CS 229, Fall 2016 PS7
Abhishek Kumar Srivastava November 29th, 2016
Student Id : 861307778

Negative Binomial Regression

Part a.
As given in the problem the negative binomial distribution is

$$p(y) = \binom{y+r-1}{y} q^y (1-q)^r$$

Step:1 Taking log of both side leaving out binomial coefficient.

$$\log(p(y)) = \binom{y+r-1}{y} y \log q + r \log(1-q)$$

Step:2 Taking exponential of both side.

$$p(y) = \binom{y+r-1}{y} \exp(y \log q + r \log(1-q))$$

As we know the exponential family are in the form of

$$p(y) = h(y) \exp(\theta^T \phi(y) - A(\theta))$$

By comparison,

$$h(y) = \binom{y+r-1}{y}$$

$$\phi(y) = y$$

$$\theta = \log(q)$$
$$\therefore q = \exp(\theta)$$

$$A(\theta) = -r \log(1-q) = -r \log(1 - \exp(\theta))$$

Therefore from above comparison it is a member of the exponential family.

Part b.
$E[y|x]$ is expected value of y given x.
Let μ be the non-linear function mapping the linear function of x to the mean.
The link function, g, is the inverse of the mean function, μ.

$$E[y|x] = \mu(w^T X) = g^{-1}(w^T X)$$

θ, the parameters of the distribution (which depends on x) are related to the mean through the log-partition function.
ψ is the map from the mean of the distribution to the parameters.

$$E[y|x] = \nabla_\theta A(\theta) = \psi^{-1}(\theta)$$

To calculate $E[y]$ we need to take partial derivative of $A(\theta)$

$$\begin{aligned}
E[y] &= \frac{\partial A(\theta)}{\partial \theta} = \frac{\partial(-r\log(1-\exp(\theta)))}{\partial \theta} \\
&= -r(\frac{-\exp(\theta)}{1-\exp(\theta)}) = r(\frac{\exp(\theta)}{1-\exp(\theta)}) \\
\text{Let } \mu &= E[y], \\
\mu &= r(\frac{\exp(\theta)}{1-\exp(\theta)}) \\
\therefore \mu &= \frac{r}{\exp(-\theta)-1} \\
\exp(-\theta) - 1 &= \frac{r}{\mu} \\
\exp(-\theta) &= \frac{r}{\mu} + 1 \\
-\theta &= \log((\frac{r}{\mu}) + 1) = \log(\frac{r+\mu}{\mu}) \\
\theta &= -\log(\frac{r+\mu}{\mu}) = \log(\frac{\mu}{r+\mu}) \\
\therefore \theta &= \log(\frac{E[y]}{r+E[y]})
\end{aligned}$$

So, canonical link function for negative binomial regression is:

$$\begin{aligned}
g(\mu) &= \psi(\mu) = \log(\frac{E[y]}{r+E[y]}) = \log(\frac{\mu}{r+\mu}) \\
g^{-1}(\theta) &= \frac{r}{\exp(-\theta)-1} = \frac{r}{\exp(-w^T X)-1}
\end{aligned}$$

To predict q from learned parameter w and point x, we can use following equation,

$$q = \exp(\theta) = \exp(w^T x)$$

using the relation between q and θ from previous part.

Part c.
Newton-Raphson equation:

$$w_{new} = w_{old} - \frac{f'(x)}{f''(x)}$$

In this case f(x) is Likelihood function. Therefore,

$$w_{new} = w_{old} - \frac{\bigtriangledown_w L(w)}{\bigtriangledown \bigtriangledown_w L(w)}$$

As given,

$$\begin{aligned}
p(y) &= \binom{y+r-1}{y} q^y (1-q)^r \\
&= h(y)\exp(\theta y - A(\theta)) \\
\therefore L(w) &= \log(p(y|x,w)) \\
&= \sum_i \log p(y_i|x_i,w) \\
&= \sum_i \theta(x_i,w)y + r\log(1-\exp(\theta(x_i,w))) \\
\text{Considering } X \text{ \& } Y &\text{ as vector and } \theta = X^T w \\
\therefore L(w) &= X^T wY + r\log(1-\exp(X^T w))
\end{aligned}$$

Calculating $\bigtriangledown_w L(w)$:

$$\begin{aligned}
&= \frac{\partial}{\partial w}(X^T wY + r\log(1-\exp(X^T w))) \\
&= X^T Y - r\frac{X^T \exp(X^T w)}{1-\exp(X^T w)} \\
\text{From Part b,}& \\
&= X^T Y - X^T E[Y] = (Y - E[Y])X^T \\
\therefore \bigtriangledown_w L(w) &= (Y-\mu)X^T
\end{aligned}$$

Calculating $\bigtriangledown \bigtriangledown_w L(w)$:

From above,

$$
\begin{aligned}
L^{'}(w) &= X^T Y - r\frac{X^T \exp(X^T w)}{1 - \exp(X^T w)} \\
\bigtriangledown \bigtriangledown_w L(w) &= \frac{\partial}{\partial w} L^{'}(w) \\
&= \frac{\partial}{\partial w}(X^T Y - r\frac{X^T \exp(X^T w)}{1 - \exp(X^T w)}) \\
&= -rX^T \frac{\partial}{\partial w}(\frac{\exp(X^T w)}{1 - \exp(X^T w)}) \\
&= -rX^T \frac{\exp(X^T w) X}{(1 - \exp(X^T w))^2} \\
&= -X^T \frac{r \exp(X^T w)}{(1 - \exp(X^T w))^2} X
\end{aligned}
$$

As we know from part b,

$$
\begin{aligned}
\mu &= r(\frac{\exp(\theta)}{1 - \exp(\theta)}) \\
\mu^{'} &= \frac{\partial \mu}{\partial \theta} = r\frac{\partial}{\partial \theta}(\frac{\exp(\theta)}{1 - \exp(\theta)}) \\
&= r\frac{\exp(\theta)}{(1 - \exp(\theta))^2} \\
&= \mu + \frac{\mu^2}{r} \\
\therefore \bigtriangledown \bigtriangledown_w L(w) &= -X^T \mu^{'} X
\end{aligned}
$$

Therefore the weight update equation will be:

$$
\begin{aligned}
w_{new} &= w_{old} - \frac{X^T (Y - \mu)}{(-X^T \mu' X)} \\
&= w_{old} - (-X^T \mu^{'} X)^{-1} X^T (Y - \mu) \\
\therefore w_{new} &= w_{old} + (X^T \mu^{'} X)^{-1} X^T (Y - \mu)
\end{aligned}
$$