4. GLMs are trained using the negative log-likelihood (NLL) as the loss function.

↳ Mathematically Equivalent to Maximum Likelihood Estimation.

⇒ Our goal is to show that NLL loss of a GLM is a convex function wrt model parameters.

↳ This is convenient because a convex function is one for which any local minimum is also a global minimum.

$$\boxed{P(y;\eta) = b(y)\exp(\eta y - a(\eta))}$$

(a) $$E[Y] = \int_{-\infty}^{\infty} y\,P(y)\,dy$$

$$= \int_{-\infty}^{\infty} y\,b(y)\,e^{(\eta y - a(\eta))}\,dy$$

$$\frac{\delta}{\delta\eta}\int_{-\infty}^{\infty} P(y)\,dy = \int_{-\infty}^{\infty} \frac{\delta}{\delta\eta} P(y)\,dy$$

0

$$= \int_{-\infty}^{\infty} (y - a'(\eta))\,b(y)\exp(\eta y - a(\eta))\,dy$$

$$\Rightarrow \int_{-\infty}^{\infty} y\, b(y) \exp(\eta y - a(\eta))\, dy$$

$$- a'(\eta) \int_{-\infty}^{\infty} b(y) \exp(\eta y - a(\eta))\, dy$$

$$\Rightarrow E[Y] - a'(\eta) = 0$$

$$\boxed{E[Y] = a'(\eta)}$$