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.

Beta Function: $B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\,d\theta = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$$\Gamma(x+1) = x\Gamma(x)$$

mean of θ: $E(\theta) = \int_0^1 \theta\, P(\theta; a, b)\, d\theta$

$$= \int_0^1 \theta \left(\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\right) d\theta$$

$$= \frac{1}{B(a,b)}\int_0^1 \theta^a (1-\theta)^{b-1}\, d\theta$$

$$= \frac{B(a+1, b)}{B(a,b)}$$

$$= \left[\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}\right]\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]$$

■

$$= \frac{a}{a+b}$$

$$\text{Var}[\theta] = E\left[(\theta - E(\theta))^2\right]$$

$$= E[\theta^2] - E[\theta]^2$$

$E[\theta^2]$ can be computed w/ similar process as above:

$$E[\theta^2] = \frac{a(a+1)}{(a+b)(a+b+1)}$$

Thus, $$\text{Var}[\theta] = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2}$$

$$= \frac{ab}{(a+b)^2(a+b+1)}$$

mode: $$\nabla_\theta P(\theta; a, b) = 0$$

$$\nabla_\theta \left[\theta^{a-1}(1-\theta)^{b-1}\right] = 0$$

$$(a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2} = 0$$

$$(a-1)\theta^{a-2}(1-\theta)^{b-1} = (b-1)\theta^{a-1}(1-\theta)^{b-2}$$

$$(a-1)(1-\theta) = (b-1)\theta$$

$$(a+b-2)\theta = a-1$$

$$\theta^* = \frac{a-1}{a+b-2}$$

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

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

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).

$$P(\vec{y};\vec{\eta}) = b(\vec{y})\exp\left(\vec{\eta}^T T(\vec{y}) - a(\vec{\eta})\right)$$ ... exponential family

$$\mathrm{Cat}(\vec{x}|\vec{\mu}) = \prod_{i=1}^{K} \mu_i^{x_i} = \exp\left[\log\left(\prod_{i=1}^{K} \mu_i^{x_i}\right)\right]$$

$$= \exp\left(\sum_{i=1}^{K} \log\left(\mu_i^{x_i}\right)\right)$$ by property of logs

$$= \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right)$$

$$\sum_{i=1}^{K} \mu_i = 1 \quad \& \quad \sum_{i=1}^{K} x_i = 1$$

$$\mu_K = 1 - \sum_{i=1}^{K-1} \mu_i \qquad x_K = 1 - \sum_{i=1}^{K-1} x_i$$

$$\mathrm{Cat}(\vec{x}|\vec{\mu}) = \exp\left(\sum_{i=1}^{K} x_i \log(\mu_i)\right)$$

$$= \exp\left(\sum_{i=1}^{K-1} x_i \log(\mu_i) + x_K \log(\mu_K)\right)$$

$$= \exp\left[\sum_{i=1}^{K-1} x_i \log\left(\frac{\mu_i}{\mu_K}\right) + \log(\mu_K)\right]$$

■

$$\vec{\eta} = \begin{bmatrix} \log\left(\frac{\mu_1}{\mu_k}\right) \\ \vdots \\ \log\left(\frac{\mu_{k-1}}{\mu_k}\right) \end{bmatrix}$$

$$\mu_i = \mu_k e^{\eta}$$

$$\mu_k = 1 - \sum_{i=1}^{k-1} \mu_i$$

$$= 1 - \mu_k \sum_{i=1}^{k-1} e^{\eta_i}$$

$$= \frac{1}{1 + \sum_{i=1}^{k-1} e^{\eta_i}}$$

$$\Rightarrow \mu_i = \mu_k e^{\eta_i} = \frac{e^{\eta_i}}{1 + \sum_{i=1}^{k-1} e^{\eta_i}}$$

$$\Rightarrow b(\vec{n}) = 1$$

$$T(\vec{x}) = \vec{x}$$

$$a(\vec{\eta}) = -\log(\mu_k) = \log\left(1 + \sum_{i=1}^{k-1} \vec{\eta}\right)$$