

早稲田大学「パターン認識」レポート 問題

2007-11-19 出題

提出期限: 2007-12-3 授業後

(ただしその後も受付はします)

担当: 赤穂昭太郎 s.akaho@aist.go.jp

1 問題

X という病気を判定するために, A, B の2種類の検査を考えた. X は病気であるとき 1 という値を取り, そうでなければ 0, A, B はそれぞれ陽性であれば 1, 陰性であれば 0 とする.

ここで, A と B は X が決まれば互いに独立であると仮定する.

検査の精度を調べるためにあらかじめ X の値がわかっている被験者 N 人をランダムに選んで検査を実施した. その結果, $X = 1$ であるような被験者 N_X 人のうち, N_A 人が $A = 1$, また, N_B 人が $B = 1$ であった. 一方, $X = 0$ である被験者 $N - N_X$ 人に対しては M_A 人が $A = 1$, また, M_B 人が $B = 1$ であった. ここで, 各被験者が $X = 1$ である確率 $P(X = 1)$ は (ほかの人によらず) 独立かつ同じ値であると仮定する.

確率的な推論を行う際に必要となる確率は以下の 5 つのパラメータ $\alpha_1, \alpha_0, \beta_1, \beta_0, \gamma$ により決めることができる.

$$\alpha_1 = P(A = 1 | X = 1), \quad \alpha_0 = P(A = 1 | X = 0), \\ \beta_1 = P(B = 1 | X = 1), \quad \beta_0 = P(B = 1 | X = 0), \quad \gamma = P(X = 1)$$

1. ある人に対する検査結果が $A = 1, B = 0$ だったとき, $X = 1$ である確率を, 上記のパラメータを使って表せ.
2. 被験者に対する実験結果からパラメータを最尤推定により求める式を書け.
3. 以下ではパラメータが確率変数であるようなベイズ推定の枠組で考える. 簡単のため, すべてのパラメータの事前分布の確率密度関数は, $q(w) = w(1-w)/Z$ という形をしているとする ($0 \leq w \leq 1$). Z は $q(w)$ が積分して 1 にするための定数である.
 - (a) Z の値を求めよ. また $q(w)$ はどのような性質の分布であるか.
 - (b) パラメータの値を固定したとき, N 人の被験者の (A, B, X) の値が得られる同時確率 $P(N \text{ 人の被験者の } (A, B, X) \text{ の値} | \alpha_0, \alpha_1, \beta_0, \beta_1, \gamma)$ を求めよ.
 - (c) ある人に対する検査結果が $A = 1, B = 0$ だったとき, $X = 1$ である事後確率 $P(X = 1 | A = 1, B = 0, N \text{ 人の被験者の } (A, B, X) \text{ の値})$ を求めよ.
 - (d) この確率値に基づいて診断を下さなければならないとすると, どのようなことに注意する必要があるか. 考えを述べよ.
4. 病気の診断の精度を改善するために考えられることをできるだけたくさん挙げ, それぞれの長短所について論じよ.