問題 1. (a) 目的関数

$$L(\beta_0, \beta) := \sum_{i=1}^{N} \log\{1 + \exp(-y_i(\beta_0 + x_i^T \beta))\}$$

を $\beta_j (j = 0, \cdots, p)$ で偏微分するとき，$x_{i,0} = 1 (i = 1, \cdots, N), v_i := \exp(-y_i(\beta_0 + x_i^T \beta))$ と置くことで

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

とかけることより

$$\nabla L = -\begin{pmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{N,1} \\ \vdots & \ddots & \vdots \\ x_{1,p} & \cdots & x_{N,p} \end{pmatrix} \begin{pmatrix} \frac{y_1 v_i}{1+v_1} \\ \vdots \\ \frac{y_1 v_N}{1+v_N} \end{pmatrix} = -X^T u$$

と書くことができる．

(b) (a) で求めた $\frac{\partial L}{\partial \beta_j}$ をさらに β_k で偏微分すると

$$\begin{aligned}\frac{\partial^2 L}{\partial \beta_j \partial \beta_k} &= -\sum_{i=1}^{N} \left\{ \frac{\partial v_i}{\partial \beta_k} \frac{\partial}{\partial v_i} \left(\frac{y_i x_{i,j} v_i}{1 + v_i} \right) \right\} \\ &= \sum_{i=1}^{N} y_i^2 x_{i,j} x_{i,k} \frac{v_i}{(1 + v_i)^2} \\ &= \sum_{i=1}^{N} x_{i,j} x_{i,k} \frac{v_i}{(1 + v_i)^2} \quad (\because y_i \in \{\pm 1\})\end{aligned}$$

となることより，

$$\begin{aligned}\nabla^2 L &= X^T \begin{pmatrix} \frac{v_1}{(1+v_1)^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{v_N}{(1+v_N)^2} \end{pmatrix} X \\ &= X^T W X\end{aligned}$$

と書くことができる．

さらに，上記で求めた行列を用いて作成した，Newton 法により値を更新していくプログラムの実行結果は下の通り．最上段が初期値 ($p+1$ 次元) を，最下段が推定された係数と実際の係数の $l2$ ノルムを表す

```
#p=2
[ 0.63094861 -0.92697651 -0.91072333]
[ 0.86047835 -1.14017869 -1.35275858]
[ 0.95114002 -1.22561286 -1.50788179]
[ 0.96108609 -1.235 -1.52382468]
0.7609464718257709

#p=3
[-0.39248015 -0.60068543 -0.6049099 -1.33818295]
[-0.46124804 -0.72404346 -0.72430081 -1.87481017]
[-0.48976562 -0.77927907 -0.771821 -2.06096778]
[-0.49271985 -0.78517205 -0.77653638 -2.0784206 ]
0.7252842383848371

#p=4
[ 0.08141478 -0.68484112 0.70987333 -1.50021169 0.0729047 ]
[ 0.0575044 -0.92584186 1.0212524 -2.04351767 0.1381232 ]
[ 0.05396007 -1.0284011 1.15461278 -2.28043797 0.16276137]
[ 0.05397496 -1.04263046 1.17324641 -2.31351461 0.16560318]
[ 0.05398124 -1.04286421 1.17355273 -2.31405542 0.16564041]
1.183505858275822

#p=5
[-0.44735853 0.1753804 0.74519156 -0.75696786 0.95888674 0.52116991]
[-0.51760096 -0.84462375 0.78877956 -1.1092826 0.70638379 -0.42738666]
[-0.74273836 -0.84191146 1.08434294 -1.39660604 1.10210054 -0.30626662]
[-0.81581402 -0.91371154 1.18648688 -1.52188011 1.2140128 -0.32811659]
[-0.82315975 -0.9212835 1.19631503 -1.53466277 1.22504241 -0.33062673]
0.8536261412241056

#p=6
[-0.30254961 -0.65441239 -0.7671653 0.89556886 0.82291769 -1.8986078
  1.09200474]
[-0.93939836 -0.85281913 0.71360078 0.24822987 0.12968842 0.23661314
  0.79192553]
[-0.78674208 -0.88892596 0.19398896 0.61617608 0.47178312 -0.73113428
```

```
  1.06501684]
[-0.98383111 -1.13192227 0.26832396 0.75029806 0.56505247 -0.8264384
  1.33542463]
[-1.03257514 -1.1857139 0.28516761 0.78091463 0.59207949 -0.84812315
  1.40043966]
[-1.03483699 -1.18803046 0.28591791 0.7822523 0.59338503 -0.8490672
  1.40340034]
1.0691750340249508
```

