確率変数 Y が実数 μ を用いて

$$P(Y=k)=\frac{\mu^k}{k!}e^{-\mu}$$

で表される分布に従うとき、Y はポアソン分布に従うという．このとき

$$\begin{aligned}E[Y]&=\sum_{i=0}^{\infty}k\cdot\frac{\mu^k}{k!}e^{-\mu}\\&=\sum_{i=1}^{i\infty}\frac{\mu\cdot\mu^{k-1}}{(k-1)!}e^{-\mu}\\&=\mu e^{-\mu}\sum_{l=0}^{\infty}\frac{\mu^l}{l!}=\mu\end{aligned}$$

となることから Y の期待値は μ になることがわかる．

ポアソン分布のパラメータ μ に関して，p を自然数とし $x\in\mathbb{R}^p$ に対してある $\beta_0\in\mathbb{R},\beta\in\mathbb{R}^p$ が存在して

$$\mu(x):=E[Y\mid X=x]=e^{\beta_0+x\beta}$$

とかけることを仮定する．この場合 N 個の観測値 $(x_1,y_1),\cdots,(x_N,y_N)$ $(x_i\in\mathbb{R}^p,y_i\in\mathbb{Z}_+)$ についての尤度関数は

$$\mu_i:=\mu(x_i)=e^{\beta_0+x_i\beta}$$

とすれば

$$\prod_{i=1}^{N}\frac{{\mu_i}^{y_i}}{y_i!}e^{-\mu_i}\tag{1}$$

とかくことができる．各 i に対して

$$\begin{aligned}\log\frac{{\mu_i}^{y_i}}{y_i!}e^{-\mu_i}&=y_i\log\mu_i+\log e^{-\mu_i}-\log(y_i!\,)\\&=y_i(\beta_0+x\beta)-e^{\beta_0+x\beta}-\log(y_i!\,)\end{aligned}$$

が成立していることより，(1) 式のマイナス対数をとると

$$
\begin{aligned}
-\log\left(\prod_{i=1}^{N}\frac{{\mu_i}^{y_i}}{y_i!}e^{-\mu_i}\right) &= -\sum_{i=1}^{N}\log\left(frac{\mu_i}^{y_i}y_i!e^{-\mu_i}\right) \\
&= -\sum_{i=1}^{N}\{y_i(\beta_0+x_i\beta)-e^{\beta_0+x_i\beta}-\log(y_i!\,)\} \\
&= -\sum_{i=1}^{N}\{y_i(\beta_0+x_i\beta)-e^{\beta_0+x_i\beta}\}+\sum_{i=1}^{N}\log(y_i!\,)
\end{aligned}
$$

となることがわかる．したがって (1) 式のマイナス対数の (β_0, β に関する) 最小化は上式の最終式第 1 項の最小化と同値になるのでこれに Lasso を適用することを考える．すなわち

$$
L(\beta_0,\beta) = -\frac{1}{N}\sum_{i=1}^{N}\{y_i(\beta_0+x_i\beta)-e^{\beta_0+x_i\beta}\}
$$

とし，これに正則化項を付けた

$$
L(\beta_0,\beta)+\lambda\|\beta\|
$$

の最小化について考える．

ロジスティック回帰の場合と同様にニュートン法を用いることを考える．そのために $x_i = (x_{i,1},\cdots,x_{i,p})^T$ とし，

$$
X = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,p} \end{pmatrix}
$$

としたうえで

$\nabla L = -\frac{1}{N}Xu, \nabla^2 L = \frac{1}{N}X^TWX$ を満たす $u \in \mathbb{R}^N, W \in \mathbb{R}^{N\times N}$ を求める．

L の $\beta_j (j = 0,\cdots,p)$ での偏微分を考えると，$x_{i,j}$ を $x_i \in \mathbb{R}^p$ の第 j 成分

として

$$\frac{\partial L}{\partial \beta_0} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - e^{\beta_0 + x_i\beta})$$

$$\begin{aligned}\frac{\partial L}{\partial \beta_j} &= -\frac{1}{N}\sum_{i=1}^{N}(x_{i,j} - x_{i,j}e^{\beta_0 + x_i\beta}) \\ &= -\frac{1}{N}\sum_{i=1}^{N}x_{i,j}(y_i - e^{\beta_0 + x_i\beta})\end{aligned}$$

と書くことができるので，∇L は

$$\begin{aligned}\nabla L &= -\frac{1}{N}\begin{pmatrix}\sum_{i=1}^{N}(y_1 - e^{\beta_0 + x_i\beta_i}) \\ \sum_{i=1}^{N}x_{i,1}(y_i - e^{\beta_0 + x_i\beta_i}) \\ \vdots \\ \sum_{i=1}^{N}x_{i,p}(y_i - e^{\beta_0 + x_i\beta_i})\end{pmatrix} \\ &= -\frac{1}{N}\begin{pmatrix}1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,p}\end{pmatrix}^T \begin{pmatrix}y_1 - e^{\beta_0 + x_1\beta} \\ \vdots \\ y_N - e^{\beta_0 + x_N\beta}\end{pmatrix}\end{aligned}$$

となり，したがって

$$u = \begin{pmatrix}y_1 - e^{\beta_0 + x_1\beta} \\ \vdots \\ y_N - e^{\beta_0 + x_N\beta}\end{pmatrix}$$

とすればよいことがわかる．

また L の二階微分を考えると，$x_{i,0} = 1(i = 1, \cdots, N)$ として

$$\frac{\partial L^2}{\partial \beta_j \beta_k} = \frac{1}{N}\sum_{i=1}^{N}x_{i,j}x_{i,k}e^{\beta_0 + x_i\beta} \quad (j, k \in \{0, 1, \cdots, p\})$$

となることより

$$
\begin{aligned}
\nabla^2 L &= \frac{1}{N}\begin{pmatrix} \sum_{i=1}^N x_{i,1}x_{i,1}e^{\beta_0+x_i\beta} & \cdots & \sum_{i=1}^N x_{i,1}x_{i,p}e^{\beta_0+x_i\beta} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^N x_{i,N}x_{i,1}e^{\beta_0+x_i\beta} & \cdots & \sum_{i=1}^N x_{i,N}x_{i,P}e^{\beta_0+x_i\beta} \end{pmatrix} \\
&= \frac{1}{N}\begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,p} \end{pmatrix}^T \begin{pmatrix} e^{\beta_0+x_1\beta} & e^{\beta_0+x_1\beta}x_{1,1} & \cdots & e^{\beta_0+x_1\beta}x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\beta_0+x_N\beta} & e^{\beta_0+x_N\beta}x_{N,1} & \cdots & e^{\beta_0+x_N\beta}x_{N,p} \end{pmatrix} \\
&= \frac{1}{N}\begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,p} \end{pmatrix}^T \begin{pmatrix} e^{\beta_0+x_1\beta} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\beta_0+x_N\beta} \end{pmatrix} \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,p} \end{pmatrix}
\end{aligned}
$$

となり，したがって

$$
W = \begin{pmatrix} e^{\beta_0+x_1\beta} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\beta_0+x_N\beta} \end{pmatrix}
$$

とすればよいことがわかる．