# Homework 3

ECE 602 – Introduction to Optimization
Due: March 18

**Exercise 1** (Lagrangian duality)

The relative entropy between two vectors $x, y \in \mathbf{R}^n_{++}$ is defined as

$$\sum_{k=1}^{n} x_k \log\left(\frac{x_k}{y_k}\right) .$$

This is a convex function, jointly in $x$ and $y$. In the following problem we calculate the vector $x$ that minimizes the relative entropy with a given vector $y$, subject to the following constraints on $x$:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \sum_{k=1}^{n} x_k \log\left(\frac{x_k}{y_k}\right) \\ \text{subject to} & Ax = b \\ & \mathbf{1}^T x = 1 \\ & x \succeq \mathbf{0} . \end{array}$$

The given parameters are $y \in \mathbf{R}^n_{++}$, $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$. Note that $\mathbf{1}^T x = 1$ and $x \geq \mathbf{0}$ mean that $x$ is a probability vector. Derive the Lagrange dual of this problem and simplify it to get

$$\underset{\nu}{\text{maximize}} \quad b^T \nu - \log \sum_{k=1}^{n} y_k e^{a_k^T \nu} ,$$

where $a_k$ is the $k$-th column of $A$ and $\nu \in \mathbf{R}^m$ is the Lagrange multiplier associated with the equality constraint.

**Exercise 2** (Support Vector Machine)

Consider a set of training data $(x_1, y_1), \ldots, (x_n, y_n)$, with $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$. In *supervised classification*, $y_i$ usually represents the class that $x_i$ belongs to, and one's goal is to find $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$ which define a hyperplane $w^T x = b$ that separates the two classes, i.e.,

$$w^T x_i - b \geq 1, \quad \text{for all } x_i \in X_{+1} := \{x_i \mid y_i = 1\},$$

while

$$w^T x_i - b \leq 1, \quad \text{for all } x_i \in X_{-1} := \{x_i \mid y_i = -1\}.$$

Note that the above approach uses an affine classification function $f(x) = w^T x - b$, in which case it belongs to the family of linear *Support Vector Machines* (SVM).

In many practical cases, the convex hulls of $X_{+1}$ and $X_{-1}$ admit some overlap and, as a result, the separating hyperplane cannot be defined. One way to overcome this difficulty is to introduce a vector of *slack variables* $\zeta \in \mathbb{R}^n_{++}$ which can be used to “relax" the classification constraints as follows:

$$w^T x_i - b \geq 1 - \zeta_i, \quad \forall\, x_i \in X_{+1} \quad \text{and} \quad w^T x_i - b \leq -1 + \zeta_i, \quad \forall\, x_i \in X_{-1}$$

or, alternatively,

$$y_i(w^T x_i - b) \geq 1 - \zeta_i, \quad \forall\, x_i.$$

Normally, we expect the values of $\zeta$ to be dominated by zeros, thus implying that $\sum_{i=1}^n \zeta_i$ is relatively small. In this case, the problem of SVM classification can be formulated as given by

$$\begin{aligned} \min_{w,b,\zeta} \quad & \frac{1}{2}||w||_2^2 + \lambda \sum_{i=1}^n \zeta_i \\ \text{s.t.} \quad & y_i(x_i^T w + b) \geq 1 - \zeta_i, \; i = 1, 2, \ldots, n, \\ & \zeta \succeq 0, \end{aligned}$$

where $\lambda > 0$ is a user-defined scalar parameter (e.g., $\lambda = 1$).

1. Write the above optimization problem in a standard form. What type of optimization problem is that?

2. Derive the dual optimization problem corresponding to the primal problem. What type of optimization problem is that?

3. Implement the dual problem in CVX and apply it to the medical diagnostic data available at `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html#breast-cancer`.

4. Given the dual optimal solution, derive a closed-form solution to the primal problem and compute the resulting $w^*$ and $b^*$.

5. Split the data set randomly into a training set ($\approx 90\%$ of the total number of points) and a validation set ($\approx 10\%$ of the total number of points). Train the SVM classifier using the former set and report the relative number of misclassified $x_i$ in the latter set.

**Exercise 3** (KKT optimality conditions)

Consider the equality constrained least-squares problem of the form

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & Cx = h, \end{array}$$

where $A \in \mathbf{R}^{m \times n}$, with $\operatorname{rank} A = n$, and $C \in \mathbf{R}^{p \times n}$, with $\operatorname{rank} C = p$. Specify the KKT conditions and derive expressions for the primal and dual optimal solutions.