CZ4001: Tutorial Week 2

Due on January 21, 2021 at 8:30am

Assoc Prof Pan, Sinno Jialin - CS4

Pang Yu Shao
U1721680D

21/01/2021

Problem 1

Show how to estimate $P(Y=0|X=0)$ and $P(Y=2|X=0)$ on the 17th page of the lecture notes "Lecture 2a".

Solution

For $\mathbf{P(Y=0|X=0)}$:

$$P(Y=0|X=0) = \frac{P(X=0|Y=0)\times P(Y=0)}{P(X=0)}$$

$$= \frac{P(X=0|Y=0)\times P(Y=0)}{P(X=0,Y=0)+P(X=0,Y=1)+P(X=0,Y=2)}$$

$$= \frac{P(X=0|Y=0)\times P(Y=0)}{P(X=0|Y=0)\times P(Y=0)+P(X=0|Y=1)\times P(Y=1)+P(X=0|Y=2)\times P(Y=2)}$$

$$= \frac{0.54\times 0.39}{0.54\times 0.39+0.44\times 0.3+0.49\times 0.31}$$

$$= \frac{0.2106}{0.4945}$$

$$= \mathbf{0.426}$$

For $\mathbf{P(Y=2|X=0)}$:

$$P(Y=2|X=0) = \frac{P(X=0|Y=2)\times P(Y=2)}{P(X=0)}$$

$$= \frac{P(X=0|Y=2)\times P(Y=2)}{P(X=0,Y=0)+P(X=0,Y=1)+P(X=0,Y=2)}$$

$$= \frac{0.49\times 0.31}{0.4945}$$

$$= \frac{0.1519}{0.4945}$$

$$= \mathbf{0.307}$$

Problem 2

Suppose that if a person has lung cancer, his/her probability of having gene X is 0.9, and if a person does not have lung cancer, his/her probability of having gene X is 0.2. The probability of a person having lung cancer is 0.01. Now, we know that a patient A has gene X

1. Please use Bayesian decision theory with 0/1 loss to predict whether the patient A has lung cancer or notes
2. Consider that costs of misclassification are different. Assume that the cost for correct decisions is 0, the cost of misclassifying a person who does not have lung cancer to be a person with lung cancer is 0.007, and the cost of misclassifying a person who has lung cancer to be a healthy patient is 1. Please use Bayesian decision theory with the predefined loss to predict whether the patient A has lung cancer or not.

Solution
Part 1
First, list down all information:

- $P(X = 1|LC = 1) = 0.9$
- $P(X = 1|LC = 0) = 0.2$
- $P(LC = 1) = 0.01$
- $P(LC = 0) = 0.99$

Using Bayes Rule:

- $P(LC = 0|X = 1) = \frac{P(X=1|LC=0) \times P(LC=0)}{P(X=1)} = \frac{0.198}{P(X=1)}$
- $P(LC = 1|X = 1) = \frac{P(X=1|LC=1) \times P(LC=1)}{P(X=1)} = \frac{0.009}{P(X=1)}$

Risk for taking action a_i: $R(a_i|x) = \sum_{c=0}^{C-1} \lambda_{ic} P(y = c|x)$

Risk of predicting LC = 1 given X = 1:

$$\begin{aligned} R(a_{LC=1}|X = 1) &= \sum_{c=0}^{C-1} \lambda_{ic} P(LC = c|X = 1) \\ &= 1 * (P(LC = 0|X = 1)) + 0 * (P(LC = 1|X = 1)) \\ &= 0.198/P(X = 1) \end{aligned}$$

Risk of predicting LC = 0 given X = 1:

$$\begin{aligned} R(a_{LC=0}|X = 1) &= \sum_{c=0}^{C-1} \lambda_{ic} P(LC = c|X = 1) \\ &= 0 * (P(LC = 0|X = 1)) + 1 * (P(LC = 1|X = 1)) \\ &= \mathbf{0.009}/P(X = 1) \end{aligned}$$

Patient A is predicted to not have Lung Cancer

Problem 2 continued on next page. . .

Part 2

Risk of predicting LC = 1 given X = 1 (misclassification cost = 0.007):

$$\begin{aligned} R(a_{LC=1}|X=1) &= \sum_{c=0}^{C-1} \lambda_{ic} P(LC=c|X=1) \\ &= 0.007 * (P(LC=0|X=1)) + 0 * (P(LC=1|X=1)) \\ &= \mathbf{0.001386}/P(X=1) \end{aligned}$$

Risk of predicting LC = 0 given X = 1 (misclassification cost = 1):

$$\begin{aligned} R(a_{LC=0}|X=1) &= \sum_{c=0}^{C-1} \lambda_{ic} P(LC=c|X=1) \\ &= 0 * (P(LC=0|X=1)) + 1 * (P(LC=1|X=1)) \\ &= 0.009/P(X=1) \end{aligned}$$

Patient A is predicted to have Lung Cancer