Exponential Family and GLM Properties

Question 4

Part (a): Expectation of the Exponential Family

Proof.

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

which implies:

$$y\, p(y;\eta) = \frac{\partial}{\partial \eta} p(y;\eta) + p(y;\eta) \frac{\partial}{\partial \eta} a(\eta)$$

Therefore,

$$\begin{aligned}
\mathbb{E}[Y;\eta] &= \int y\, p(y;\eta)\, dy \\
&= \int \left(\frac{\partial}{\partial \eta} p(y;\eta) + p(y;\eta) \frac{\partial}{\partial \eta} a(\eta) \right) dy \\
&= \frac{\partial}{\partial \eta} \int p(y;\eta)\, dy + \frac{\partial}{\partial \eta} a(\eta) \int p(y;\eta)\, dy \\
&= \frac{\partial}{\partial \eta}(1) + \frac{\partial}{\partial \eta} a(\eta) \cdot 1 \\
&= \frac{\partial}{\partial \eta} a(\eta)
\end{aligned}$$

Hence, the mean of an exponential family distribution is the first derivative of the log-partition function with respect to the natural parameter.

Part (b): Variance of the Exponential Family

Proof.

$$\begin{aligned}
\frac{\partial^2}{\partial \eta^2} p(y;\eta) &= \frac{\partial}{\partial \eta}\left(y p(y;\eta) - p(y;\eta)\frac{\partial}{\partial \eta} a(\eta)\right) \\
&= p(y;\eta) + y^2 p(y;\eta) - 2y p(y;\eta)\frac{\partial}{\partial \eta} a(\eta) \\
&\quad + p(y;\eta)\left(\frac{\partial}{\partial \eta} a(\eta)\right)^2 - p(y;\eta)\frac{\partial^2}{\partial \eta^2} a(\eta) \\
&= p(y;\eta) - p(y;\eta)\frac{\partial^2}{\partial \eta^2} a(\eta) + \left(y - \frac{\partial}{\partial \eta} a(\eta)\right)^2 p(y;\eta)
\end{aligned}$$

which implies:

$$\left(y - \frac{\partial}{\partial \eta} a(\eta)\right)^2 p(y;\eta) = \frac{\partial^2}{\partial \eta^2} p(y;\eta) - p(y;\eta) + p(y;\eta)\frac{\partial^2}{\partial \eta^2} a(\eta)$$

Hence,

$$\begin{aligned}
\mathrm{Var}[Y;\eta] &= \int (y - \mathbb{E}[Y;\eta])^2\, p(y;\eta)\, dy \\
&= \int \left(y - \frac{\partial}{\partial \eta} a(\eta)\right)^2 p(y;\eta)\, dy \\
&= \int \left(\frac{\partial^2}{\partial \eta^2} p(y;\eta) - p(y;\eta) + p(y;\eta)\frac{\partial^2}{\partial \eta^2} a(\eta)\right) dy \\
&= \frac{\partial^2}{\partial \eta^2}\int p(y;\eta)\, dy - \int p(y;\eta)\, dy + \frac{\partial^2}{\partial \eta^2} a(\eta)\int p(y;\eta)\, dy \\
&= 0 + 1 \cdot \frac{\partial^2}{\partial \eta^2} a(\eta) \\
&= \frac{\partial^2}{\partial \eta^2} a(\eta)
\end{aligned}$$

Thus, the variance of an exponential family distribution equals the second derivative of the log-partition function with respect to the natural parameter.

Part (c): Convexity of the NLL in GLMs

Recall the NLL:

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

Gradient:

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

Hessian:

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

For any $z \in \mathbb{R}^n$:

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

Thus, the Hessian is positive semidefinite (PSD), proving that the NLL loss for generalized linear models (GLMs) is convex.