p を大きくしていくごとに，シミュレーションの計算が発散してしまう回数が増えていった． □

問題 2. (a) $\gamma_0 \in \mathbb{R}, \gamma \in \mathbb{R}^p$ を任意にとって固定し，(2.25) 式の指数部全てから $\gamma_0 + x^T\gamma$ を引いた式を変形していくと

$$\begin{aligned}
&\frac{\exp(\beta_{0,k} + x^T\beta^{(k)} - (\gamma_0 + x^T\gamma))}{\sum_{l=1}^K \exp(\beta_{0,k} + x^T\beta^{(l)} - (\gamma_0 + x^T\gamma))} \\
&\qquad = \frac{\exp(-(\gamma_0 + x^T\gamma))\exp(\beta_{0,k} + x^T\beta^{(k)})}{\exp(-(\gamma_0 + x^T\gamma))\sum_{l=1}^K \exp(\beta_{0,k} + x^T\beta^{(l)})} \\
&\qquad = \frac{\exp(\beta_{0,k} + x^T\beta^{(k)})}{\sum_{l=1}^K \exp(\beta_{0,k} + x^T\beta^{(l)})} = P(Y = k \mid x) \\
&\therefore P(Y = k \mid x) = \frac{\exp(\beta_{0,k} + x^T\beta^{(k)} - (\gamma_0 + x^T\gamma))}{\sum_{l=1}^K \exp(\beta_{0,k} + x^T\beta^{(l)} - (\gamma_0 + x^T\gamma))}
\end{aligned}$$

となり，等号が得られることがわかる．

(b) $\beta_{j,1}, \cdots, \beta_{j,K}$ の値が求まったとき $(j = 1, \cdots, p)$, $\displaystyle\sum_{k=1}^K |\beta_{j,k} - \gamma_j|$ の値を最小にする $\gamma_j \in \mathbb{R}$ を求めればよい．またそのような γ_j は $\beta_{j,1}, \cdots, \beta_{j,K}$ の中央値になると文中で述べられているので，$\beta_{j,1}, \cdots, \beta_{j,K}$ を求めた後にそれらすべてから中央値を引けばよいことがわかる．

(c) $\displaystyle\sum_{k=1}^K \beta_{0,k}$ となるように設定されていることから，最初に $\beta_{0,1}, \cdots, \beta_{0,K}$ の値を求めた後，それらすべてから $\bar{\beta} := \dfrac{1}{K}\displaystyle\sum_{k=1}^K \beta_{0,k}$ を引いていることがわかる． □

問題 3. (a) (2.26) 式は観測値 $(x_1, y_1), \cdots, (x_N, y_N)$ から得られる尤度関数で

$$\prod_{i=1}^{N} \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} \quad (\mu_i = e^{\beta_0 + x_i^T \beta})$$

となっている．この尤度関数のマイナス対数をとると

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

となる．最下段の β_0, β に関する最小化は $\frac{1}{N} L(\beta_0, \beta)$ の β_0, β に関する最小化と同値であり，この式に正則化項をつけると

$$\frac{1}{N} L(\beta_0, \beta) + \lambda \|\beta\|_1$$

となる．このようにして (2.27) 式が導出される．

(b) $L(\beta_0, \beta) = \sum_{i=1}^{N} \{y_i(\beta_0 + x_i^T \beta) - e^{\beta_0 + x_i^T \beta}\}$ を $\beta_j (j = 0, \cdots, p)$ で微分すると

