Advanced Topics in Machine Learning (600.692)
Homework 1: Linear Algebra, Optimization and Statistics Review

Instructor: René Vidal

Due Date: 02/21/2014, 11.59PM Eastern

READING MATERIAL: The Matrix Cookbook, Appendix A and Appendix B.1 of GPCA book.

1. **Properties of Symmetric Matrices.** Let $S \in \mathbb{R}^{n \times n}$ be a real symmetric matrix. Show that:
   (a) All the eigenvalues of S are real, i.e., $\sigma(S) \subset \mathbb{R}$.
   (b) Let (λ, v) be an eigenvalue-eigenvector pair. If $\lambda_i \neq \lambda_j$, then $v_i \perp v_j$; i.e., eigenvectors corresponding to distinct eigenvalues are orthogonal.
   (c) There always exist n orthonormal eigenvectors of S, which form a basis of $\mathbb{R}^n$.
   (d) S is positive definite (positive semidefinite) if and only if all of its eigenvalues are positive (non-negative), i.e., $S \succ 0$ ($S \succeq 0$), iff $\forall i = 1, 2, \ldots, n$, $\lambda_i > 0$ ($\lambda_i \geq 0$).
   (e) If $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the sorted eigenvalues of S , then $\max_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^\top S \boldsymbol{x} = \lambda_1$ and $\min_{\|x\|_2=1} \boldsymbol{x}^\top S \boldsymbol{x} = \lambda_n$.

2. **Pseudo-Inverse of a Matrix.**
   (a) Let $A = U_r \Sigma_r V_r^\top$ be the compact SVD of a matrix A of rank r. Show that the pseudo-inverse of A is given by $A^\dagger = V_r \Sigma_r^{-1} U_r^\top$.
   (b) Consider the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$, where the matrix $A \in \mathbb{R}^{m \times n}$ is of rank $r = \text{rank}(A) = \min\{m, n\}$. Show that $\boldsymbol{x}^* = A^\dagger \boldsymbol{b}$ minimizes $\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2$, where $A^\dagger$ is the pseudo-inverse of A. When is $\boldsymbol{x}^*$ the unique solution?

3. **Convex Sets and Functions.** Show that:
   (a) The intersection of two convex sets is convex.
   (b) The convex hull of a set is convex and the convex hull of a convex set is the set itself.
   (c) Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function defined over a convex domain $\mathcal{X} \subseteq \mathbb{R}^n$. Show that for any $c \in \mathbb{R}$, the set $\{\boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}) \leq c\}$ is convex.
   (d) A convex function is pseudoconvex and quasi convex.

4. **Derivatives of Traces and Logarithms.** Show that
   (a) $\frac{\partial}{\partial X} \text{trace}(AX^{-1}B) = -(X^{-1}BAX^{-1})^\top$
   (b) $\frac{\partial}{\partial X} \text{trace}(A \otimes X) = \text{trace}(A)I$
   (c) $\frac{\partial}{\partial X} \text{trace}(X^\top BX) = BX + B^\top X$
   (d) $\frac{\partial}{\partial X} \text{trace}(XBX^\top) = XB^\top + XB$
   (e) $\frac{\partial}{\partial X} \log|\det(X)| = (X^{-1})^\top$

5. **Maximum Likelihood Estimates of the Parameters of a Gaussian.** Let $\boldsymbol{x} \in \mathbb{R}^D$ be a random vector. Let $\boldsymbol{\mu}_{\boldsymbol{x}} = \mathbb{E}(\boldsymbol{x}) \in \mathbb{R}^D$ and $\Sigma_{\boldsymbol{x}} = \mathbb{E}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\top \in \mathbb{R}^{D \times D}$ be, respectively, the mean and the covariance of $\boldsymbol{x}$. Show that the maximum likelihood estimates of $\boldsymbol{\mu}_{\boldsymbol{x}}$ and $\Sigma_{\boldsymbol{x}}$ are, respectively, given by

$$\widehat{\boldsymbol{\mu}}_N \doteq \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{x}_j \quad \text{and} \quad \widehat{\Sigma}_N \doteq \frac{1}{N} \sum_{j=1}^{N} (\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}}_N)(\boldsymbol{x}_j - \widehat{\boldsymbol{\mu}}_N)^\top. \tag{1}$$

Submission instructions. Please send an email to advml2014@gmail.com with subject 600.692:HW1 and attachment `firstname-lastname-hw1-learning14.zip` or `firstname-lastname-hw1-learning14.tar.gz`. The attachment should have the following content:

1. For analytical questions, please submit a file called `hw1.pdf` containing your answers to each one of the analytical questions. If at all possible, you should generate this file using the latex template hw1-learning14.tex. If not possible, you may use another editor, or scan your handwritten solutions. But note that you must submit a single PDF file with all your answers.

2. For coding questions, please submit a file called `README`, which contains instructions on how to run your code. Please use separate directories for each coding problem. Each directory should contain all the functions and scripts you are asked to write in separate files. For example, for HW1 the structure of what you should submit could look like

   (a) `README`

   (b) `hw1.pdf`

   (c) `hw1q3`: `hw1q3c.m`, `hw1q3e.m`

   (d) `hw1q4`: `hw1q4b.m`, `hw1q4c.m`

   The TA will run your scripts to generate the results. Thus, your script should include all needed plotting commands so that figures pop up automatically. Please make sure that the figure numbers match those you describe in `hw1.pdf`. You do not need to submit input or output images. The output images should be automatically generated by your scripts so that the TA can see the results by just running the scripts. In writing your code, you should assume that the TA will place the input images in the directory that is relevant to the question solved by your script. Also, make sure to comment your code properly.