Problem Set #1

4. (c)

① $l(\theta)$: the negative log-likelihood of the distribution (loss)

② Hessian of the loss w.r.t θ → ③ show that it is always PSD.

NLL loss of GLM is always convex

① $l(\theta) = -\log \prod_{i=1}^{m} p(y^{(i)} | x^{(i)}; \theta)$

$= -\sum_{i=1}^{m} \log p(y^{(i)} | x^{(i)}; \theta)$

$= -\sum_{i=1}^{m} \log b(y^{(i)}) \exp(\eta y^{(i)} - a(\eta))$

$= \sum_{i=1}^{m} \left[-\log(b(y^{(i)})) - (\eta y^{(i)} - a(\eta)) \right]$

$= \sum_{i=1}^{m} \left[-\log(b(y^{(i)})) - \eta y^{(i)} + a(\eta) \right]$ $\quad \eta \in \mathbb{R}$, $x \in \mathbb{R}^n$ $\theta \in \mathbb{R}^n$ $\eta = \theta^T x$

$= \sum_{i=1}^{m} \left[-\log(b(y^{(i)})) - \theta^T x^{(i)} y^{(i)} + a(\theta^T x^{(i)}) \right]$

② $\frac{\partial}{\partial \theta} l(\theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^{m} \left[-\log(b(y^{(i)})) - \theta^T x^{(i)} y^{(i)} + a(\theta^T x^{(i)}) \right]$

$= \sum_{i=1}^{m} \frac{\partial}{\partial \theta} \left[-\log(b(y^{(i)})) - \theta^T x^{(i)} y^{(i)} + a(\theta^T x^{(i)}) \right]$

#1 derivative $= \sum_{i=1}^{m} \left\{ -x_j^{(i)} y^{(i)} + \frac{\partial}{\partial \theta_j} a(\theta^T x^{(i)}) \right\}$

$= \sum_{i=1}^{m} \left\{ -x_j^{(i)} y^{(i)} + x_j^{(i)} a'(\theta^T x^{(i)}) \right\}$

$= \sum_{i=1}^{m} \left\{ a'(\theta^T x^{(i)}) - y^{(i)} \right\} x_j^{(i)}$

#2

derivative $= \frac{\partial}{\partial \theta_k} \sum_{i=1}^{m} \{ a'(\theta^T x^{(i)}) - y^{(i)} \} x_j^{(i)}$

$$= \sum_{i=1}^{m} \left\{ \frac{\partial}{\partial \theta_k} a'(\theta^T x^{(i)}) \right\} x_j^{(i)}$$

$$= \sum_{i=1}^{m} \{ x_k^{(i)} \times a''(\theta^T x^{(i)}) \} x_j^{(i)}$$

Hessian $= \sum_{i=1}^{m} \left\{ \frac{\partial^2 a(\eta)}{\partial \eta^2} \right\} x_j^{(i)} x_k^{(i)}$

$\Updownarrow$

$\frac{\partial^2 \ell(\theta)}{\partial \theta_j \partial \theta_k}$ $= \frac{\partial^2 a(\eta)}{\partial \eta^2} \sum_{i=1}^{m} x_j^{(i)} x_k^{(i)}$

$$H = Var(Y|X;\theta) \sum_{i=1}^{m} x^{(i)} x^{(i)T}$$

③ For any $z \in \mathbb{R}^n$, we have

$$z^T H z = z^T Var(Y|X;\theta) \sum_{i=1}^{m} x^{(i)} x^{(i)T} z$$

$$= Var(Y|X;\theta) \sum_{i=1}^{m} z^T x^{(i)} x^{(i)T} z$$

$$= Var(Y|X;\theta) \sum_{i=1}^{m} (z^T x^{(i)})^2 \geq 0$$

$\therefore$ Hessian of negative log likelihood of exponential family is always positive semi-defitite

$\Rightarrow$ Any GLM model is convex in its model parameters.