- Nikhil Kamthe
- 861245635
- 11/24/2016
- CS 229
- PS 7

Solution(part a): The negative binomial distribution described the number of successes before r failures if the probability of success is q:

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

Consider

$$log(q) = \theta$$
$$q = e^{\theta}$$
$$1 - q = 1 - e^{\theta}$$

Therefore,

$$p(y) = \binom{y+r-1}{y} \exp\left(y\theta + r\ log(1-e^{\theta})\right)$$

Comparing with equation for exponential family,

$$\phi(y) = y$$
$$\theta = log(q)$$
$$A(\theta) = -r\ log(1-e^{\theta})$$
$$h(y) = \binom{y+r-1}{y}$$
$$Z(\theta) = e^{-r\ log(1-e^{\theta})}$$

Solution(part b):

$$\begin{aligned}
E(y|x) &= \frac{d}{d\theta}(A(\theta)) \\
\mu &= \frac{-r}{1-e^{\theta}}(-e^{\theta}) \\
\mu &= \frac{re^{\theta}}{1-e^{\theta}} \\
\frac{\mu}{r} &= \frac{e^{\theta}}{1-e^{\theta}} \\
\frac{r}{\mu} &= \frac{1-e^{\theta}}{e^{\theta}} \\
\frac{r}{\mu} &= \frac{1}{e^{\theta}} - 1 \\
\frac{1}{e^{\theta}} &= \frac{r}{\mu} + 1 \\
e^{-\theta} &= \frac{r+\mu}{\mu} \\
\theta &= log(\frac{\mu}{r+\mu}) \\
\psi(\mu) &= log(\frac{\mu}{r+\mu})
\end{aligned}$$

Using cononical link function: $g = \psi$

$$\therefore g(\mu) = log(\frac{\mu}{r+\mu})$$

as $g = \psi$,

$$\theta = x^T w$$

Using results from part a,

$$\begin{aligned}
log(q) &= x^T w \\
q &= e^{x^T w}
\end{aligned}$$

Solution(part c): Loss function in case a canonical link function can be given as as follows:

$$L(w) = \frac{1}{\sigma^2} \sum_i (x_i^T w y_i - A(x_i^T w))$$

Gradient:

$$\begin{aligned}
\bigtriangledown_w L(w) &= \frac{1}{\sigma^2} \bigtriangledown_w \left(\sum_i (x_i^T w y_i - A(x_i^T w))\right) \\
&= \frac{1}{\sigma^2} \bigtriangledown_w \left(\sum_i (x_i^T w y_i + rlog(1 - e^{x_i^T w}))\right) \\
&= \frac{1}{\sigma^2} \sum_i (x_i y_i + x_i (rlog(1 - e^{x_i^T w}))) \\
&= \frac{1}{\sigma^2} YX - \mu X \\
&= \frac{1}{\sigma^2} (Y - \mu) X
\end{aligned}$$

Hessian:

$$\begin{aligned}
\bigtriangledown \bigtriangledown_w L(w) &= \bigtriangledown\left(\frac{1}{\sigma^2}(Y - \mu)X\right) \\
&= \frac{1}{\sigma^2} \frac{d}{dw}((Y - \mu)X) \\
&= -\frac{1}{\sigma^2} \frac{d}{dw}(\mu X) \\
&= -\frac{1}{\sigma^2} \frac{d\mu}{dw} \frac{d}{d\mu}(\mu X) \\
&= -\frac{1}{\sigma^2} \frac{d\mu}{dw} X^T \\
&= -\frac{1}{\sigma^2} \frac{d\theta}{dw} \frac{d\mu}{d\theta} X^T \\
&= -\frac{1}{\sigma^2} X \frac{d\mu}{d\theta} X^T \\
&= -\frac{1}{\sigma^2} X \mu' X^T \qquad \text{where } \mu' = \frac{d\mu}{d\theta}
\end{aligned}$$

Update rule:

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