx$$

on luokkien 1 ja 2 välinen *Chernoffin etäisyys*.

Menettele seuraavasti:

1. Osoita kaavan (1) avulla, että

$$e(g^*, x) = \min\{P(1|x), P(2|x)\}.$$

2. Päättelä tästä, että kaikilla $u \in]0, 1[$ pätee

$$e(g^*, x) \leq P(1|x)^u P(2|x)^{1-u}, \quad x \in \mathbb{R}^d.$$

Sovella sitten $P(j|x)$:n määritelmää (kaava (3.3)) ja kaavaa (2).

3. Olkoot f_1 ja f_2 normaalisti jakautuneiden luokkien $N(\mu_1, \Sigma_1)$ ja $N(\mu_2, \Sigma_2)$ tiheysfunktiot. Voidaan osoittaa, että tällöin luokkien Chernoffin etäisyydelle saadaan lauseke

$$J_C(u) = -\log \int f_1^u(x) f_2^{1-u}(x) dx = \frac{1}{2} u(1-u)(\mu_2 - \mu_1)^T [(1-u)\Sigma_1 + u\Sigma_2]^{-1} (\mu_2 - \mu_1) + \frac{1}{2} \log \frac{\det((1-u)\Sigma_1 + u\Sigma_2)}{(\det \Sigma_1)^{1-u} (\det \Sigma_2)^u},$$

kun $0 < u < 1$. Osoita, että jos $\Sigma_1 = \Sigma_2$, niin $J_C(u)$ maksimoituu, kun $u = 1/2$. Lukua $J_B = J_C(1/2)$ sanotaan luokkien *Bhattacharyyan etäisyydeksi*. Totea myös, että jos lisäksi vielä $P_1 = P_2$, saadaan Bayesin virhetodennäköisyydelle pienin mahdollinen Chernoffin yläraja, kun $u = 1/2$.

4. Olkoot f_1 ja f_2 normaalisti jakautuneiden luokkien $N(0, I_d)$ ja $N(0, \Lambda)$ tiheysfunktiot, jossa

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d), \quad \lambda_i > 0.$$

- Osoita, että edellisessä tehtävässä annettu luokkien Bhattacharyyan etäisyys voidaan saattaa muotoon

$$J_B = \frac{1}{2} \sum_{i=1}^d \log \frac{1 + \lambda_i}{2\sqrt{\lambda_i}}.$$

- Olkoon $\varepsilon > 0$ ja olkoon

$$|\lambda_i - 1| > \varepsilon, \quad i = 1, \dots, d.$$

Osoita, että tällöin Bayesin luokittimen virhetodennäköisyys $e(g^*) \rightarrow 0$, kun $d \rightarrow \infty$. Onko tulos oman intuitiosi mukainen?

5. Olkoon g luokitin. Sen virhetodennäköisyyttä $e(g)$ arvioidaan erillisellä otannalla saadulla aineistolla

$$X_{ji}, \quad j = 1, \dots, c, \quad i = 1, \dots, m_j.$$

Määritellään

$$\hat{e}_j(g) = \frac{1}{m_j} \sum_{i=1}^{m_j} 1\{g(X_{ji}) \neq j\}, \quad j = 1, \dots, c$$

ja käytetään $e(g)$:lle estimaattoria $\hat{e}(g) = \sum_{j=1}^c P_j \hat{e}_j(g)$.

Todista, että $\mathbb{E} \hat{e}_j(g) = e_j(g)$, missä $e_j(g) = P(g(X) \neq j \mid J = j)$, (vrt. luku 3.2) ja päättelee edelleen, että $\mathbb{E} \hat{e}(g) = e(g)$, eli että $\hat{e}(g)$ on harhaton. Todista vielä, että estimaattorin varianssille pätee kaava

$$D^2(\hat{e}(g)) = \sum_{j=1}^c \frac{P_j^2}{m_j} e_j(g)(1 - e_j(g)).$$

(Tämän takia varianssia voidaan suurissa otoksissa arvioida estimaattorilla $\sum_{j=1}^c P_j^2 / m_j \hat{e}_j(g)(1 - \hat{e}_j(g))$.)