Note:

1. Due 11:59PM, December 15, 2016.
2. Electronic submission to your instructor's email.
3. You are VERY MUCH encouraged to form teams to discuss proofs and program algorithms. If so, please acknowledge your teammate(s)' contributions at the beginning of your submitted homework. You must independently write your homework based on your own understanding.
4. Choose any programming language you like, R, Python, Matlab, C/C++, Julia, etc.

Examples and Implementations

[**Bayesian Linear Regression with Automatic Relevance Determination**]
Suppose we have a data set of the form $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$, where $y \in R$ and $\boldsymbol{x} \in R^d$. We assume d is large and not all dimensions of $\boldsymbol{x}$ are useful for predicting y. Consider the following regression model for this problem:

$$y_i \sim^{independent} Normal(x_i' w, \lambda^{-1}), \qquad w \sim Normal(0, \lambda^{-1} diag(\alpha_1, \dots, \alpha_d)^{-1}),$$
$$\alpha_k \sim^{iid} Gamma(a_0, b_0), \qquad \lambda \sim Gamma(e_0, f_0).$$

Use the density function: $Gamma(\eta \mid \tau_1, \tau_2) = \frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} \eta^{\tau_1 - 1} \exp\{-\tau_2 \eta\}$. In this homework, you will derive a variational inference algorithm for approximating the posterior distribution with

$$q(w, \alpha_1, \dots, \alpha_d, \lambda) \approx p(w, \alpha_1, \dots, \alpha_d, \lambda \mid y, \boldsymbol{x})$$

a) Using the factorization $q(w, \alpha_1, \dots, \alpha_d, \lambda) = q(w) q(\lambda) \prod_k q(\alpha_k)$, derive the optimal form of each q distribution. Use these optimal q distributions to derive a variational inference algorithm for approximating the posterior.

b) Summarize the algorithm derived in Part (a) using pseudo-code.

c) Using these q distributions, calculate the variational objective function. (In practice, you will need to evaluate this function to assess the convergence of your algorithm.)

Hint:

1) More background can be found at **Appendix A** of Blei DM, Kucukelbir A and McAuliffe JD (2016). Variational Inference: A Review for Statisticians.
2) A very good read on **Automatic Relevance Determination**: Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1(Jun):211–244.