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

となる．ただし $x_{i,0} = 1 (i = 1, \cdots, N)$ とした．よって

$$\nabla L = -\begin{pmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{N,1} \\ \vdots & \ddots & \vdots \\ x_{1,p} & \cdots & x_{N,p} \end{pmatrix} \begin{pmatrix} y_1 - e^{\beta_0 + x_1^T \beta} \\ \vdots \\ y_N - e^{\beta_0 + x_N^T \beta} \end{pmatrix}$$

となるので，

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

とすれば $\nabla L = -X^T u$ とかくことができる．

(c) (b) で得られた $\frac{\partial L}{\partial \beta_j}$ をさらに β_k で偏微分すると

$$\frac{\partial}{\partial \beta_k}\left(\frac{\partial L}{\partial \beta_j}\right) = -\sum_{i=1}^{N} x_{i,j}x_{i,k}e^{\beta_0 + x_i^T\beta}$$

となることより

$$\nabla^2 L = X^T \begin{pmatrix} e^{\beta_0 + x_1^T\beta} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\beta_0 + x_N^T\beta} \end{pmatrix} X$$

となるので,

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

とすれば $\nabla^2 L = X^T W X$ と書くことができる. □

また掲載されているプログラムの空欄を埋めてコード，およびそれを実行した結果は下のようになる

```
def poisson_lasso(X, y, lam):
    p = X.shape[1] # pはすべて1の列を含んでいる
    beta = np.random.randn(p)
    gamma = np.random.randn(p)
    while np.sum((beta - gamma) ** 2) > 0.0001:
        beta = gamma
        s = np.dot(X, beta)
        w = np.exp(s) #空欄1
        u = y - w #空欄2
        z = s + u / w ##空欄3
        gamma_0, gamma_1 = W_linear_lasso(X[:, range(1, p)],
                                          z, np.diag(w), lam)
        gamma = np.block([gamma_0, gamma_1]).copy()
        print(gamma)
    return gamma

N = 100
p = 3
X = np.random.randn(N, p)
X = np.concatenate([np.ones(N).reshape(N, 1), X], axis=1)
beta = np.random.randn(p + 1)
```

```
s = np.dot(X, beta)
y = np.random.poisson(lam=np.exp(s))
print(beta)
```

```
[-0.47085992 -0.31067144 0.30882489 -0.00316935]  #真のbの値
[ 0.7654697 0. -0. -0. ]
[ 0.05384126 -0. 0. 0. ]
[-0.35865758 -0. 0. 0. ]
[-0.4711855 -0. 0. 0. ]
[-0.47801239 -0. 0. 0. ]#推定されたb
```

推定された β がスパースなものであることが確認できた． □

問題 4. (a) 関数 L は

$$L(\beta) := -\sum_{i:\delta_i=1} \log \frac{e^{x_i^T\beta}}{\sum_{j\in R_i} e^{x_j^T\beta}}$$

で定義されている．この関数 L を β_k で偏微分すると

$$\begin{aligned}
\frac{\partial L}{\partial \beta_k} &= -\sum_{i:\delta_i=1}\left\{x_{i,k} - \frac{\sum_{j\in R_i x_{j,k}} \exp(x_j^T\beta)}{\sum_{h\in R_i}\exp(x_h^T\beta)}\right\} \\
&= -\sum_{i=1}^{N} x_{i,k}\left\{\delta_i - \sum_{j\in C_i}\frac{\exp(x_i^T\beta)}{\sum_{h\in R_j}\exp(x_h^T\beta)}\right\}
\end{aligned}$$

とかくことができる．ここで，$S_i\alpha = \sum_{h\in R_i}\exp(x_h^T\beta)$ としたときに

$$\begin{aligned}
\sum_{i;\delta_i=1}\sum_{j\in R_j}\frac{x_{j,k}\exp(x_j^T\beta)}{S_i} &= \sum_{i=1}^{N}\sum_{i\in C_i}\frac{x_{j,k}\exp(x_j^T\beta)}{S_i} \\
&= \sum_{i=1}^{N} x_{i,k}\sum_{j\in C_i}\frac{\exp(x_i^T\beta)}{S_j}
\end{aligned}$$

