Group 8 - GLM Exercises

November 10, 2015

1. A density that belongs in the one-parameter exponential family has the following (canonical) representation:

$p(z) = e^{(q[z\theta + c(\theta)] + h(z,q)}$

where θ is called the natural parameter, and c and h are functions whose exact form depends on the particular density. For many members of this family, q > 0, is a further unknown parameter, often called a precision parameter.

1. The following densities belong to the exponential family. Identify θ, q and an appropriate $c(\theta)$, for each of them (take into account that for some z might be a transformation of t):

Normal

$t \backsim N(\mu, -q) = exp\left\{\frac{-(t-\mu)^2}{2q}\right\}$

$= exp\left\{\frac{-(t^2-2\mu t+\mu^2)}{2q}\right\}$

$= exp\left\{-\frac{1}{2q}\left[t^2 - 2\mu t + \mu^2\right]\right\}$

$= exp\left\{-\frac{1}{q}\left[\frac{t^2}{2} - \mu t + \frac{\mu^2}{2}\right]\right\}$

$= exp\left\{\frac{1}{q}\left[-\frac{t^2}{2} + \mu t - \frac{\mu^2}{2}\right]\right\}$

$= exp\left\{\frac{1}{q}\left[\mu t - \frac{\mu^2}{2}\right] - \frac{t^2}{2q}\right\}$

Result:

$\theta = \mu$

$q = q$ (since q^{-1} is the variance parameter passed, than q the paramter is the precision and the q above is the variance for the normal distribution)

$c(\theta) = \frac{\mu^2}{2}$

Bernoulli

$t \backsim Bern(n,p) = p^z(1-p)^{1-z}$

$= exp\left\{log(p^z) + log((1-p)^{1-z})\right\}$

$= exp\left\{ zlog(p) + (1-z)log(1-p) \right\}$

$= exp\left\{ zlog(p) + log(1-p) - zlog(1-p) \right\}$

$= exp\left\{ zlog(\frac{p}{1-p}) + log(1-p) \right\}$

Result:

$\theta = log(\frac{p}{1-p})$

$q = 1$

$c(\theta) = -log(1-p)$

Binomial

$t \backsim Bin(n,p) = \binom{n}{k} p^k (1-p)^{n-k}$

$= exp\left\{ log(\binom{n}{k}) + log(p^k) + log((1-p)^{n-k}) \right\}$

$= exp\left\{ log(\binom{n}{k}) + klog(p) + (n-k)log(1-p) \right\}$

$= exp\left\{ log(\binom{n}{k}) + klog(\frac{p}{1-p}) + nlog(1-p) \right\}$

$= exp\left\{ n\left[\frac{k}{n} log(\frac{p}{1-p}) + log(1-p) \right] + log(\binom{n}{k}) \right\}$

Result:

$\theta = log(\frac{p}{1-p})$

$q = n$

$c(\theta) = log(\binom{n}{k})$

Poisson

$t \backsim Poisson(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$

$= exp\left\{ log(\lambda^k e^{-\lambda}) - log(k!) \right\}$

$= exp\left\{ k * log(\lambda) - \lambda log(e) - log(k!) \right\}$

$= exp\left\{ k * log(\lambda) - \lambda - log(k!) \right\}$

Result:

$\theta = log(\lambda)$

$q = 1$

$c(\theta) = -\lambda$

2. Identify the canonical link function for each of the models given above.

Normal

μ

Bernoulli

$logit(\mu)$

Binomial

$logit(\mu)$

Poisson

$log(\mu)$

3. Consider now the generalised linear model:

$t_n|x_n \backsim NdEF(\theta(x_n, w), q\gamma_n)$ with

$\theta(x_n, w) = (c')^{-1}(g^{-1}(\phi(x_n)^T w)) =: f(\phi(x_n)^T w)$

We know the log-likehood of the NdEF family of distributions takes the form:

$2log(p(t|x, \gamma, w, q)) = 2q \sum_n \gamma_n \Bigg[t_n\theta(x_n, w) - c(\theta(x_n, w)) \Bigg]$

