Question 5

1 Given $D = \{x_i, y_i\}_{i=1}^{n}$, what is the log likelihood function (of μ_i's) given D?

Given $P_r(Y_i = k \mid X_i) = \frac{\mu_i^k}{k!} e^{-\mu_i}$

Let us consider a datapoint $\{(x_i, y_i) \in D$

$$\text{Then } P(Y_i = y_i \mid x_i) = \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i}$$

Finding log ~~likelihood~~ $\Rightarrow$ $= \log\left(\frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i}\right)$

$$= \log(\mu_i^{y_i}) + \log(e^{-\mu_i}) - \log(y_i!)$$

$$= y_i \log(\mu_i) - \mu_i - \log(y_i!)$$

This function has parameter μ_i and observable data y_i. $\therefore$ Log likelihood of the entire dataset is a summation of the log of each individual datapoint

$$\therefore f(\mu_i) = \sum_{i=1}^{n} (y_i \log(\mu_i) - \mu_i - \log(y_i!))$$

2 Give justification of $\log \mu_i = w^T x_i + b$

The parameterization of μ_i is justified due to the fact that μ_i is the mean of the Poisson Distribution because of which it must always be ≥ 0.

$\therefore$ $\log \mu_i$ is valid because μ_i will never be negative while $\log(\mu_i)$ ~~log~~ can be negative

As such μ_i is always a valid ~~mean~~ mean and parameter as long as
$\log(\mu_i) = w^T x_i + b$

3. Write down the objective function for Poisson regression

We aim to maximize log likelihood of data and we learned $\log(\mu_i) = w^T x_i + b$ is a valid parameter

$\therefore$ Objective function $= \sum_{i=1}^{n} (y_i \log(\mu_i) - \mu_i - \log(y_i!))$

$= \sum_{i=1}^{n} (y_i(w^T x_i + b) - e^{w^T x_i + b} - \log(y_i!))$

We are trying to maximize w.r.t. w & b.

4. Compute gradient

$\cancel{\frac{\partial}{\partial w} \sum_{i=1}^{n} (y_i \log (}$

With respect to w

$$\nabla f_w = \frac{\partial}{\partial w} \sum_{i=1}^{n} y_i w^T x_i + y_i b - e^{w^T x_i} e^{b} - \log(y_i!)$$

$$= \sum_{i=1}^{n} (y_i x_i - x_i e^{w^T x_i + b})$$

With respect to b

$$\nabla f_b = \frac{\partial}{\partial b} \sum_{i=1}^{n} y_i w^T x_i + y_i b - e^{b} e^{w^T x} - \log(y_i!)$$

$$= \sum_{i=1}^{n} (y_i - e^{w^T x + b})$$

To maximize objective function, we need to constant update w and b using gradient descent similar to Question 1 part 3 of the assignment

$$\eta = \text{step size}$$

$$w_{new} = w_{old} - \eta \nabla f_w$$

$$= w_{old} - \eta \sum_{i=1}^{n} \left(y_i x_i - x_i e^{w^T x + b}\right)$$

$$b_{new} = b_{old} - \eta \nabla f_b$$

$$= b_{old} - \eta \sum_{i=1}^{n} \left(y_i - e^{(w^T x_i + b)}\right)$$

Repeat untile w & b converge to a maximum value