Solutions to Exam 1

Rongfei Jin

March 11, 2025

Problem 1

(a) See code above

(b) See code above

(c) (1) Gender

　　(i) Faster multiplication on 0

　　(ii) gender encoded in 0,1 is easiser to interpret because the coefficients represents the effect of being male, or no effect if being female

　(2) Income and travel

　　inc25p: 1 if income is greater than 25k, 0 otherwise
　　inc55p: 1 if income is greater than 55k, 0 otherwise
　　inc95p: 1 if income is greater than 95k, 0 otherwise
　　tra025p: 1 if travel is greater than .25h, 0 otherwise
　　tra400p: 1 if travel is greater than 4h, 0 otherwise

　　(i) Design matrix achieves full rank. These transformations solve the collinearity problem since if the model has all the condition except the last one, then the last one is determined, so numerically it is more stable.

　　(ii) Easier to interpret as the coefficients are simply the effect of income greater than a certain threshold

(d) The new model has 7 coefficients (without intercept) and NO coefficients are NA. Gender, income greater than 55k, travel time greater than 0.25h and travel time greater than 4h are significant.

Based on the significant coefficients, we can make the following interpretation.

　(i) Males are associated with 0.36 more visit than female

　(ii) Income greater than 55k is associated with 0.5 less visit than income less than 55k

　(iii) Travel time greater than 0.25h is associated with 0.6 less visit than travel time less than 0.25h

　(iv) Travel time greater than 4h is associated with 1.8 less visit than travel time less than 4h

(e) The possion regression has the following probability density function

$$p(y|\eta) = \frac{\eta^y e^{-\eta}}{y!}$$

where $\eta = \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p)$ is the mean of the possion distribution. Therefore, we have

$$\mathrm{E}(y|x_1, \ldots, x_p) = \eta = \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)$$

Given female, earning \$65,000 annually, and living two miles from the park. We have the following data gen = 0, inc25p = 1 inc55p = 1, inc95p = 0, tra025p = 0, tra400p = 0.

Therefore, we have

$$\mathrm{E}(y|x_1, \ldots, x_p) = \exp(\beta_0 + \beta_1 0 + \beta_2 1 + \beta_3 1 + \beta_4 0 + \beta_5 0 + \beta_6 0) = \exp(\beta_0 + \beta_2 + \beta_3) = 34.49$$

Given male, earning \$65,000 annually, and living two miles from the park. We have the following data gen $= 1$, inc25p $= 1$ inc55p $= 1$, inc95p $= 0$,tra025p $= 0$, tra400p $= 0$.

$$\mathrm{E}(y|x_1,\ldots,x_p) = \exp(\beta_0 + \beta_1 1 + \beta_2 1 + \beta_3 1 + \beta_4 0 + \beta_5 0) = \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3) = 49.11$$

$$\frac{E(y|\text{male with given conditions})}{E(y|\text{female with given conditions})} = \frac{49.11}{34.49} \approx 1.42$$

$$\frac{E(y|\text{female with given conditions})}{E(y|\text{male with given conditions})} = \frac{34.49}{49.11} \approx 0.703$$

Problem 2

(a) See code above

(b) (i) the center from both approaches are very close

    (ii) the bootstrapped CI width is wider than the normal CI, which is expected since the normal CI assumes the distribution is normal, but the bootstrapped CI does not make this assumption

Problem 3

(a) Since $Y \sim \text{Poisson}(\lambda)$ we have $P(Y = y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$

We compute the Momement Generating Function

$$
\begin{aligned}
M_Y(t) &= \mathrm{E}(e^{tY}) \\
&= \sum_{y=0}^{\infty} e^{ty} P(Y = y; \lambda) \\
&= \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y e^{-\lambda}}{y!} \\
&= e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!} \\
&= e^{-\lambda} e^{\lambda e^t} \qquad\qquad\qquad \text{Taylor expansion} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}
$$

Now we compute mean by deriving the first moment

$$
\begin{aligned}
\mathrm{E}(Y) &= M_Y'(0) \\
&= \frac{d}{dt} e^{\lambda(e^t - 1)} \Big|_{t=0} \\
&= \lambda e^{\lambda(e^0 - 1)} \\
&= \lambda
\end{aligned}
$$

Now we compute the variance by first deriving the second moment

$$
\begin{aligned}
\mathrm{E}(Y^2) &= M_Y''(0) \\
&= \frac{d^2}{dt^2} e^{\lambda(e^t - 1)} \Big|_{t=0} \\
&= \lambda e^{\lambda(e^0 - 1)} + \lambda^2 e^{\lambda(e^0 - 1)} \\
&= \lambda + \lambda^2
\end{aligned}
$$

Now we compute the variance

$$\text{Var}(Y) = \text{E}(Y^2) - \text{E}(Y)^2$$
$$= (\lambda + \lambda^2) - \lambda^2$$
$$= \lambda$$

(b)

$$p(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$$
$$= e^{\log(\frac{\lambda^y e^{-\lambda}}{y!})}$$
$$= e^{y\log(\lambda) - \lambda - \log(y!)}$$

(c) since $\log(\lambda) = \beta_0 + \boldsymbol{\beta} \cdot \mathbf{x} = \beta_0 + \boldsymbol{\beta}^T\mathbf{x}$, we have

$$p(y|\mathbf{x}, \beta_0, \boldsymbol{\beta}) = \exp\left\{[\beta_0 + \boldsymbol{\beta}^T\mathbf{x}]y - \exp\left\{\beta_0 + \boldsymbol{\beta}^T\mathbf{x}\right\} - \log(y!)\right\}$$

$$E(y|\mathbf{x}, \beta_0, \boldsymbol{\beta}) = \lambda = \exp\left\{\beta_0 + \boldsymbol{\beta}^T\mathbf{x}\right\}$$

