1. Poisson Regression

(a)

$$\begin{aligned}
p(y;\lambda) &= \frac{e^{-\lambda}\lambda^y}{y!} \\
&= \frac{1}{y!}\exp(\log(\lambda^y)-\lambda) \\
&= \frac{1}{y!}\exp(y\log\lambda-\lambda)
\end{aligned}$$

取

$$b(y)=\frac{1}{y},\eta=\log\lambda,\lambda=a(\eta)=\exp(\eta)$$

则有

$$p(y;\eta)=\frac{1}{y!}\exp(\eta^T y-\exp(\eta))$$

(b)

$$\begin{aligned}
g(\eta) &= E[T(y);\eta] \\
&= E[p(y;\lambda)] \\
&= \lambda \\
&= \exp(\eta)
\end{aligned}$$

(c)

$$\begin{aligned}
\ell(\theta) &= \log(p(y^{(i)}|x^{(i)},\theta)) \\
&= \log(p(y^{(i)};\eta)) \\
&= \log\left(\frac{1}{y^{(i)}!}\right)+\eta^T y^{(i)}-\exp(\eta) \\
&= \log\left(\frac{1}{y^{(i)}!}\right)+\theta^T x^{(i)} y^{(i)}-\exp(\theta^T x^{(i)})
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta}\left(\log\left(\frac{1}{y^{(i)}!}\right) + \theta^T x^{(i)} y^{(i)} - \exp(\theta^T x^{(i)})\right) \\
&= x^{(i)} y^{(i)} - x^{(i)} \exp(\theta^T x^{(i)}) \\
&= y^{(i)} - h_\theta(x^{(i)})
\end{aligned}
$$

$$
\begin{aligned}
\therefore \theta &:= \theta + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x^{(i)} \\
&:= \theta + \alpha \left(x^{(i)} y^{(i)} - x^{(i)} \exp(\theta^T x^{(i)})\right) x^{(i)}
\end{aligned}
$$

2. Convexity of Generalized Linear Models

(a)

由

$$
E(Y;\eta) = \int y p(y;\eta) dy
$$

得

$$
\begin{aligned}
\frac{\partial}{\partial \eta} \int p(y;\eta) dy &= \int \frac{\partial}{\partial \eta} p(y;\eta) dy \\
&= \int \frac{\partial}{\partial \eta} \left(b(y) \exp(\eta y - a(\eta))\right) dy \\
&= \int b(y) \exp(\eta y - a(\eta)) \left(y - \frac{\partial a(\eta)}{\partial \eta}\right) dy \\
&= \int p(y;\eta) \left(y - \frac{\partial a(\eta)}{\partial \eta}\right) dy \\
&= E(Y;\eta) - \int p(y;\eta) \frac{\partial a(\eta)}{\partial \eta} dy
\end{aligned}
$$

又由

$$
\int p(y;\eta) dy = 1
$$

$$
\begin{aligned}
\int p(y;\eta) \frac{\partial a(\eta)}{\partial \eta} dy &= \frac{\partial a(\eta)}{\partial \eta} \int p(y;\eta) dy \\
&= \frac{\partial a(\eta)}{\partial \eta}
\end{aligned}
$$

得

$$E(Y;\eta) = \frac{\partial a(\eta)}{\partial \eta}$$

(b)

由(a)得

$$E(Y;\eta) = \frac{\partial a(\eta)}{\partial \eta}$$

$$\frac{\partial E(Y;\eta)}{\partial \eta} = \frac{\partial^2 a(\eta)}{\partial \eta^2}$$

又由

$$\begin{aligned}
\frac{\partial E(Y;\eta)}{\partial \eta} &= \frac{\partial}{\partial \eta}\int yp(y;\eta)dy \\
&= \int y\frac{\partial p(y;\eta)}{\partial \eta}dy \\
&= \int yp(y;\eta)\left(y - \frac{\partial a(\eta)}{\partial \eta}\right)dy \\
&= \int p(y;\eta)\left(y^2 - y\frac{\partial a(\eta)}{\partial \eta}\right)dy \\
&= E(Y^2;\eta) - E(Y;\eta)\frac{\partial a(\eta)}{\partial \eta} \\
&= E(Y^2;\eta) - E^2(Y;\eta) \\
&= Var(Y;\eta)
\end{aligned}$$

得

$$Var(Y;\eta) = \frac{\partial^2 a(\eta)}{\partial \eta^2}$$

(c)

由

$$
\begin{aligned}
\ell(\theta) &= -\log p(y;\eta) \\
&= -\log(b(y)\exp(\eta y - a(\eta))) \\
&= a(\eta) - \eta y - \log b(y) \\
&= a(\theta^T x) - \theta^T x y - \log b(y)
\end{aligned}
$$

得

$$
\nabla_\theta \ell(\theta) = x \frac{\partial}{\partial \eta} a(\theta^T x) - xy
$$

$$
H = \nabla_\theta^2 \ell(\theta) = x x^T \frac{\partial^2}{\partial \eta^2} a(\theta^T x)
$$

