Solutions to Week 5 Exercise Questions: Maximum Likelihood Estimation and Logistic Regression

March 16, 2023

1. (2021 AI2 Exam question) As a data scientist in a telecommunication company, your task is to analyse a customer dataset to predict whether a customer will terminate his/her contract. The dataset consists of around 8000 customer records, each consisting of one binary dependent variable Y, indicating whether the customer terminates the contract ($Y = 1$) or not ($Y = 0$), and 19 independent variables, which include the customer's information, e.g., age, subscription plan, extra data plan, etc., and the consumer behaviour such as average numbers of calls and hours per week. Since your boss needs some actionable insights to retain customers, you decided to use interpretable machine learning methods. Design your interpretable machine learning method by answering the following questions:

   (a) You have implemented a feature selection algorithm based on mutual information to select the most informative features from the 19 independent variables. To validate the implementation of your mutual information calculation function, you use a small subset of the data to calculate mutual information manually. You select one independent variable 'subscription plan', denoted as S, which takes two values, $S \in \{1, 2\}$. Please use the following Probability Mass Function table

$p(S,Y)$	$S=1$	$S=2$
$Y=0$	$\frac{2}{12}$	$\frac{5}{12}$
$Y=1$	$\frac{2}{12}$	$\frac{3}{12}$

   to calculate

   - Entropies $H(S)$ and $H(Y)$
   - Conditional entropies $H(S|Y)$ and $H(Y|S)$
   - Joint entropy $H(S,Y)$
   - Mutual information $I(S;Y)$

   Show all your working. Discuss what mutual information means and whether this feature will be selected or not.

   (b) After applying your algorithm you selected two variables: 1) extra data plan E, which is a binary random variable that indicates whether the customer subscribes to the extra data plan ($E = 1$) or not ($E = 0$); and 2) averaged hours used per week H, which is a continuous random variable. You then built a logistic regression model to classify customers into 'low risk' or 'high risk' of terminating the contract. The fitted model is

$$\log\left(\frac{p}{1-p}\right) = -0.77 + 0.23H - 1.18E$$

- Given a customer who has the extra data plan ($E = 1$) and spent on average 0.5 hours per week, calculate the odds and the probability the customer will terminate the contract ($Y = 1$). (**4 marks**)
- Using this fitted model, explain to your boss what actions should be taken to retain customers. (**10 marks**)

Answer:

(c) Question a: **We will solve this next week.**

Question b: **First question**:

- Odds:

$$\begin{aligned} odds = \frac{p}{1-p} &= \exp(-0.77 + 0.23H - 1.18E) = \exp(-0.77 + 0.115 - 1.18) \\ &= 0.1596 \end{aligned} \tag{1}$$

- Probability: (Note: $p = \frac{1}{1+\frac{1}{odds}}$)

$$P(Y = 1) = \frac{1}{1 + \frac{1}{0.1596}} = 0.137 \tag{2}$$

Second question: We need to extend the logistic model with one independent variable as learned in the lecture to two independent variables. We then investigate the effect or each variable by deriving the odd ratios by fixing the value of the other variable:

$$OR_E = \frac{\text{odds when } E = 1}{\text{odds when } E = 0} = \frac{\exp(-0.77 + 0.23 - 1.18)}{\exp(-0.77 + 0.23)} = \exp(-1.18) \approx 0.31$$

$$\begin{aligned} OR_H &= \frac{\text{odds when } H = h + \Delta}{\text{odds when } H = h} & (3) \\ &= \frac{\exp(-0.77 + 0.23(h + \Delta) - 1.18E)}{\exp(-0.77 + 0.23h - 1.18E))} & (4) \\ &= \frac{\exp(-0.77)\exp(0.23h)\exp(0.23\Delta)\exp(-1.18E)}{\exp(-0.77)\exp(0.23h)\exp(-1.18E)))} & (5) \\ &= (\exp(0.23))^{\Delta} & (6) \\ &\approx 1.25^{\Delta} & (7) \end{aligned}$$

The following key points should be mentioned:

- If a customer add the extra data plan, the odds of terminating the contract will increase is 0.31, which means that odds the customer terminating the contract will decrease by a factor of 3.
- If a customer increase the average time by one hour, the odds of terminating the contract increase by a factor of 1.25.

From the analysis, we can suggest to the boss that, the more hours the customers spent, the more likely the customers will terminate, which means the company should improve its telecommunication service/price. However, by simply persuade them to subscribe to the extra data plan, they are more likely to stay.

