Justin Jiang
Math189R SU2020
Homework 3
Wednesday, June 17, 2020

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.

By computing the mean of θ, we see that:

$\mathbb{E}[\theta] = \int_0^1 \theta \mathbb{P}(\theta; a, b) d\theta = \int_0^1 \theta (\frac{1}{B(a,b)} \theta^{a-1} (1 - \theta)^{b-1}) d\theta$

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

Integrate, such that: $\frac{B(a+1,b)}{B(a,b)}$

Plug in the gamma function to get the mean of theta:

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

To compute the variance, we need to find $\mathbb{E}[\theta^2]$ using the same integration method as above. We find that:

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

Thus, the variance is, after a bit of fraction simplifying:

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

To find the mode, we need to find the gradient of the P function with respect to a , and then we simplify, such that the mode:

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

∎

2 (Murphy 9) Show that the multinoulli 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 multinoulli logistic regression (softmax regression).

I had to refer to the solutions for this question. Again, because there is no point in me rewriting the work on the solutions page, I will summarize:

To prove the problem statement, we need to apply the form of the exponential family as well as the property of logarithms, because we need to rewrite the distribution to have both a log and an expoential. The result you get from that is:

$Cat(\mathbf{x}|\mu) = \prod_K^{i=1} \mu_i^{x_i} = exp[log(\prod_{i=1}^{K} \mu_i^{x_i}]$

Simplify to get: $= exp(\sum_{i=1}^{K} x_i log(\mu_i))$

Following the solutions, we can split up the above summation:

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

Let vector n be defined as it is in the solutions and make the substitution such that:

$= \frac{1}{1+\sum_{K-1}^{i=1} e^{\mathbf{n}_i}}$

Thus, we can write it in the form of the expoential family:

$b(\mathbf{n}) = 1$

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

$a(\mathbf{n}) = -log(\mu_K) = log(1 + \sum_{i=1} K - 1e^{\mathbf{n}_i})$ QED

∎