(d)

$$\ell(\beta_0, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \log\prod_{i=1}^{n} p(y_i|\mathbf{x}_i, \beta_0, \boldsymbol{\beta})$$
$$= \sum_{i=1}^{n} \log p(y_i|\mathbf{x}_i, \beta_0, \boldsymbol{\beta})$$
$$= \sum_{i=1}^{n} \left\{[\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i]y_i - \exp\left\{\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i\right\} - \log(y_i!)\right\}$$

Problem 4

(a)

$$\frac{\partial}{\partial\beta_0}[-\ell(\beta_0, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X})] = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \exp(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i)$$

$$\frac{\partial}{\partial\boldsymbol{\beta}}[-\ell(\beta_0, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X})] = -\sum_{i=1}^{n} y_i\mathbf{x}_i + \sum_{i=1}^{n} \exp(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i)\mathbf{x}_i$$

(b)

$$\frac{\partial^2}{\partial\beta_0^2}[-\ell(\beta_0, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X})] = \sum_{i=1}^{n} \exp(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i)$$

$$\frac{\partial^2}{\partial\boldsymbol{\beta}^2}[-\ell(\beta_0, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X})] = \sum_{i=1}^{n} \exp(\beta_0 + \boldsymbol{\beta}^T\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T$$

3

$$\frac{\partial^2}{\partial\beta_0\partial\boldsymbol{\beta}}[-\ell(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X})] = \sum_{i=1}^{n}\exp(\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i)\mathbf{x}_i$$

$$\mathrm{H} = \begin{bmatrix} \frac{\partial}{\partial\beta_0^2} & \frac{\partial}{\partial\beta_0\partial\boldsymbol{\beta}} \\ \frac{\partial}{\partial\beta_0\partial\boldsymbol{\beta}} & \frac{\partial}{\partial\boldsymbol{\beta}^2} \end{bmatrix}$$

where $\frac{\partial}{\partial\boldsymbol{\beta}^2}$ is a matrix with $\frac{\partial}{\partial\beta_i\partial\beta_j}$ where $i = 1,\ldots,p, j = 1,\ldots p$

(c) for the function to be convex, the Hessian matrix must be positive definite.

(d) for the ease of computation, we will express the gradient and Hessian in matrix form and let $\boldsymbol{\beta}' = \begin{bmatrix} \beta_0 & \beta_1 & \ldots & \beta_p \end{bmatrix}^T$,

$$\frac{\partial}{\partial\boldsymbol{\beta}'} = -\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\exp(\mathbf{X}\boldsymbol{\beta}')$$

$$\frac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}'^T} = \mathbf{X}^T\mathrm{diag}(\exp\{\mathbf{X}\boldsymbol{\beta}'\})\mathbf{X}$$

Problem 5

To get the loss function, We first remove the constant term $\log(y!)$ from negative log-likelihood since it does not affect the optimization problem.

(a)

$$\mathcal{L}(\beta_0\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n}\left\{[\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i]y_i - \exp\left\{\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i\right\}\right\}$$

Then we add the L2 regularization term to the loss function

$$\mathcal{L}_\lambda(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n}\left\{[\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i]y_i - \exp\left\{\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i\right\}\right\} + \lambda||\boldsymbol{\beta}||_2^2$$

We then derive the gradient and Hessian of the loss function

$$\frac{\partial}{\partial\beta_0}\mathcal{L}_\lambda(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = -\sum_{i=1}^{n}y_i + \sum_{i=1}^{n}\exp(\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i)$$

$$\frac{\partial}{\partial\boldsymbol{\beta}}\mathcal{L}_\lambda(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = -\sum_{i=1}^{n}y_i\mathbf{x}_i + \sum_{i=1}^{n}\exp(\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i)\mathbf{x}_i - 2\lambda\boldsymbol{\beta}$$

$$\frac{\partial^2}{\partial\beta_0^2}\mathcal{L}_\lambda(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n}\exp(\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i)$$

$$\frac{\partial^2}{\partial\boldsymbol{\beta}^2}\mathcal{L}_\lambda(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n}\exp(\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T - 2\lambda\mathbf{I}$$

$$\frac{\partial^2}{\partial\beta_0\partial\boldsymbol{\beta}}\mathcal{L}_\lambda(\beta_0,\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n}\exp(\beta_0+\boldsymbol{\beta}^T\mathbf{x}_i)\mathbf{x}_i$$

we can write the gradient and Hessian in matrix form as

$$\frac{\partial}{\partial \boldsymbol{\beta}'} \mathcal{L}_\lambda(\beta_0, \boldsymbol{\beta}, \boldsymbol{\beta}' | \mathbf{y}, \mathbf{X}) = -\mathbf{X}^T \mathbf{y} + \mathbf{X}^T \exp(\mathbf{X}\boldsymbol{\beta}') - 2\lambda[0; \boldsymbol{\beta}]$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}'^T} \mathcal{L}_\lambda(\beta_0, \boldsymbol{\beta}\boldsymbol{\beta}' | \mathbf{y}, \mathbf{X}) = \mathbf{X}^T \operatorname{diag}(\exp\{\mathbf{X}\boldsymbol{\beta}'\})\mathbf{X} - 2\lambda(\mathbf{I} - e_1 e_1^T)$$

It is required that the Hessian matrix is positive definite for the loss function to be convex.