Math 160 (De Loera) **Homework 3**
Due date: May 16, 2016

INSTRUCTIONS

All homeworks will have many problems, both theoretical and practical.

Programming exercises need to be submitted via SMARTSITE using the assignment boxes NOT DROPBOX.

Other methods of submission without prior approval will receive zero points.

Write legibly preferably using word processing if your hand-writing is unclear. Be organized and use the notation appropriately. Show your work on every problem. Correct answers with no support work will not receive full credit.

1. PROJECT 3: Finding out the top features that determine what is an "spam email"

   You will write an algorithm in SCIP for deciding what are the most important features that distinguishe an spam email from regular email. You can download the data files from `http://archive.ics.uci.edu/ml/machine-learning-databases/spambase/`

   In there you can find the data file spambase.data that contains information on 4601 emails (1813 are Spam!!) Each row contains 58 attributes: 58 (57 continuous, 1 nominal class label). The last attribute in the very last column of 'spambase.data' denotes whether the e-mail was considered spam (1) or not (0).

   Most of the attributes indicate whether a particular word or character was frequently occurring in the e-mail. The run-length attributes (55-57) measure the length of sequences of consecutive capital letters. For more information see the documentation.

   GOAL use the LASSO method to figure out what are the most significant of the characteristics that define the spam emails (e.g., maybe number of consecutive capital letters is overly excessive?). Can you run LASSO directly? You can rewrite the LASSO convex model as a linear program by adding some extra variables!

2. A number of communities want to build a central warehouse. The location of each of the communities is described by city A $(3, 9)$, city B $(10, 6)$, city C $(0, 12)$ and city D $(4, 9)$. Goods will be delivered to cities by plane, the price of the delivery is proportional to the distance between the warehouse and the city. City A will need 20 deliveries, City B 14, City C 8, and City D 24. GOAL: Find the location of the warehouse (i.e., explicit coordinates), so that the total cost of the deliveries is minimized.

3. Consider the optimization problem

$$\min \ 2x_1 \arctan(x_1) - \ln(x_1^2 + 1) + x_2^4 + (x_3 - 1)^2$$

$$\textbf{subject to: } x_1^2 + x_2^2 + x_3^2 - 4 \leq 0, \quad x_1 \geq 0$$

   .

   Prove that this problem is a convex optimization problem. Use the Karush -Kuhn-Tucker theorem to find an optimal solution.

4. Let F be a closed compact convex set in R^n, show that for all u there is a unique point in F which is closest to u.

5. Consider the optimization problem

$$\min\ x_3$$
$$\textbf{subject to:} (x_1 - 1)^2 + x_2^2 + x_3^2 - 1 \geq 0,$$
$$(x_1 + 1)^2 + x_2^2 + x_3^2 - 1 \geq 0,$$
$$x_1^2 + x_2^2 + x_3^2 - 4 \leq 0$$

.

(a,b,c) and (d,e,f) feasible solutions; prove t(a,b,c) + (1-t)(d,e,f)is feasible

- Is the feasible region convex?
- What is the convex hull of the feasible region?
- Solve the optimization problem over the convex hull.

6. Suppose you have a model in which all variables are meant to be binary, i.e., can only be 0 or 1. Show that it can be expressed as non-linear optimization problem with *real* variables.

7. Find the maximum value of the function $f(x, y) = x^2 + y^4 + xy$ inside the convex hull of the points $(-1, 1)$, $(-1, 2)$, $(-2, 2)$, $(-3, 1)$. Is the maximum unique? What about the minimum value of $f(x, y)$? Does this function have a unique global minimum in R^2.