

DM-04: 課題 1

dm-04-assign1

答えの確率 (%) は小数点以下第 1 位までで答えよ。

1. ある感染症の罹患率は 200人に1人であることが知られている。この感染症の検査は、罹患している人の70%を正しく「陽性」と判定。また罹患していない人の99%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: $P(\text{陽性})$ の計算には全確率の公式を用いる

2. 同様に、検査結果が陰性のとき、実際に罹患していない確率を求めよ。

現在の日本におけるCOVID-19 / PCR検査の状況は、不顕性感染者を含めると、これに近い可能性がある

dm-04-assign1

3. 検査対象を絞り、検査人数中のおよその罹患率を10人に1人にした。この場合、検査結果が陽性だったときに、実際に罹患している確率、および、検査結果が陰性だったときに、実際に罹患していない確率を、ベイズ推定の考え方をを用いて求めよ。
4. 1回目の検査が陰性だったとき、もう一度同じ検査を受けた。再検査結果も陰性だったときに、実際に罹患していない確率を、ベイズ推定の考え方をを用いて求めよ。

ヒント:

3. の事後確率の1つを使って、4. の事前確率（罹患率）を求める。

現在の日本におけるCOVID-19 / PCR検査の状況は、不顕性感染者を含めると、これに近い可能性がある

dm-04-assign1 解答例

dm-04-assign1-ans.ipynb

1.

- 事前確率 $P(\text{罹患}) = 1 / 200 = 0.005$

- 尤度 $P(\text{陽性}|\text{罹患}) = 0.70$

- $P(\text{陽性}) = P(\text{陽性}|\text{罹患}) \times P(\text{罹患}) + P(\text{陽性}|\text{非罹患}) \times P(\text{非罹患})$

全確率の公式 $= 0.70 \times 0.005 + (1-0.99) \times (1-0.005) = 0.01345$

- 事後確率 $P(\text{罹患}|\text{陽性})$

$$= P(\text{陽性}|\text{罹患}) \times P(\text{罹患}) / P(\text{陽性})$$

$$= 0.70 \times 0.005 / 0.01345$$

$$= 0.26022\cdots = \mathbf{26.0\% (Ans.)}$$

2.

- 事前確率 $P(\text{罹患}) = 1 / 200 = 0.005$
- 尤度 $P(\text{陽性}|\text{罹患}) = 0.70$
- $P(\text{陰性}) = P(\text{陰性}|\text{罹患}) \times P(\text{罹患}) + P(\text{陰性}|\text{非罹患}) \times P(\text{非罹患})$

全確率の公式 $= (1-0.70) \times 0.005 + 0.99 \times (1-0.005) = 0.98655$

- 事後確率 $P(\text{非罹患}|\text{陰性})$
 $= P(\text{陰性}|\text{非罹患}) \times P(\text{非罹患}) / P(\text{陰性})$
 $= 0.99 \times (1-0.005) / 0.98655$
 $= 0.99847\cdots = \mathbf{99.8\% (Ans.)}$

3-1.

- 事前確率 $P(\text{罹患}) = 1 / 10 = 0.1$

- 尤度 $P(\text{陽性}|\text{罹患}) = 0.70$

- $P(\text{陽性}) = P(\text{陽性}|\text{罹患}) \times P(\text{罹患}) + P(\text{陽性}|\text{非罹患}) \times P(\text{非罹患})$

全確率の公式 $= 0.70 \times 0.1 + (1-0.99) \times (1-0.1) = 0.079$

- 事後確率 $P(\text{罹患}|\text{陽性})$

$$= P(\text{陽性}|\text{罹患}) \times P(\text{罹患}) / P(\text{陽性})$$

$$= 0.70 \times 0.1 / 0.079$$

$$= 0.88607\cdots = \mathbf{88.6\% \text{ (Ans.)}}$$

3-2.

- 事前確率 $P(\text{罹患}) = 1 / 10 = 0.1$
- 尤度 $P(\text{陽性}|\text{罹患}) = 0.70$
- $P(\text{陰性}) = P(\text{陰性}|\text{罹患}) \times P(\text{罹患}) + P(\text{陰性}|\text{非罹患}) \times P(\text{非罹患})$

全確率の公式 $= (1-0.70) \times 0.1 + 0.99 \times (1-0.1) = 0.921$

- 事後確率 $P(\text{非罹患}|\text{陰性})$
 $= P(\text{陰性}|\text{非罹患}) \times P(\text{非罹患}) / P(\text{陰性})$
 $= 0.99 \times (1-0.1) / 0.921$
 $= 0.96742\cdots = \mathbf{96.7\% (Ans.)}$

4.

- 事前確率 $P(\text{罹患}) = 1 - 0.96742 \dots = 0.03258\dots$
- 尤度 $P(\text{陽性}|\text{罹患}) = 0.70$
- $P(\text{陰性}) = P(\text{陰性}|\text{罹患}) \times P(\text{罹患}) + P(\text{陰性}|\text{非罹患}) \times P(\text{非罹患})$
 $\text{全確率の公式} = (1-0.70) \times 0.03258 + 0.99 \times (1-0.03258) = 0.96751\dots$
- 事後確率 $P(\text{非罹患}|\text{陰性})$
 $= P(\text{陰性}|\text{非罹患}) \times P(\text{非罹患}) / P(\text{陰性})$
 $= 0.99 \times 0.96742 / 0.96751\dots$
 $= 0.98990\dots = \mathbf{99.0\% (Ans.)}$

DM-04: 課題 2

dm-04-assign2

果物に関する文書を分類予測する。以下の6語を解析対象とする。すでに分類されている文章における単語の出現は以下の通りであった。

りんご: 赤 甘味 皮 種

いちご: 赤 甘味 甘味 酸味

みかん: 橙 皮 酸味 甘味

「赤 甘味 酸味」を含む文書を、単純ベイズ分類器の考え方により分類予測し、各クラスの事後確率、分類予測結果を報告せよ。なお、事前確率は $1/3$ ずつとし、 $\alpha = 1$ の Laplace smoothing を用いること。確率の計算に対数を用いる必要はなく、事後確率は約分しない分数のままでよい。

dm-04-assign2 解答例

尤度

10 のうちの 6 はLaplace smoothing分

- $P(\text{赤}|\text{りんご}) \times P(\text{甘味}|\text{りんご}) \times P(\text{酸味}|\text{りんご}) = 2/10 \times 2/10 \times 1/10$
- $P(\text{赤}|\text{いちご}) \times P(\text{甘味}|\text{いちご}) \times P(\text{酸味}|\text{いちご}) = 2/10 \times 3/10 \times 2/10$
- $P(\text{赤}|\text{みかん}) \times P(\text{甘味}|\text{みかん}) \times P(\text{酸味}|\text{みかん}) = 1/10 \times 2/10 \times 2/10$

事後確率 = 尤度 × 事前確率

- | | 割合 |
|---|-----|
| ■ $P(\text{りんご} \text{単語群}) = 2/10 \times 2/10 \times 1/10 \times 1/3 =$ 4 / 3000 | 0.2 |
| ■ $P(\text{いちご} \text{単語群}) = 2/10 \times 3/10 \times 2/10 \times 1/3 =$ 12 / 3000 | 0.6 |
| ■ $P(\text{みかん} \text{単語群}) = 1/10 \times 2/10 \times 2/10 \times 1/3 =$ 4 / 3000 | 0.2 |

最大の事後確率に対応するのは 「いちご」