Let $\begin{cases} y_i \text{ be the response of a datum, } \vec{y} = [y_1, y_2, \ldots, y_n]^T \\ \overrightarrow{X_k} = [1, x_{k1}, x_{k2}, \ldots, x_{kp}] \text{ be a datum, } X = [\overrightarrow{X_1}, \overrightarrow{X_2}, \ldots, \overrightarrow{X_n}]^T \\ \vec{\beta} = [\beta_0, \beta_1, \ldots, \beta_p]^T \text{ be the coefficient vector} \\ \lambda_k = \lambda(\overrightarrow{X_k}) = e^{\beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \ldots + \beta_p x_{kp}} = e^{\overrightarrow{X_k} \cdot \vec{\beta}} \rightarrow \nabla_\beta \lambda_k = \overrightarrow{X_k} e^{\overrightarrow{X_k} \cdot \vec{\beta}} = \overrightarrow{X_k} \lambda_k \\ \vec{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]^T \\ P_k = P[Y = y_k] = \frac{e^{-\lambda_k} \lambda_k^{y_k}}{y_k!} \end{cases}$

Likelihood function l

$$l = \prod_{i=1}^{n} P_i = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

$$\rightarrow ln(l) = \sum_{i=1}^{n} ln(\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}) = \sum_{i=1}^{n} [y_i ln(\lambda_i) - \lambda_i - ln(y_i!)]$$

To maximize l is to maximize $ln(l) \rightarrow \nabla_\beta ln(l) = 0$, then

$$\begin{aligned} \nabla_\beta ln(l) &= \sum_{i=1}^{n} [y_i \nabla_\beta ln(\lambda_i) - \nabla_\beta \lambda_i] \\ &= \sum_{i=1}^{n} (\frac{y_i}{\lambda_i} \nabla_\beta \lambda_i - \nabla_\beta \lambda_i) \\ &= \sum_{i=1}^{n} (\frac{y_i}{\lambda_i} - 1) \lambda_i \overrightarrow{X_i} \\ &= \sum_{i=1}^{n} (y_i - \lambda_i) \overrightarrow{X_i} \\ &= X^T \cdot (\vec{y} - \vec{\lambda}) \end{aligned}$$

With learning rate $\alpha > 0$ and proper number of iterations, we can fit β by

$$\beta_{new} = \beta_{old} + \alpha \nabla_\beta ln(l) = \beta_{old} + \alpha X^T \cdot (\vec{y} - \vec{\lambda})$$