Machine Learning from Data: Homework 3 - Probabilities

227367455 and 323081950

May 7, 2021

Question 1

Given a random sample $\{x_1, x_2, ..., x_n\}$, derive the maximum likelihood estimator $\hat{p}$ of the Binomial distribution.

$$B(x, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

We first want to calculate the likelihood:

$L = P(x_1, ...x_n \mid p) = \prod_{i=1}^{n} P(x_i \mid p)$

$= \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$

$= \prod_{i=1}^{n} p^{x_i} (1-p)^{n-x_i} \prod_{i=1}^{n} \binom{n}{x_i}$

$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n^2 - \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{n}{x_i}$

From the likelihood we calculate the log-likelihood:

$$ln(L) = ln(p^{\sum_{i=1}^{n} x_i}(1-p)^{n^2-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{n}{x_i})$$

$$= ln(p^{\sum_{i=1}^{n} x_i}) + ln((1-p)^{n^2-\sum_{i=1}^{n} x_i}) + ln(\prod_{i=1}^{n} \binom{n}{x_i})$$

$$= ln(p) \sum_{i=1}^{n} x_i + ln(1-p)(n^2 - \sum_{i=1}^{n} x_i) + \sum_{i=1}^{n} ln(\binom{n}{x_i})$$

We will take the derivative in respect to p our given value:

$$\frac{\partial[ln(L)]}{\partial p} = \frac{\partial[ln(p) \sum_{i=1}^{n} x_i]}{\partial p} + \frac{\partial[ln(1-p)(n^2 - \sum_{i=1}^{n} x_i)]}{\partial p}$$

$$= \frac{\sum_{i=1}^{n} x_i}{p} - \frac{(n^2 - \sum_{i=1}^{n} x_i)}{1-p}$$

To find the a maximum we set the derivative to 0 obtaining:

$$\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n^2 - \sum_{i=1}^{n} x_i}{1-p} = 0$$

$$(1-p) \sum_{i=1}^{n} x_i - p(n^2 - \sum_{i=1}^{n} x_i) = 0$$

$$\sum_{i=1}^{n} x_i - p \sum_{i=1}^{n} x_i - pn^2 + p \sum_{i=1}^{n} x_i) = 0$$

$$\sum_{i=1}^{n} x_i - pn^2 + = 0$$

$$pn^2 = \sum_{i=1}^{n} x_i$$

Thus we obtain:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n^2}$$

Question 2

A student wants to know her chances to pass and fail an exam if she studies and if she doesn't study. From last year's results, she sees that $P(Pass) = 60\%$. She also found out that $P(Studied \mid Pass) = 95\%$, $P(Studied \mid Failed) = 60\%$. You can assume that every student either studied or didn't study, and either passed or failed.

a.

What is her probability of passing the exam if she studies?

$$P(Pass|Studied) = \frac{P(Studied \mid Pass)P(Pass)}{P(Studied)}$$

b.

What is her probability of passing if she doesn't study?

$$P(Pass \mid \overline{Studied}) = \frac{P(\overline{Studied} \mid Pass)P(Pass)}{P(\overline{Studied})}$$

Question 3

a.

b.

c.

i.

ii.

iii.

Question 4

a.

b.

c.

d.