と変形できることより

$$\frac{\partial L}{\partial \beta_k} = -\sum_{i=1}^{N} x_{i,k}\left\{\delta_i - \sum_{j\in C_i}\frac{\exp(x_i^T\beta)}{\sum_{h\in R_j}\exp(x_h^T\beta)}\right\}$$

となることがわかる．したがって

$$\nabla L = -X^T \begin{pmatrix} \delta_1 - \sum_{j \in C_1} \frac{\exp(x_1^T \beta)}{\sum_{h \in R_j} \exp(x_h^T \beta)} \\ \vdots \\ \delta_N - \sum_{j \in C_N} \frac{\exp(x_1^T \beta)}{\sum_{h \in R_j} \exp(x_h^T \beta)} \end{pmatrix}$$

となることより，求める u は

$$u = \begin{pmatrix} \delta_1 - \sum_{j \in C_1} \frac{\exp(x_1^T \beta)}{\sum_{h \in R_j} \exp(x_h^T \beta)} \\ \vdots \\ \delta_N - \sum_{j \in C_N} \frac{\exp(x_1^T \beta)}{\sum_{h \in R_j} \exp(x_h^T \beta)} \end{pmatrix}$$

とかける．

(b) $\nabla^2 L$ の各成分を変形していくと以下のようになる

$$\begin{aligned} \frac{\partial L}{\partial \beta_k \partial \beta_l} &= \sum_{i=1}^{N} \sum_{j \in C_i} \frac{\partial}{\partial \beta_l} \left(\frac{\exp(x_i^T \beta)}{\sum_{h \in R_j} \exp(x_h^T \beta)} \right) \\ &= \sum_{i=1}^{N} x_{i,k} \sum_{j \in C_i} \frac{1}{(\sum_{r \in R_j} \exp(x_r^T \beta))^2} \Bigg\{ x_{i,l} \exp(x_i^T \beta) \sum_{s \in R_j} \exp(x_s^T \beta) \\ &\qquad - \exp(x_i^T \beta) \sum_{h \in R_j} x_{h,l} \exp(x_h^T \beta) \Bigg\} \\ &= \sum_{i=1}^{N} \sum_{h=1}^{N} x_{i,k} x_{h,l} \sum_{j \in C_i} \frac{\exp(x_i^T \beta)}{(\sum_{t \in R_j} \exp(x_t^T \beta))^2} \left\{ I(i = h) \sum_{s \in R_j} \exp(x_s^T \beta) - I(h \in R_j) \exp(x_h^T \beta) \right\} \end{aligned}$$

とかける．ここで W の対角成分，すなわち $i = h$ となる成分は

$$w_i = \sum_{j \in C_i} \frac{\exp(x_i^T \beta)}{(\sum_{t \in R_j} \exp(x_t^T \beta))^2} \left\{ \sum_{s \in R_j} \exp(x_s^T \beta) - I(h \in R_j) \exp(x_i^T \beta) \right\}$$

とかくことができ，

$$\delta_i = 1, j \in R_i \Leftrightarrow i \in C_i$$

であることより $i = h$ ならば $j \in C_i \Leftrightarrow i = h \in R_j$ が得られる．したがって

$$\pi_{i,j} := \frac{\exp(x_i^T \beta)}{\sum_{r \in R_j} \exp(x_i^T \beta)}$$

とすることで，W の各対角成分 w_i は

$$w_i = \sum_{j \in C_i} \frac{\exp(x_i^T \beta)}{\sum_{r \in R_j} \exp(x_i^T \beta)} \left\{ 1 - \frac{\exp(x_i - T\beta)}{\sum_{r \in R_j} \exp(x_i^T \beta)} \right\} = \sum_{j \in C_i} \pi_{i,j}(1 - \pi_{i,j})$$

とかける． □