Ludwig-Maximilians-Universitaet Muenchen
Institute for Informatics
Prof. Dr. Volker Tresp
Rajat Koner
Andreas Lohrer
Sebastian Schmoll

11.06.2021

Machine Learning

Summer 2021

Exercise Sheet 8

Exercise 8-1 Conditional Probability

If screening for a disease, there are several possible outcomes. Let T, $\neg T$ denote the events that the test is positive and negative, respectively, and D, $\neg D$ denote the events of having and not having the disease, respectively. There are two major criteria to evaluate tests by:

- Sensitivity: Probability (in practice more likely: ratio) of positively tested people having the disease, i.e., $P(T \mid D)$.
- Specificity: Probability (or ratio) of negatively tested people not having the disease, i.e., $P(\neg T \mid \neg D)$.

Now, assume a (realistic) test for HIV with a sensitivity and specificity of 99.9%. Suppose that a person is randomly selected from a population where 1% are infected with HIV and tested with the aforementioned test. Compute the probability that the person has HIV if:

(a) The test is positive.

(b) The test is negative.

Possible Solution

(a) $P(D \mid T) = \frac{P(T|D) \cdot P(D)}{P(T)}$
where

$$\begin{aligned} P(T) &= P(T \mid D) \cdot P(D) + P(T \mid \neg D) \cdot P(\neg D) \\ &= P(T \mid D) \cdot P(D) + (1 - P(\neg T \mid \neg D)) \cdot P(\neg D) \\ &= (0.999 \cdot 0.01) + ((1 - 0.999) \cdot 0.99) = 0.01098 \end{aligned}$$

It follows: $P(D \mid T) = \frac{(0.999 \cdot 0.01)}{0.01098} \sim 0.9098$

(b) $P(D \mid \neg T) = \frac{P(\neg T|D) \cdot P(D)}{P(\neg T)}$
where

$$\begin{aligned} P(\neg T) &= P(\neg T \mid D) \cdot P(D) + P(\neg T \mid \neg D) \cdot P(\neg D) \\ &= (1 - P(T \mid D)) \cdot P(D) + P(\neg T \mid \neg D) \cdot P(\neg D) \\ &= (0.001 \cdot 0.01) + (0.999 \cdot 0.99) = 0.98902 \end{aligned}$$

It follows: $P(D \mid \neg T) = \frac{0.001 \cdot 0.01}{0.98902} \sim 1 \cdot 10^{-5}$

Exercise 8-2 Maximum Likelihood Estimator I

Suppose that X is a discrete random variable with the following probability mass function, where $0 < \theta < 1$ is a parameter.

X	0	1	2	3
P(X)	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

The following 10 independent observations were taken from such a distribution: $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$. What is the maximum likelihood estimate of θ?

Possible Solution

Since the sample is $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$, the likelihood is

$$L(\theta) = \prod_{i=1}^{n} P(X_i \mid \theta) = \tag{1}$$

$$= P(X=3)P(X=0)P(X=2)P(X=1)P(X=3)P(X=2)P(X=1)P(X=0)P(X=2)P(X=1) = \tag{2}$$

$$= \left(\frac{2\theta}{3}\right)^2 \left(\frac{\theta}{3}\right)^3 \left(\frac{2(1-\theta)}{3}\right)^3 \left(\frac{1-\theta}{3}\right)^2 \tag{3}$$

Taking the log likelihood function of it, yields:

$$\mathcal{L}(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log P(X_i|\theta) \tag{4}$$

$$= 2\left(\log\frac{2}{3} + \log\theta\right) + 3\left(\log\frac{1}{3} + \log\theta\right) + 3\left(\log\frac{2}{3} + \log(1-\theta)\right) + 2\left(\log\frac{1}{3} + \log(1-\theta)\right) \tag{5}$$

$$= \log\left(\frac{2^2 \cdot 1^3 \cdot 2^3 \cdot 1^2}{3^2 \cdot 3^3 \cdot 3^3 \cdot 3^2}\right) + 5\log\theta + 5\log(1-\theta) \tag{6}$$

where the first summand does not depend on θ. It can be seen that the log likelihood function is easier to maximize compared to the likelihood function. Let the derivative of $\mathcal{L}(\theta)$ with respect to θ be zero:

$$\mathcal{L}' = \frac{\partial\mathcal{L}(\theta)}{\partial\theta} = \frac{5}{\theta} - \frac{5}{1-\theta} = 0 \tag{7}$$

and the solution gives us the MLE, which is $\hat{\theta}^{ML} = \frac{1}{2}$.

Note also for the second derivative (considering the curvature):

$$\mathcal{L}''(\theta) = -\frac{5}{\theta^2} - \frac{5}{(1-\theta)^2} < 0 \tag{8}$$

Exercise 8-3 Frequentist versus Bayesian Statistics

Consider this example to illustrate the difference between frequentist and bayesian statistics: Alice and Bob play a game in which the first person to get 6 points wins. The points are scored in the following way: A referee is standing at a pool table Alice and Bob cannot see. Before the game begins, the referee rolls a ball onto the table coming to rest at a random position. Each point scored is decided by the referee rolling another ball. If the ball comes to rest left of (the middle of) the initial ball, Alice scores, if it comes to rest right, Bob scores. The players know nothing but who scored a point. If the portion left of the initial ball is denoted as p, it is obvious that the probability of Alice scoring a point is p.