对 $\forall z \in \mathbb{R}^n$, 有

$$
\begin{aligned}
z^T H z &= z^T x x^T \frac{\partial^2}{\partial \eta^2} a(\theta^T x) z \\
&= (x^T z)^T \frac{\partial^2}{\partial \eta^2} a(\eta)(x^T z) \\
&= (x^T z)^2 Var(Y;\eta) \\
&\geq 0
\end{aligned}
$$

$$
\therefore H \succeq 0, \text{GLM的NLL损失是凸函数}
$$

3. Multivariate Least Squares

(a)

$$
\begin{aligned}
J(\Theta) &= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{p} ((\Theta^T x^{(i)})_j - y_j^{(i)})^2 \\
&= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{p} (X\Theta - Y)_{ij}^2 \\
&= \frac{1}{2} \sum_{i=1}^{p} [(X\Theta - Y)^T (X\Theta - Y)]_{ii} \\
&= \frac{1}{2} \operatorname{tr}[(X\Theta - Y)^T (X\Theta - Y)]
\end{aligned}
$$

(b)

$$\begin{aligned}
\frac{\partial J(\Theta)}{\partial \Theta} &= \frac{\partial}{\partial \Theta}\frac{1}{2}\operatorname{tr}[(X\Theta - Y)^T(X\Theta - Y)] \\
&= \frac{1}{2}\frac{\partial}{\partial \Theta}\operatorname{tr}[\Theta^T X^T X\Theta - \Theta^T X^T Y - Y^T X\Theta + Y^T Y] \\
&= \frac{1}{2}\frac{\partial}{\partial \Theta}\operatorname{tr}[\Theta^T X^T X\Theta - 2\Theta^T X^T Y + Y^T Y] \\
&= \frac{1}{2}(2X^T X\Theta - 2X^T Y) \\
&= X^T X\Theta - X^T Y
\end{aligned}$$

$\nabla_\Theta J(\Theta) = \mathbf{0}$时，有

$$X^T X\Theta - X^T Y = \mathbf{0}$$

即

$$\Theta = (X^T X)^{-1} X^T Y$$

(c)

由题意得

$$Y = X(\theta_1, \theta_2, \dots, \theta_p)$$

记

$$Y_j = (y_j^{(1)}, y_j^{(2)}, \dots, y_j^{(m)})^T$$

则

$$Y_j = X\theta_j$$

$$\theta_j = (X^T X)^{-1} X^T Y_j$$

$$(\theta_1, \theta_2, \dots, \theta_p) = (X^T X)^{-1} X^T Y = \Theta$$

$\therefore$ 与上题结论一致

4. Incomplete, Positive-Only Labels

(a)

$$
\begin{aligned}
p(t^{(i)}=1|y^{(i)}=1,x^{(i)}) &= \frac{p(y^{(i)}=1|t^{(i)}=1,x^{(i)})p(t^{(i)}=1,x^{(i)})}{p(y^{(i)}=1,x^{(i)})} \\
&= \frac{p(y^{(i)}=1|t^{(i)}=1,x^{(i)})p(t^{(i)}=1,x^{(i)})}{p(y^{(i)}=1|t^{(i)}=1,x^{(i)})p(t^{(i)}=1,x^{(i)})+p(y^{(i)}=1|t^{(i)}=0,x^{(i)})p(t^{(i)}=0,x^{(i)})} \\
&= \frac{\alpha p(t^{(i)}=1,x^{(i)})}{\alpha p(t^{(i)}=1,x^{(i)})+0} \\
&= 1
\end{aligned}
$$

(b)

$$
\begin{aligned}
p(y^{(i)}=1 \mid x^{(i)}) &= p(y^{(i)}=1, t^{(i)}=1 \mid x^{(i)}) + p(y^{(i)}=1, t^{(i)}=0 \mid x^{(i)}) \\
&= p(y^{(i)}=1 \mid t^{(i)}=1, x^{(i)})\, p(t^{(i)}=1 \mid x^{(i)}) + p(y^{(i)}=1 \mid t^{(i)}=0, x^{(i)})\, p(t^{(i)}=0 \mid x^{(i)}) \\
&= p(y^{(i)}=1 \mid t^{(i)}=1, x^{(i)})\, p(t^{(i)}=1 \mid x^{(i)}) \\
&= \alpha\, p(t^{(i)}=1 \mid x^{(i)})
\end{aligned}
$$

即

$$
p(t^{(i)}=1 \mid x^{(i)}) = \frac{1}{\alpha}\, p(y^{(i)}=1 \mid x^{(i)})
$$

(c)

$$
\begin{aligned}
h(x^{(i)}) &= p(y^{(i)}=1|x^{(i)}) \\
&= \alpha p(t^{(i)}=1|x^{(i)}) \\
&= \alpha
\end{aligned}
$$

则

$$
\begin{aligned}
E[h(x^{(i)})|y^{(i)}=1] &= E[\alpha|y^{(i)}=1] \\
&= \alpha
\end{aligned}
$$