Coursework **Spring 2019**
Statistical Pattern Recognition (COMP136) **Roni Khardon**
Department of Computer Science, Tufts University **Rishit Sheth**

Written Assignment 1

This assignment is due by the **start** of class on Tuesday, January 29.

1. (15 pts.) Solve problem 2.6 in the textbook. Please consult the solutions of problems 1.17 and 2.5 that are available on the text's web page.

2. (10 pts.) Solve the first part of 2.10 in the textbook, i.e., solve only $E[\mu_j]$.

3. Bayesian unigram model for text learning.

   We have a simple probabilistic model for generating an N-word document, $Y = (y_1, y_2, \ldots, y_N)$. Assuming a vocabulary of K words $w_1, \ldots, w_K$, we sample each word y_i of the document i.i.d. from a discrete distribution over the vocabulary with parameter $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$; hence, $p(y_i = w_j|\boldsymbol{\mu}) = \mu_j$. The discrete distribution is specified by Eq. (B.54) in the textbook Appendix.

   We will empirically contrast the maximum likelihood, maximum a posteriori, and Bayesian approaches for this model in the first programming project. For now, we focus on the calculations associated with the Bayesian approach. As shown in Section 2.2 of the textbook, we can use a Dirichlet prior $p(\boldsymbol{\mu}|\boldsymbol{\alpha})$ to yield a Dirichlet posterior $p(\boldsymbol{\mu}|Y)$ given in Eq. (2.41).

   (i) (10 pts.) Calculate the predictive distribution for this problem. Specifically, show

   $$p(y_* = w_j|Y) = \int_{\boldsymbol{\mu}} p(y_* = w_j|\boldsymbol{\mu})p(\boldsymbol{\mu}|Y)\ d\boldsymbol{\mu} = \frac{m_j + \alpha_j}{N + \alpha_0}.$$

   (ii) (10 pts.) Calculate the evidence function. Specifically, show

   $$p(Y|\boldsymbol{\alpha}) = \int_{\boldsymbol{\mu}} p(Y|\boldsymbol{\mu})p(\boldsymbol{\mu}|\boldsymbol{\alpha})\ d\boldsymbol{\mu} = \frac{\Gamma(\alpha_0)\prod_{j=1}^{K}\Gamma(\alpha_j + m_j)}{\Gamma(\alpha_0 + N)\prod_{j=1}^{K}\Gamma(\alpha_j)}.$$

   Later in the course, we'll see how the evidence function can be used to select a suitable value for $\boldsymbol{\alpha}$.