DM-05: 課題 1

dm-05-assign1

1. ある疾病の罹患率は 1000人に1人であることが知られている。この疾病の検査は、罹患している人の95%を正しく「陽性」と判定できる。また罹患していない人の90%を正しく「陰性」と判定できる。検査結果が陽性だったときに、実際に罹患している確率を、ベイズ推定の考え方を用いて求めよ。

   ヒント: P(陽性) の計算には全確率の公式を用いる

2. 1回目の検査が陽性だったとき、罹患している人の98%を正しく「陽性」と、また罹患していない人の95%を正しく「陰性」と判定できる再検査を受けた。再検査結果も陽性だったときに、実際に罹患している確率を、ベイズ推定の考え方を用いて求めよ。

   ヒント: 1. の事後確率を、2. の事前確率とする。

dm-05-assign1 解答例

dm-05-assign1-ans.ipynb

1.

- 事前確率 P(罹患) = 1 / 1000 = 0.001
- 尤度 P(陽性|罹患) = 0.95
- P(陽性) = P(陽性|罹患) x P(罹患) + P(陽性|非罹患) x P(非罹患)

全確率の公式 = 0.95 x 0.001 + (1-0.90) x (1-0.001) = 0.100…

- 事後確率 P(罹患|陽性)

  = P(陽性|罹患) x P(罹患) / P(陽性)

  = 0.95 x 0.001 / 0.100…

  = 0.00941… (= 0.941…%)

dm-05-assign1 解答例

2.

- 事前確率 P(罹患) = 0.00941…
- 尤度 P(陽性|罹患) = 0.98
- P(陽性) = P(陽性|罹患) x P(罹患) + P(陽性|非罹患) x P(非罹患)

全確率の公式 = 0.98 x 0.00941… + (1-0.95) x (1-0.00941…) = 0.05876..

- 事後確率 P(罹患|陽性)

  = P(陽性|罹患) x P(罹患) / P(陽性)

  = 0.98 x 0.00941… / 0.05876…

  = 0.1571… (= 15.71…%)

DM-05: 課題 2

dm-05-assign2

Mr. O、Ms. H、Mr. Tの3氏の、最近2週間のつぶやきを、解析対象を10語に絞って調べたところ、以下の出現があった。

O氏(つぶやき数10): USA great Democrats care mother

H氏(つぶやき数15): mother love Democrats care mother USA

T氏(つぶやき数25): Russia fake USA Mexico great Mexico haters

「USA Democrats mother」を含む新たなつぶやきが発信された。上のデータと単純ベイズ分類器の考え方をもとに、この新たなつぶやきが、3氏のうちの誰のものと考えられるか予測せよ。なお、各氏の最近2週間のつぶやき数を、事前確率とすること。また、全度数の初期値を1とするLaplace smoothingを用いよ。

dm-05-assign2

提出するもの

1. 3氏それぞれの事前確率
2. 3氏それぞれの尤度
3. 3氏それぞれの事後確率 (P(単語群)=1とみなす)
4. 3.をpost['O'], post['H'], post['T'], この3つの和をpost['tot'] としたとき、post['O'] / post['tot'], post['H'] / post['tot'], post['T'] / post['tot']の値
5. 予測結果 (3氏のうち誰のものと予測するか)

dm-05-assign2 解答例

dm-05-assign2-ans.ipynb

1. 事前確率
   - P(O) = 10 / (10+15+25) = 1 / 5 = 0.2
   - P(H) = 15 / (10+15+25) = 3 / 10 = 0.3
   - P(T) = 25 / (10+15+25) = 1 / 2 = 0.5
2. 尤度
   - P(USA|O) x P(Democrats|O) x P(mother|O) = 2/15 x 2/15 x 2/15
   - P(USA|H) x P(Democrats|H) x P(mother|H) = 2/16 x 2/16 x 3/16
   - P(USA|T) x P(Democrats|T) x P(mother|T) = 2/17 x 1/17 x 1/17

dm-05-assign2 解答例

dm-05-assign2-ans.ipynb

3. 事後確率 = 尤度 x 事前確率
   - P(O|単語群) = 2/15 x 2/15 x 2/15 x 0.2 = 8 / 16875 = 0.0004744…
   - P(H|単語群) = 2/16 x 2/16 x 3/16 x 0.3 = 9 / 10240 = 0.0008789…
   - P(T|単語群) = 2/17 x 1/17 x 1/17 x 0.5 = 1 / 4913 = 0.0002035…
4. 割合
   - P(O|単語群) / (P(O|単語群) + P(H|単語群) + P(T|単語群)) = 0.3045…
   - P(H|単語群) / (P(O|単語群) + P(H|単語群) + P(T|単語群)) = 0.5646…
   - P(T|単語群) / (P(O|単語群) + P(H|単語群) + P(T|単語群)) = 0.1307…
5. 最大の事後確率に対応するのは Ms. H