We will prove it is log-concave if it is the negative of a log-convex function, so will prove:

$f(\theta(x_n, w), q\gamma_n) = -2q \sum_n \gamma_n \Bigg[t_n\theta(x_n, w) - c(\theta(x_n, w)) \Bigg]$

is log-covex, taking into account the link function constructed as:

$g(y) = c'^{-1}(\theta) => \theta(x_n, w) = \phi(x_n^T w$

We will prove it is log-convex by showing the second derivative is positive semi-definite for all $\theta(x_n, w)$

The first derivative of the negative log likelihood is:

$f'(\theta(x_n, w), q\gamma_n) = -2q \Bigg[\sum_n \gamma_n t_n \phi(x_n)^T - \sum_n \gamma_n c'(\phi(x_n^T w)) \Bigg]$

$f''(\theta(x_n, w), q\gamma_n) = 2q \sum_n \gamma_n c''(\phi(x_n)^T w)$

We know that c'' is the variance function and γn and q are positive by definition. So the second derivative is always positive and therefore:

$x^T f'' x \geq 0$ for all x.

2. R^2 and deviance

1. Show that for any linear regression model:

$-2logp(t|X, w_{MLE}, q_{MLE}) = N log e^T e + const$

where "const" does not depend on M or X.

We know that:

$w_{mle} = (\phi^T \phi)^{-1} \phi^T t$

and

$q_{mle} = \left(\frac{1}{N} e^T e \right)^{-1}$

We use this to solve for the deviance of the maximum likelihood function:

$-2logp(t|X, w_{mle}, q_{mle}) = -Nlog(q_{mle}) + q_{mle} e^T e$

$-Nlog(q_{mle}) + q_{mle} e^T e = -Nlog\left((\frac{1}{N} e^T e)^{-1} \right) + \left(\frac{1}{N} e^T e \right)^{-1} e^T e$

$= Nlog\left((\frac{1}{N} e^T e) \right) + N$

$= N[log(e^T e) - log(N)] + N$

$= Nlog(e^T e) - N(log(N) + 1)$

$= Nlog(e^T e) + const$

2. Show that in the null model,

$w_{0,MLE} = \bar{t}$

We know from **1**:

$w_{mle} = (\phi^t \phi)^{-1} \phi t$

In the case of the null model, ϕ is a vector of ones so this becomes:

$\frac{1}{N} \sum_{n=1}^{N} t_n = \bar{t}$

3. The null model is nested within the saturated model, and it corresponds to the special case where $w_1 = ...w_M = 0$. Let D_0 be the deviance of the null model and D_1 be that of the saturated model. Show that:

$D_0 - D_1 = -Nlog(1 - R^2)$

where R^2 is the coefficient R^2 for the saturated model.

We know from **part 1** that:

$-2logp(t|X, w_{mle}, q_{mle}) = Nlog(e^T e) + const$

Which also what we equate as D_M for model M, so:

$D_0 - D_1 = Nlog(e_0^T e_0) - Nlog(e_1^T e_1)$

$= Nlog(\frac{e_0^T e_0}{e_1^T e_1})$

We use the following properties:

$e = t - \hat{t}$, and,

$R^2 = 1 - \frac{\sum_i (t_i - \hat{t}_i)^2}{\sum_i (t_i - \bar{t})^2}$

$1 - R^2 = \frac{\sum_i (t_i - \hat{t}_i)^2}{\sum_i (t_i - \bar{t})^2}$

To show that

$= Nlog(\frac{e_0^T e_0}{e_1^T e_1})$

$$= Nlog(\frac{\sum_i (t_i - \bar{t})^2}{\sum_i (t_i - \hat{t_i})^2})$$

$$= Nlog(\sum_i (t_i - \bar{t})^2) - Nlog(\sum_i (t_i - \hat{t_i})^2)$$

$$= -Nlog(\frac{\sum_i (t_i - \hat{t_i})^2}{\sum_i (t_i - \bar{t})^2})$$

$$= -Nlog(1 - R^2)$$