Eli Weissler

Math189R SU17
Homework 3
Wednesday, May 24, 2017

Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though.

1 (**Murphy 2.16**) Suppose $\theta \sim \text{Beta}(a, b)$ such that

$$\mathbb{P}(\theta; a, b) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the Beta function and $\Gamma(x)$ is the Gamma function. Derive the mean, mode, and variance of θ.

■

2 (**Murphy 9**) Show that the multinomial distribution

$$\text{Cat}(\mathbf{x}|\boldsymbol{\mu}) = \prod_{i=1}^{K} \mu_i^{x_i}$$

is in the exponential family and show that the generalized linear model corresponding to this distribution is the same as multinomial logistic regression (softmax regression).

First we write the multinomial distribution in a single line:

$$P(y|\phi) = \phi_1^{y_1} \phi_2^{y_2} \dots \phi_K^{y_K} \qquad y = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

only 1 at the category it is

We can reduce one dof by noting $\sum y_i = 1$

$$P(y|\phi) = \phi_1^{y_1} \phi_2^{y_2} \dots (\phi_K)^{1-\sum_{i=1}^{K-1} y_i}$$

Now force it into exponential family

$$P(y|\phi) = \exp\left(\log\left[\prod_{i=1}^{K-1} \phi_i^{y_i} \cdot (\phi_K)^{1-\sum_{i=1}^{K-1} y_i}\right]\right)$$

$$= \exp\left(\sum_i y_i \log(\phi_i) + \log\phi_K - \sum_i y_i \log(\phi_K)\right)$$

$$= \exp\left(\sum_i y_i \log\frac{\phi_i}{\phi_K} + \log(\phi_K)\right) \qquad \text{let } \vec{\eta} = \begin{bmatrix} \log\frac{\phi_i}{\phi_K} \\ \vdots \end{bmatrix}$$

Then $\phi_i = \phi_K e^{\eta_i}$ & $1 - \sum \phi_i = \phi_K = 1 - \sum_{i=1}^{K-1} \phi_K e^{\eta_i} = 1 - \phi_K e^{\eta_i}$

$$\rightarrow \phi_K = \frac{1}{1+\sum e^{\eta_i}}$$

Therefore

$$\boxed{P(y|\phi) = \exp\left(\eta^T \vec{y} + \log\frac{1}{1+\sum e^{\eta_i}}\right), \text{ with } \begin{array}{l} b(y)=1 \\ T(y) = y \quad a(\eta) = \frac{1}{1+\sum e^{\eta_i}} \end{array}}$$ ∎

We also note that $\phi_i = \phi_K e^{\eta_i} = \dfrac{e^{\eta_i}}{1+\sum e^{\eta_i}}$ is softmax regression