Now, consider the following situation within the game: Alice has 5 points, Bob has 3. Let us investigate the probability of Bob winning.

(a) Assume that the initial ball came to rest such that $p = 2/3$. What is the probability that Bob wins?

Possible Solution

Obviously, $P(\text{Bob wins}) = (1-p)^3 = 1/27 \approx 0.037$

(b) Unfortunately, we do not know p – we only have some data we can try to estimate it from. Follow a *frequentist* approach: compute the maximum likelihood estimator for p and the probability of Bob winning.

Possible Solution

$$\mathcal{L}(p) = \log L(p) = \sum_{i=1}^{8} \log P_p(x_i) = 3\log(1-p) + 5\log(p)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial p} &= \frac{-3}{1-p} + \frac{5}{p} \stackrel{!}{=} 0 \\ &\Leftrightarrow 5 - 5p = 3p \\ &\Leftrightarrow \hat{p}^{\text{ML}} = 5/8 \end{aligned}$$

Therefore, $P(\text{Bob wins}) = (1 - \hat{p}^{\text{ML}})^3 \approx 0.053$.

(c) Now, let us follow a *bayesian* approach: We know that p is only dependent on the position of the initial ball which we assume to be uniformly distributed on the table, i.e., $unif[0,1]$. Note that we compute the expected probability of Bob winning, as p itself is now drawn from a distribution. *Hint:* You will need the beta function:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{(x-1)!(y-1)!}{(x+y-1)!}$$

Possible Solution

The expected probability of Bob winning is an integral over all possible values of p:

$$E(\text{Bob wins}) = \int_0^1 (1-p)^3 \cdot P(p \mid A=5, B=3)dp$$

In order to compute the posterior probability $P(p \mid A=5, B=3)$ we need Bayes' formula:

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)} = \frac{P(Y \mid X)P(X)}{\sum_Z P(Y \mid Z)P(Z)}$$

where the Z are a partition of the event space. In this particular case:

$$P(p \mid A=5, B=3) = \frac{P(A=5, B=3 \mid p)P(p)}{\int_0^1 P(A=5, B=3 \mid p)P(p)dp}$$

We can compute $P(A=5, B=3 \mid p)$ (binomial):

$$P(A=5, B=3 \mid p) = \frac{8!}{5!3!}p^5(1-p)^3 = 56p^5(1-p)^3 \quad (\star)$$

A crucial feature of the 'table game' is that $P(p)$ is well-defined: the game is contrived such that p is picked from a uniform distribution. Because it is uniform, it is a constant, and it cancels out of the Bayes equation, i.e. we can pull it out of the integral and cancel it out. It remains:

$$P(p \mid A = 5, B = 3) = \frac{P(A = 5, B = 3 \mid p)}{\int_0^1 P(A = 5, B = 3 \mid p) dp}$$

which we plug into the original formula:

$$\begin{aligned} E(\text{Bob wins}) &= \frac{\int_0^1 (1-p)^3 \cdot P(A = 5, B = 3 \mid p) dp}{\int_0^1 P(A = 5, B = 3 \mid p) dp} \\ &\underset{\star}{=} \frac{\int_0^1 p^5 (1-p)^6 dp}{\int_0^1 p^5 (1-p)^3 dp} \\ &\underset{\text{Beta}}{=} \frac{5!6!}{12!} \frac{9!}{5!3!} \\ &= 1/11 \approx 0.09 \end{aligned}$$

Exercise 8-4 Maximum Likelihood Estimator - Exponential distribution

Measuring the decays of radioactive material, we observe the probability distribution of

$$f(t \mid \tau) = \frac{1}{\tau} \cdot e^{\frac{-t}{\tau}} \tag{9}$$

where τ denotes the mean lifetime. In an experiment we observe n decays with measured time of t_i, $i = 1, \ldots, n$. Set up the likelihood function and specify the maximum likelihood estimator.

Possible Solution

The likelihood function of the sample is given by:

$$L(t_1, \ldots, t_n \mid \tau) = \prod_{i=1}^{n} \frac{1}{\tau} \cdot e^{\frac{-t_i}{\tau}} \tag{10}$$

$$\overset{(*)}{\Rightarrow} \mathcal{L}(t_1, \ldots, t_n \mid \tau) = \sum_{i=1}^{n} (-ln\tau - \frac{t_i}{\tau}) \tag{11}$$

$$(*)\ log(\prod_{i=1}^{n} a_i) = \sum_{i=1}^{n} log(a_i) \tag{12}$$

Maximizing of $\mathcal{L}$ yields the ML-estimator of τ:

$$\frac{\partial \mathcal{L}}{\partial \tau} = \sum_{i=1}^{n} \left(-\frac{1}{\tau} + \frac{t_i}{\tau^2} \right) = 0 \tag{13}$$

$$\Rightarrow \hat{\tau}^{ML} = \frac{1}{n} \sum_{i=1}^{n} t_i = \bar{t} \tag{14}$$

The ML-estimator of the mean lifetime is therefore the arithmetic mean of the observered lifetimes.