GLM Cheatsheet

Pavan Ramkumar

July 14, 2016

Abstract

We derive the gradients and Hessians of the penalized log likelihood loss to use as updates in the Newton coordinate descent algorithm for GLMs.

1 Cheatsheet

distr	μ	$\mathcal{L}$	J	$\partial J/\partial \beta_j$	$\partial^2 J/\partial \beta_j^2$
poisson	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (4)	Eq. (5)
poissonexp	Eq. (6)	Eq. (7)	Eq. (8)	Eq. (9)	Eq. (10)
normal	Eq. (11)	Eq. (12)	Eq. (13)	Eq. (14)	Eq. (15)
binomial	Eq. (16)	Eq. (17)	Eq. (18)	Eq. (19)	Eq. (20)
multinomial	Eq. (21)	Eq. (22)	Eq. (23)	Eq. (24)	Eq. (25)

2 Poisson: Softplus

2.1 Mean function

$$\begin{aligned} z_i &= \beta_0 + \sum_j \beta_j x_{ij} \\ \mu_i &= \log(1 + \exp(z_i)) \end{aligned} \tag{1}$$

2.2 Log-likelihood function

$$\mathcal{L} = \sum_i y_i \log(\mu_i) - \sum_i \mu_i \tag{2}$$

2.3 L2-penalized loss function

$$\begin{aligned} J = \frac{1}{n} \sum_i \left\{ \log(1 + \exp(\beta_0 + \sum_j \beta_j x_{ij})) \right\} \\ -\frac{1}{n} \sum_i \left\{ y_i \log(\log(1 + \exp(\beta_0 + \sum_j \beta_j x_{ij}))) \right\} \\ +\lambda(1-\alpha)\frac{1}{2} \sum_j \beta_j^2 \end{aligned} \tag{3}$$

2.4 Gradient

$$
\begin{aligned}
\mu(z_i) &= \log(1+\exp(z_i)) \\
\sigma(z_i) &= \frac{1}{1+\exp(-z_i)} \\
\frac{\partial J}{\partial \beta_0} &= \frac{1}{n}\sum_i \sigma(z_i) - \frac{1}{n}\sum_i y_i \frac{\sigma(z_i)}{\mu(z_i)} \\
\frac{\partial J}{\partial \beta_j} &= \frac{1}{n}\sum_i \sigma(z_i) x_{ij} - \frac{1}{n}\sum_i \sigma(z_i) y_i \frac{\sigma(z_i)}{\mu(z_i)} x_{ij} + \lambda(1-\alpha)\beta_j
\end{aligned} \tag{4}
$$

2.5 Hessian

$$
\begin{aligned}
\mu(z_i) &= \log(1+\exp(z_i)) \\
\sigma(z_i) &= \frac{1}{1+\exp(-z_i)} \\
\frac{\partial^2 J}{\partial \beta_0^2} &= \frac{1}{n}\sum_i \sigma(z_i)(1-\sigma(z_i)) - \frac{1}{n}\sum_i y_i \left\{ \frac{\sigma(z_i)(1-\sigma(z_i))}{\mu(z_i)} - \frac{\sigma(z_i)}{\mu(z_i)^2} \right\} \\
\frac{\partial^2 J}{\partial \beta_j^2} &= \frac{1}{n}\sum_i \sigma(z_i)(1-\sigma(z_i)) x_{ij}^2 - \frac{1}{n}\sum_i y_i \left\{ \frac{\sigma(z_i)(1-\sigma(z_i))}{\mu(z_i)} - \frac{\sigma(z_i)}{\mu(z_i)^2} \right\} x_{ij}^2 + \lambda(1-\alpha)
\end{aligned} \tag{5}
$$

3 Poisson: Linearized

3.1 Mean function

$$
\begin{aligned}
z_i &= \beta_0 + \sum_j \beta_j x_{ij} \\
\mu_i &= \begin{cases} \exp(z_i), & z_i \leq \eta \\ \exp(\eta) z_i + (1-\eta)\exp(\eta), & z_i > \eta \end{cases}
\end{aligned} \tag{6}
$$

3.2 Log-likelihood function

$$
\mathcal{L} = \sum_i y_i \log(\mu_i) - \sum_i \mu_i \tag{7}
$$

3.3 L2-penalized loss function

$$
J = -\frac{1}{n}\mathcal{L} + \lambda(1-\alpha)\frac{1}{2}\sum_j \beta_j^2 \tag{8}
$$

