Faculty of Arts and Sciences
Department of Computer Science
CMPS 351 – Numerical Optimization
Assignment 8 – Due Mar 18, 2019

Background reading / reference material:

- Duality and optimality conditions. Sections 5.1, 5.2, 5.3 and 5.5 of Reference 2.
- Background on classification: section 8.6 (pp. 422–431) in reference 2.

1. Problem in R Consider the problem:

$$\begin{array}{ll} \text{minimize} & x^2+1 \\ \text{subject to} & (x-2)(x-4) \leq 0 \end{array}$$

- Derive an expression for the dual function $g(\lambda)$. Plot $g(\lambda)$.
- Solve the dual problem to find λ^* and g^*. Verify that this problem has strong duality.

2. Dual in R^n. Consider the following optimization problem where $x \in R^n$, $\mathbf{1}$ is the one vector in R^n, and b is a scalar.

$$\begin{array}{ll} \text{minimize} & p^t x \\ \text{subject to} & 0 \leq x \leq 1 \\ & \mathbf{1}^t x = b \end{array}$$

- Let ν denote the multiplier of the equality constraint, λ denote the multipliers of the inequalities $x \leq 1$, and λ_2 denote the multipliers of the non-negativity constraints. Write the Lagrangian of this problem.
- Show that the dual may be written as:

$$\begin{array}{ll} \text{maximize} & b\nu + \mathbf{1}^t\lambda \\ \text{subject to} & \nu\mathbf{1} + \lambda \geq -p \\ & \lambda \geq 0 \end{array}$$

3. Dual of a quadratic problem. Consider the following least squares solution of the set of underdetermined linear equations.

$$\begin{array}{ll} \text{minimize} & x^t x \\ \text{subject to} & Ax = b \end{array}$$

where $A \in R^{p\times n}$. Derive the dual problem.

4. Piecewise-linear minimization. Consider the problem:

$$\text{minimize } \max_{i=1,\dots,m}(a_i^t x - b_i)$$

- Show that it can be expressed as the smooth linear problem:

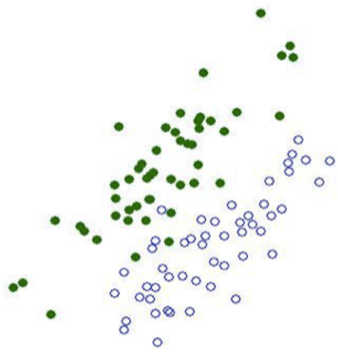
$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} A & -\mathbf{1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq b \end{array} \tag{1}$$

- Show that the dual of (1) may be written as:

$$\begin{array}{ll} \text{maximize} & -b^t\lambda \\ \text{subject to} & A^t\lambda = 0 \\ & \mathbf{1}^t\lambda = 1 \\ & \lambda \geq 0 \end{array} \tag{2}$$

- Write the KKT conditions of the dual problem.

5. Classification. Given two sets of points in R^n, the classification problem asks to which set a new point belongs. The answer may be found by first finding a hyperplane $a^t x - b$ that "best" separates the two sets, being positive on the first set and negative on the second. A new point can be classified by checking which side of the hyperplane it lies on. This problem appears in a wide variety of contexts in machine learning, pattern recognition, data mining, and related domains where one seeks to learn from a set of observations or samples. The figure below shows an illustration of the problem in R^2.



- When the two sets are separable by a hyperplane $a^t x - b$, the problem can be formulated as that of finding the "thickest slab" that achieves the separation. Formulate this problem as a quadratic optimization problem of the form:

$$\begin{aligned} \underset{a,b}{\text{minimize}} \quad & \frac{1}{2} a^t a \\ \text{subject to} \quad & D(Xa - b\mathbf{1}) \geq \mathbf{1} \end{aligned}$$

  and use `scipy.optimize.minimize` or `cvxpy` to solve it for the data posted on moodle. Show your solution as the pair $a^t x - b = \pm 1$.

- Add a data point that prevents the sets from being separable and verify (with `scipy.optimize` or `cvxpy`) that the problem above becomes infeasible.

- In general the two sets of points cannot be separated by a hyperplane, so we seek to find the best classifier that maximizes the width of a slab that separates the two point sets while minimizing the amount of misclassification (as measured, for example, by the violation of the constraints above). A weighted combination of these objectives can be used. Formulate the problem as:

$$\begin{aligned} \underset{a,b,u}{\text{minimize}} \quad & \frac{1}{2} a^t a + \alpha \mathbf{1}^t u \\ \text{subject to} \quad & D(Xa - b\mathbf{1}) \geq \mathbf{1} - u \\ & u \geq 0 \end{aligned}$$

  and solve it to generate the tradeoff curve of Pareto optimal points for the given data. Comment on the shape of the curve. Plot the solution. Does it make sense?

6. Dual. Consider the dual of the problem above

- Write the Lagrangian:

$$L(a, b, u, \lambda, \sigma) = \tfrac{1}{2} a^t a - \lambda^t D X a + \lambda^t D \mathbf{1} b + (\alpha \mathbf{1}^t - \lambda^t - \sigma^t) u + \lambda^t \mathbf{1}$$

  where λ and σ are the Lagrange multipliers of the classification and non-negativity constraints, respectively.
- Express the dual problem as:

$$\begin{aligned} \underset{\lambda}{\text{maximize}} \quad & -\tfrac{1}{2} \lambda^t D X X^t D \lambda + \mathbf{1}^t \lambda \\ \text{subject to} \quad & \mathbf{1}^t D \lambda = 0 \\ & 0 \leq \lambda \leq \alpha \mathbf{1} \end{aligned}$$

  (Hint: The dual variables σ can be eliminated)
- Choose a value of α (say, $\alpha = 0.5$) and solve the primal and dual problems. Do you get the same optimal value of the objective value? Comment.