2. We flipped a coin 100 times. Given that there were 55 heads, find the maximum likelihood estimate for the probability p of heads on a single toss.

Answer:

Step 1 Write down the likelihood function. First, we need to determine the probability distribution function. Since counting the number of heads in 100 tosses is an experiment with 100 trials, therefore, we should use binomial distribution. The likelihood function, or the probability of 55 heads given that the probability of heads on a single toss is p, can be written:

$$P(55 \text{ heads}|p) = \binom{100}{55} p^{55}(1-p)^{45}$$

Explanation:

- **Experiments**: flip the coin 100 times and count the number of heads
- **Data**: The data is the result of the experiments, i.e., '55 heads'
- **Parameter(s)**: we want to estimate the value of the unknown parameter p
- **Likelihood or likelihood function**: $P(\text{data} \mid \text{p})$
- $\binom{100}{55}$: the binomial coefficient, which is called "Combination" and can be calculated using
$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

Step 2 Write down the log likelihood function:

$$\log(P(55 \text{ heads}|p)) = \log\left(\binom{100}{55} p^{55}(1-p)^{45}\right) \tag{8}$$
$$= \log\left(\binom{100}{55}\right) + 55\log(p) + 45\log(1-p) \tag{9}$$
$$\tag{10}$$

Step 3 Take the derivative of the log likelihood function with respect to the unknown parameter p and set it to 0

$$\frac{d}{dp}(\text{log likelihood function}) = \frac{d}{dp}\left[\log\left(\binom{100}{55}\right) + 55\log(p) + 45\ln(1-p)\right] \tag{11}$$

$$= \frac{55}{p} - \frac{45}{1-p} = 0 \tag{12}$$

$$\Rightarrow 55(1-p) = 45p \Rightarrow \hat{p} = 0.55$$

3. Let X be independent and identically distributed (i.i.d.) Poisson (λ) distributed. Find the maximum likelihood estimator for λ, $\hat{\lambda}$. Calculate an estimate using this estimator when $x_1 = 1$, $x_2 = 2$, $x_3 = 5$, $x_4 = 3$.

   Answer We can calculate the maximum likelihood estimator for λ, $\hat{\lambda}$ as follows

Step 1 Poisson PMF: We first write the probability density function of the Poisson distribution:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Step 2 Write the likelihood function:

$$L(\lambda|x_1, \ldots, x_n) = \prod_{j=1}^{n} \frac{\lambda^{x_j} e^{-\lambda}}{x_j!}$$

Step 3 By taking the natural logarithm on both sides, we have

$$\log L(\lambda|\boldsymbol{x}) = \log\left(\prod_{j=1}^{n} \frac{\lambda^{x_j} e^{-\lambda}}{x_j!}\right) \tag{13}$$

$$= \sum_{j=1}^{n} \log\left(\frac{\lambda^{x_j} e^{-\lambda}}{x_j!}\right) \tag{14}$$

$$= \sum_{j=1}^{n} \left[\log(\lambda^{x_j} e^{-\lambda}) - \log(x_j!)\right] \tag{15}$$

$$= \sum_{j=1}^{n} \left[\log(\lambda^{x_j}) + \log(e^{-\lambda}) - \log(x_j!)\right] \tag{16}$$

$$= \sum_{j=1}^{n} \left[x_j \log(\lambda) - \lambda - \log(x_j!)\right] \tag{17}$$

$$= -n\lambda + \log(\lambda)\sum_{j=1}^{n} x_j - \sum_{j=1}^{n} \log(x_j!) \tag{18}$$

Step 4 Calculate the derivative of the log likelihood function with respect to

λ:

$$\frac{d}{d\lambda}\log L(\lambda|\boldsymbol{x}) = \frac{d}{d\lambda}\left(-n\lambda + \log(\lambda)\sum_{j=1}^{n} x_j - \sum_{j=1}^{n}\log(x_j!)\right) \tag{19}$$

$$= -n + \frac{1}{\lambda}\sum_{j=1}^{n} x_j \tag{20}$$

Step 5 Set the derivative $\frac{d}{d\lambda}\log L(\lambda|\boldsymbol{x}) = 0$

$$-n + \frac{1}{\lambda}\sum_{j=1}^{n} x_j = 0 \tag{21}$$

which yield

$$\lambda = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{22}$$

For the given data, the estimate $\hat{\lambda}$ is

$$\hat{\lambda} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{4}(1+2+5+3) = 2.75$$