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 func: $B(a,b) = \Gamma(a)\,\Gamma(b) / \Gamma(a+b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\,d\theta$.

Recall the gamma function's property $\Gamma(x+1) = x\,\Gamma(x)$.

The mean of θ is

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

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

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

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

■

The variance of θ is $\mathrm{var}(\theta) = E(\theta^2) - E(\theta)^2$.

$$E(\theta^2) = \int_0^1 \theta^2 \left(\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\right) d\theta$$

$$= \frac{1}{B(a,b)} \int_0^1 \theta^{a+1}(1-\theta)^{b-1} d\theta = \frac{B(a+2,b)}{B(a,b)}$$

$$= \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

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

So $\mathrm{var}(\theta) = \boxed{\dfrac{ab}{(a+b)^2(a+b+1)}}$.

The mode of θ can be calculating using a MAP estimate, giving

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

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).

We write

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

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

Observe $\mu_k = 1 - \sum_{i=1}^{k-1} \mu_i$ and $x_k = 1 - \sum_{i=1}^{k-1} x_i$.

We therefore split our summation into

$$\mathrm{Cat}(x|\mu) = \exp\left(\sum_{i=1}^{k-1} x_i \log(\mu_i) + x_k \log(\mu_k)\right)$$

$$= \exp\left(\sum_{i=1}^{k-1} x_i \log(\mu_i) + \left(1 - \sum_{i=1}^{k-1} x_i\right) \log(\mu_k)\right)$$

$$= \exp\left(\sum_{i=1}^{k-1} x_i \log\left(\frac{\mu_i}{\mu_k}\right) + \log(\mu_k)\right).$$

■

Let vector $v = \left(\log\left(\frac{\mu_1}{\mu_k}\right) \dots \log\left(\frac{\mu_{k-1}}{\mu_k}\right) \right)$.

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

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

$$= \frac{1}{1 + \sum_{i=1}^{k-1} e^{v^T}}.$$

We conclude that $\text{Cat}(x|\mu)$ is in the exponential function, where $\mu = S(v^T)$ and $S(v^T)$ is the softmax function.