INT 3405: Machine Learning	**First Semester 2023-2024**
Week 2	
Student Name: Le Hoang Lan	*UID: 22028013*

- Feel free to talk to other students in the class when doing the homework. You should, however, write down your solution yourself. You also must indicate on each homework with whom you collaborated and cite any other sources you use including Internet sites.
- You will write your solution in LaTeX and submit the pdf file in zip files, including relevant materials, through courses.uet.vnu.edu.vn
- Dont be late.

For both exercises, $L(\theta)$ is the likelihood function of θ with respect to D.

1 Homework 1 - 10pts

Extend to Binomial Distribution

We have the probability of the data D is $P(D)$.

$$L(\theta) = P(D) = \prod_{i=1}^{n} P(s_i) = \prod_{i=1}^{n} C_N^{s_i}\theta^{s_i}(1-\theta)^{N-s_i}$$

$$l(\theta) = logL(\theta) = \sum_{i=1}^{n} logC_N^{s_i} + \sum_{i=1}^{n} log\theta^{s_i} + \sum_{i=1}^{n} log(1-\theta)^{N-s_i}$$

$$\frac{\partial l}{\partial \theta} = \frac{\sum_{i=1}^{N} s_i}{\theta} - \frac{\sum_{i=1}^{N}(N-s_i)}{1-\theta} = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{N} s_i}{\theta} = \frac{\sum_{i=1}^{N}(N-s_i)}{1-\theta}$$

$$\Rightarrow \sum_{i=1}^{n} s_i = \theta N n$$

$$\Rightarrow \theta = \frac{\sum_{i=1}^{n} s_i}{nN}$$

Hence, $\theta^{MLE} = \frac{\sum_{i=1}^{n} s_i}{nN}$

2 Homework 2 - 10pts

Extend to Categorical Distribution

We have the probability of the data D is $P(D)$, the number of x_i equals to k is N_k, and $\sum_{k=1}^{K} \theta_k = 1$. It can be seen that $\sum_{k=1}^{K} N_k = n$.

$[x_i = k]$ is the Iverson bracket. https://en.wikipedia.org/wiki/Iverson$_b racket$

$$L(\theta) = P(D) = \prod_{i=1}^{n} P(x_i) = \prod_{i=1}^{n} \prod_{k=1}^{K} \theta_k^{[x_i=k]}$$
$$l(\theta) = logL(\theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} log(\theta_k^{[x_i=k]})$$

Let $g(\theta) = \sum_{k=1}^{K} \theta_k - 1$. Using Lagrange multiplier for $l(\theta)$:

$$l(\theta, \lambda) = \sum_{i=1}^{n} \sum_{k=1}^{K} log(\theta_k^{[x_i=k]}) + \lambda g(\theta)$$
$$= \sum_{i=1}^{n} \sum_{k=1}^{K} log(\theta_k^{[x_i=k]}) + \lambda(\sum_{k=1}^{K} \theta_k - 1)$$

Calculating the derivative of each θ_k and λ:

$$\frac{\partial l}{\partial \lambda} = \sum_{k=1}^{K} \theta_k - 1 = 0$$
$$\frac{\partial l}{\partial \theta_k} = \frac{\sum_{i=1}^{n} [x_i = k]}{\theta_k} + \lambda = 0$$
$$\Rightarrow -\frac{\sum_{i=1}^{n} [x_i = k]}{\theta_k} = \lambda$$
$$\Rightarrow -\frac{N_k}{\theta_k} = \lambda$$

We can gather that

$$\frac{N_1}{\theta_1} = \frac{N_2}{\theta_2} = ... = \frac{N_K}{\theta_K} = -\lambda$$
$$\Rightarrow \frac{N_k}{\theta_k} = \frac{\sum_{k=1}^{K} N_k}{\sum_{k=1}^{K} \theta_k} \Rightarrow \frac{N_k}{\theta_k} = \frac{n}{1} \Rightarrow \theta_k = \frac{N_k}{n}$$

Hence, $\theta_k^{MLE} = \frac{N_k}{n}$