DM-04: Assignment1

dm-04-assign1

The probability of the answer (%) should be rounded off to one decimal place.

1. The incidence rate of a infectious disease is known to be 1 per 200. Diagnostic assay of this disease can correctly determine 70% of being illness as positive, and 99% of not being illness as negative. When the result of a diagnostic assay was positive, determine the probability of being illness with the way of thinking of Bayesian Inference.

   Hint: Use the formula of total probability to compute P(posi)

2. Similarly, determine the probability of not being illness when the result of a diagnostic assay was negative.

The current state of COVID-19 / PCR testing in Japan is likely to be similar to this, including those with subclinical infection.

dm-04-assign1

3. Select the target of assay to increase the incidence rate of the infectious disease to 1 per 10. In this situation, determine the probability of being illness when the result of the diagnostic assay was positive, and that of not being illness when the result of the diagnostic assay was negative, with the way of thinking of Bayesian Inference
4. When the result of the first diagnostic assay was negative, and the result of the re-examination was also negative, determine the probability of not being illness with the way of thinking of Bayesian Inference.

Hint: Obtain the prior probability of 4 (ratio of infection) from one of the posterior probabilities of 3.

The current state of COVID-19 / PCR testing in Japan is likely to be similar to this, including those with subclinical infection.

Ans. of dm-04-assign1

dm-04-assign1-ans.ipynb

1.

- Prior Probability P(ill) = 1 / 200 = 0.005
- Likelihood P(posi | ill) = 0.70
- P(posi) = P(posi | ill) x P(ill) + P(posi | no_ill) x P(no_ill)
  = 0.70 x 0.005 + (1-0.99) x (1-0.005) = 0.01345

formula of total probability

- Posterior probability P(ill | posi)
  = P(posi | ill) x P(ill) / P(posi)
  = 0.70 x 0.005 / 0.01345
  = 0.26022··· = **26.0% (Ans.)**

2.

- Prior probability P(ill) = 0.00941…
- Likelihood P(posi | ill) = 0.98
- P(nega) = P(nega|ill) x P(ill) + P(nega|no_ill) x P(no_ill)

  = (1-0.70) x 0.005 + 0.99 x (1-0.005) = 0.98655

formula of total probability

- Posterior probability P(no_ill|nega)

  = P(nega|no_ill) x P(no_ill) / P(nega)

  = 0.99 x (1-0.005) / 0.98655

  = 0.99847… = **99.8% (Ans.)**

3-1.

- Prior probability $P(ill) = 1 / 10 = 0.1$
- Likelihood $P(posi \mid ill) = 0.70$
- $P(posi) = P(posi|ill) \times P(ill) + P(posi|no\_ill) \times P(no\_ill)$

  $= 0.70 \times 0.1 + (1-0.99) \times (1-0.1) = 0.079$
- Posterior probability $P(ill|posi)$

  $= P(posi|ill) \times P(ill) / P(posi)$

  $= 0.70 \times 0.1 / 0.079$

  $= 0.88607\cdots =$ **88.6% (Ans.)**

formula of total probability

3-2.

- Prior probability P(ill) = 1 / 10 = 0.1
- Likelihood P(posi | ill) = 0.70
- P(nega) = P(nega|ill) x P(ill) + P(nega|no_ill) x P(no_ill)

  = (1-0.70) x 0.1 + 0.99 x (1-0.1) = 0.921
- Posterior probability P(no_ill|nega)

  = P(nega|no_ill) x P(no_ill) / P(nega)

  = 0.99 x (1-0.1) / 0.921

  = 0.96742… = **96.7% (Ans.)**

formula of total probability

4.

- Prior probability P(ill) = 1 - 0.96742 ··· = 0.03258···
- Likelihood P(posi | ill) = 0.70
- P(nega) = P(nega|ill) x P(ill) + P(nega|no_ill) x P(no_ill)

  = (1-0.70) x 0.03258 + 0.99 x (1-0.03258) = 0.96751···

formula of total probability

- Posterior probability P(no_ill|nega)

  = P(nega|no_ill) x P(no_ill) / P(nega)

  = 0.99 x 0.96742 / 0.96751···

  = 0.98990··· = **99.0% (Ans.)**

DM-04: Assignment 2

dm-04-assign2

Make classification prediction for documents about fruits. The following 6 words are subjects for analysis: red, orange, sweetness, sourness, skin, seeds. The occurrence of words in the classified documents was as follows.

Apple: red sweetness skin seeds

Strawberry: red sweetness sweetness sourness

Orange: orange skin sourness sweetness

Make classification prediction by thinking of naive Bayes classifier for a document containing "red sweetness sourness", report posterior probability of each class and classification result. Note: the prior probability is 1/3 for each class, and alpha = 1 for Laplace smoothing. It is not necessary to use logarithms to calculate the probability, and the posterior probability could be a fraction without reducing.

Ans. of dm-04-assign2

Likelihood

6 out of 10 come from Laplace smoothing

- P(red|apple) x P(sweetness|apple) x P(sourness|apple) = 2/10 x 2/10 x 1/10
- P(red|strawberry) x P(sweetness|strawberry) x P(sourness|strawberry) = 2/10 x 3/10 x 2/10
- P(red|orange) x P(sweetness|orange) x P(sourness|orange) = 1/10 x 2/10 x 2/10

posterior prob. = Likelihood x prior prob.

		ratio
P(apple|words) = 2/10 x 2/10 x 1/10 x 1/3	**= 4 / 3000**	0.2
P(strawberry|words) = 2/10 x 3/10 x 2/10 x 1/3	**= 12 / 3000**	0.6
P(apple|words) = 1/10 x 2/10 x 2/10 x 1/3	**= 4 / 3000**	0.2

The largest posterior probability: **"strawberry"**