Name: ______________________________ NetID: ____________

STATISTICS AND DATA SCIENCE 355 / 555

Introductory Machine Learning

Quiz 2 (practice), Thursday, October 31, 2019

No notes or computers are allowed

1. *Bayesian inference* (5 points)

   Suppose that X is a random variable denoting a coin flip, where $X = 1$ is "heads" with probability θ, and $X = 0$ is "tails" with probability $1 - \theta$. We want to carry out Bayesian inference on θ, using a Beta(α, α) prior $p(\theta)$. Suppose that we flip the coin five times and observe $X_1, X_2, \ldots, X_5$, with three heads and two tails.

   (a) Give the formula for $p(\theta)$. You can state it up to a constant of proporationality.

   $$p(\theta) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}\,\theta^{\alpha-1}(1-\theta)^{\alpha-1} \propto \theta^{\alpha-1}(1-\theta)^{\alpha-1}$$

   (b) Give an expression for the likelihood of the data given θ.

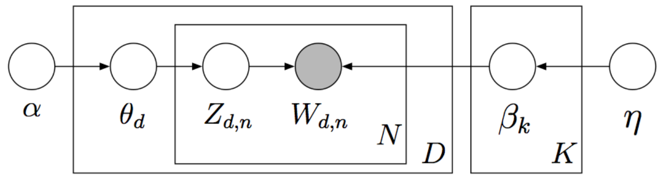
   $$p(X \,|\, \theta) = \theta^3(1-\theta)^2$$

   (c) What is the posterior distribution $p(\theta \,|\, X_1, \ldots, X_5)$?

   $$\text{Beta}(3 + \alpha, 2 + \alpha)$$

2. *Topic modeling* (5 points)

The latent Dirichlet allocation topic model is represented by the diagram



where $\theta_d \sim \text{Dirichlet}(\alpha)$ are the per-document topic proportions, $Z_{d,n} \sim \text{Multinomial}(\theta_d)$ are the per-word topic assignments, $W_{d,n} \sim \text{Multinomial}(\beta_{Z_{d,n}})$ are the observed words, and $\beta_k \sim \text{Dirichlet}(\eta)$ are the topics.

Circle the correct answers:

$\boxed{\text{TRUE}}$ FALSE (1) The model is generative, and can assign a probability to documents that are not in the training data.

TRUE $\boxed{\text{FALSE}}$ (2) According to the model, each document is generated by a single topic.

TRUE $\boxed{\text{FALSE}}$ (3) According to the model, the words are generated independently.

$\boxed{\text{TRUE}}$ FALSE (4) As α decreases from one toward zero, the topic proportions vector θ_d tends to have small values for a larger number of topics.

TRUE $\boxed{\text{FALSE}}$ (5) The Gibbs sampling algorithm chooses the most probable topic $Z_{d,n}$ for a selected word $W_{d,n}$ while holding all of the other Z values fixed.