3.4 Gradient

$$
\begin{aligned}
\mu_i &= \begin{cases} \exp(z_i), & z_i \leq \eta \\ \\ \exp(\eta) z_i + (1-\eta)\exp(\eta), & z_i > \eta \end{cases} \\
\frac{\partial J}{\partial \beta_0} &= \frac{1}{n} \sum_{i; z_i \leq \eta} (\mu_i - y_i) + \frac{1}{n} \sum_{i; z_i > \eta} \eta (1 - y_i/\mu_i) \\
\frac{\partial J}{\partial \beta_j} &= \frac{1}{n} \sum_{i; z_i \leq \eta} (\mu_i - y_i) x_{ij} + \frac{1}{n} \sum_{i; z_i > \eta} \eta (1 - y_i/\mu_i) x_{ij}
\end{aligned} \tag{9}
$$

3.5 Hessian

$$
\begin{aligned}
\mu_i &= \begin{cases} \exp(z_i), & z_i \leq \eta \\ \\ \exp(\eta) z_i + (1-\eta)\exp(\eta), & z_i > \eta \end{cases} \\
\frac{\partial^2 J}{\partial \beta_0^2} &= \frac{1}{n} \sum_{i; z_i \leq \eta} \mu_i - \frac{1}{n} \sum_{i; z_i > \eta} \exp(\eta)(1-\eta) \frac{y_i}{\mu_i^2} \\
\frac{\partial^2 J}{\partial \beta_j^2} &= \frac{1}{n} \sum_{i; z_i \leq \eta} \mu_i x_{ij}^2 - \frac{1}{n} \sum_{i; z_i > \eta} \exp(\eta)(1-\eta) \frac{y_i}{\mu_i^2} x_{ij}^2 + \lambda(1-\alpha)
\end{aligned} \tag{10}
$$

4 Normal

4.1 Mean function

$$
\begin{aligned}
z_i &= \beta_0 + \sum_j \beta_j x_{ij} \\
\mu_i &= z_i
\end{aligned} \tag{11}
$$

4.2 Log-likelihood function

$$
\mathcal{L} = -\frac{1}{2} \sum_i (y_i - \mu_i)^2 \tag{12}
$$

4.3 L2-penalized loss function

$$
J = \frac{1}{2n} \sum_i (y_i - (\beta_0 + \sum_j \beta_j x_{ij}))^2 + \lambda(1-\alpha) \frac{1}{2} \sum_j \beta_j^2 \tag{13}
$$

4.4 Gradient

$$\begin{aligned}
\mu(z_i) &= z_i \\
\frac{\partial J}{\partial \beta_0} &= -\frac{1}{n}\sum_i (y_i - \mu_i) \\
\frac{\partial J}{\partial \beta_j} &= -\frac{1}{n}\sum_i (y_i - \mu_i)x_{ij} + \lambda(1-\alpha)\beta_j
\end{aligned} \tag{14}$$

4.5 Hessian

$$\begin{aligned}
\frac{\partial^2 J}{\partial \beta_0^2} &= 1 \\
\frac{\partial^2 J}{\partial \beta_j^2} &= \frac{1}{n}\sum_i x_{ij}^2 + \lambda(1-\alpha)
\end{aligned} \tag{15}$$

5 Binomial

5.1 Mean function

$$\begin{aligned}
z_i &= \beta_0 + \sum_j \beta_j x_{ij} \\
\mu_i &= \frac{1}{1+\exp(-z_i)}
\end{aligned} \tag{16}$$

5.2 Log-likelihood function

$$\mathcal{L} = \sum_i \{y_i \log(\mu_i) + (1-y_i)\log(1-\mu_i)\} \tag{17}$$

5.3 L2-penalized loss function

$$J = -\frac{1}{n}\sum_i \left\{ y_i \log(\beta_0 + \sum_j \beta_j x_{ij}) + (1-y_i)\log(1-(\beta_0 + \sum_j \beta_j x_{ij})) \right\} + \lambda(1-\alpha)\frac{1}{2}\sum_j \beta_j^2 \tag{18}$$

5.4 Gradient

$$\begin{aligned}
\mu(z_i) &= \frac{1}{1+\exp(-z_i)} \\
\frac{\partial J}{\partial \beta_0} &= -\frac{1}{n}\sum_i (y_i - \mu_i) \\
\frac{\partial J}{\partial \beta_j} &= -\frac{1}{n}\sum_i (y_i - \mu_i)x_{ij} + \lambda(1-\alpha)\beta_j
\end{aligned} \tag{19}$$

5.5 Hessian

$$
\begin{aligned}
\frac{\partial^2 J}{\partial \beta_0^2} &= \frac{1}{n} \sum_i \mu_i (1 - \mu_i) \\
\frac{\partial^2 J}{\partial \beta_j^2} &= \frac{1}{n} \sum_i \mu_i (1 - \mu_i) x_{ij}^2 + \lambda (1 - \alpha)
\end{aligned} \tag{20}
$$

6 Multinomial