ECE M146 Discussion 4
Introduction to Machine Learning
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen, Ruiyi (John) Wu

1. Introduction to optimization problem

   (a) Convex sets.

   (b) Convex functions.

   (c) Optimization problem in standard form.

      - Convex optimization.

   (d) Globally and locally optimal.

   (e) Duality.

      - Lagrange dual problem.
      - Geometric interpretation.
      - KKT conditions.

   Notes: One can read the book Convex Optimization by Boyd and Vandenberghe (freely available on-line) for more extensive coverage of the above topics.

2. Find the dual problem of the following Quadratic program

$$\begin{array}{ll} \text{minimize}_x & x^T P x \\ \text{subject to} & Ax \leq b \end{array}$$

Assume $P \in \mathcal{S}^n_{++}$.

3. Quadratic program example Consider the objective function

$$J(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2.$$

Find the optimal x that minimize $J(x)$ under the following constrains:

(a) No constrain.

(b) $x_1 + x_2 + 2 = 0$.

(c) $x_1 + x_2 + 2 \leq 0$.

(d) $x_1 + x_2 + 2 \geq 0$.

4. **Multi-class Classification Least Squares** In this section, you will determine the parameter matrix $\mathbf{W} \in \mathbb{R}^{m \times p}$ for the Multi-class Least Squares classification.

   Given a data matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ and target matrix $\mathbf{T} \in \mathbb{R}^{n \times p}$, the sum-of-squares error function can be written as

   $$Er(\mathbf{W}) = \mathbf{Tr}\{(\mathbf{XW} - \mathbf{T})^T(\mathbf{XW} - \mathbf{T})\}$$

   where **Tr** is the trace of a matrix. You can assume that $\mathbf{X}$ has full rank.

   We will solve this problem by setting the derivative with respect to $\mathbf{W}$ to be zero and solve for $\mathbf{W}$. To do this we must first know some matrix derivative properties.

   (a) Let $\mathbf{A}, \mathbf{Z}$ be two matrices. Prove

   $$\frac{d\mathbf{Tr}(\mathbf{AZ})}{d\mathbf{Z}} = \mathbf{A}^T$$

   (b) Let $\mathbf{A}, \mathbf{Z}$ be two matrices. Prove

   $$\frac{d\mathbf{Tr}(\mathbf{ZAZ}^T)}{d\mathbf{Z}} = \mathbf{ZA}^T + \mathbf{ZA}$$

   (c) Now, we can take the derivative of $Er(\mathbf{W})$ and set it to zero. Show that this results in

   $$\mathbf{W} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}$$