Coursework	**Fall 2012**
Statistical Pattern Recognition (COMP136)	**Roni Khardon**
Department of Computer Science, Tufts University	

Assignment 1

This assignment is due by Tuesday September 18.

1. Solve problems 1.6, 1.13, 1.40, 2.6, 2.10, 2.38 in the textbook.

   Notes: (1) For 1.6 show in addition that the inverse is not true, i.e. $cov(x, y) = 0$ does not imply that x, y are independent. (2) For 2.6: please consult the solutions of problems 1.17 and 2.5 that are available on the text's web page. (3) For 2.10 solve only $E[\mu_j]$. (4) Problem 2.38 repeats and completes some calculations from class.

2. Bayesian Unigram model for text learning.

   We have a simple probabilistic model for text generation using a vocabulary of K words $w_1, \ldots, w_K$, specified by a multinomial distribution over the words with parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$, so that $Pr[\text{next word is } w_j] = \mu_j$.

   To generate a document with N word tokens using this model we sample each token independently from $\boldsymbol{\mu}$. As shown in Section 2.2 of the textbook we can use a Dirichlet prior for this problem to yield a Dirichlet posterior $Pr(\boldsymbol{\mu}|\text{Data})$, given in Eq (2.41), where the prior is specified by the vector of counts $\boldsymbol{\alpha}$.

   Our data is one document of length N, given by the token sequence Data= $x_1, \ldots, x_N$, where each x_i is some word w_j in the vocabulary.

   (i) Calculate the predictive distribution for this problem. Specifically calculate

   $$Pr[\text{next word is } w_j|\text{Data}] = \int_{\boldsymbol{\mu}} Pr(\boldsymbol{\mu}|\text{Data}) Pr([\text{next word is } w_j|\boldsymbol{\mu})\; d\boldsymbol{\mu}.$$

   (ii) Calculate the evidence function

   $$Pr[\text{Data}|\boldsymbol{\alpha}] = \int_{\boldsymbol{\mu}} Pr(\boldsymbol{\mu}|\boldsymbol{\alpha}) Pr(\text{Data}|\boldsymbol{\mu})\; d\boldsymbol{\mu}.$$

   As we discuss later in the course the evidence function can be used to select a suitable value for $\boldsymbol{\alpha}$. For now we just focus on the calculation.