Exercise 1:

The implementations and answers are in "code_part.ipynb".

Exercise 2

1. Likelihood function is : $\prod_{i=1}^{n} Pr(Y_i=y_i|x_i) = \prod_{i=1}^{n} \frac{\mu_i^{y_i}}{y_i!} \exp(-\mu_i)$

Log-likelihood function is : $\log \prod_{i=1}^{n} Pr(Y_i=y_i|x_i)$

$$= \log\left[\prod_{i=1}^{n} \frac{\mu_i^{y_i}}{y_i!} \exp(-\mu_i)\right]$$

$$= \sum_{i=1}^{n} \log\left[\frac{\mu_i^{y_i}}{y_i!} \exp(-\mu_i)\right]$$

$$= \sum_{i=1}^{n} \log(\mu_i^{y_i}) - \log(y_i!) + \log[\exp(-\mu_i)]$$

$$= \sum_{i=1}^{n} \left[y_i \log(\mu_i) - \log(y_i!) - \mu_i\right]$$

2. Since $\mu_i \geq 0$ and $w^T x_i + b \in \mathbb{R}$,

We need to transform domain of μ_i to $\mathbb{R}$ i.e. $\mu_i \in \mathbb{R}$.

Log Transformation can take input >0 and output the real number,

So after using log transformation. $\log(\mu_i) \in \mathbb{R}$.

Therefore, the parameterization $\log \mu_i = w^T x_i + b$ is reasonable.

3. Since $\log \mu_i = w^T x_i + b$, $\mu_i = \exp(w^T x_i + b)$.

Simplifying log likelihood function :

$$\sum_{i=1}^{n} \left[y_i \log(\mu_i) - \log(y_i!) - \mu_i\right]$$

$$= \sum_{i=1}^{n} \left[y_i \log(\exp(w^T x_i + b)) - \log(y_i!) - \mu_i\right] \text{ as } \mu_i = \exp(w^T x_i + b)$$

$$= \sum_{i=1}^{n} \left[y_i (w^T x_i + b) - \log(y_i)! - \exp(w^T x_i + b)\right]$$

Since $\log(y_i!)$ is a constant, we drop it.

We want to maximize log likelihood function,

equivalently is to maximize $\sum_{i=1}^{n} \left[y_i (w^T x_i + b) - \exp(w^T x_i + b)\right]$

with respect to variables w, b.

So, $$(w, b) = \underset{w \in \mathbb{R}^d, b \in \mathbb{R}}{\arg\max} \underbrace{\sum_{i=1}^{n} \left[y_i (w^T x_i + b) - \exp(w^T x_i + b)\right]}_{\text{objective function.}}$$

Therefore, the objective function is the $\sum_{i=1}^{n}\left[y_i\,(w^T x_i + b) - \exp(w^T x_i + b)\right]$

the optimization variables are w and b, and

it's a maximization problem.

4. Let $f(w,b) = \sum_{i=1}^{n}\left[y_i\,(w^T x_i + b) - \exp(w^T x_i + b)\right]$

$$\frac{\partial f}{\partial w} = \sum_{i=1}^{n}\left[y_i \frac{\partial (w^T x_i + b)}{\partial w} - \frac{\partial(\exp(w^T x_i + b))}{\partial w}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \cdot x_i - \exp(w^T x_i + b)\,x_i\right]$$

$$= \sum_{i=1}^{n}\left[(y_i - \exp(w^T x_i + b))\,x_i\right]$$

$$\frac{\partial f}{\partial b} = \sum_{i=1}^{n}\left[y_i \frac{\partial (w^T x_i + b)}{\partial b} - \frac{\partial(\exp(w^T x_i + b))}{\partial b}\right]$$

$$= \sum_{i=1}^{n}\left[y_i - \exp(w^T x_i + b)\right]$$

Gradient Algorithm:

Input: $X \in \mathbb{R}^{n \times d}$, $y \in \mathbb{R}^n$, $\vec{w_0} = \vec{0_d}$, $b_0 = 0$, $max\_pass \in \mathbb{N}$, $\eta > 0$, $tol > 0$

Output: $w \in \mathbb{R}^d$, $b \in \mathbb{R}$

for $t = 1, 2, 3, \ldots, max\_pass$:

$$\nabla w_{t-1} = \sum_{i=1}^{n}\left[(y_i - \exp(w_{t-1}^T x_i + b_{t-1}))\,x_i\right]$$

$$\nabla b_{t-1} = \sum_{i=1}^{n}\left[y_i - \exp(w_{t-1}^T x_i + b_{t-1})\right]$$

$$w_t = w_{t-1} - \eta \nabla w_{t-1}$$

$$b_t = b_{t-1} - \eta \nabla b_{t-1}$$

if $\|w_t - w_{t-1}\| \le tol$: break

$w = w_t$

$b = b_t$

return w, b