Problem set 5
Convex Optimization (Winter 2018)

Due Friday, February 9th at 5pm

- Upload both a pdf writeup of your work and a zip file containing your code to Canvas by the due date. Late homeworks will be penalized 25% per day and will not be accepted after the next Monday at 5pm unless you have made arrangements with us well before the due date. We strongly encourage you to typeset your solutions.

- You may collaborate with other students to complete this assignment, but all submitted work should be yours. When you go to write down your solutions, please do so on your own.

1 Regularized Regression Problems

In class, we derived dual problems for minimizing classification losses. Here, we will look at regression problems with regularization. In a linear regression model, we would like to find the optimal linear estimator of responses \( y_1, \ldots, y_m \) as a function of input points (feature/covariate vectors) \( x_1, \ldots, x_m \in \mathbb{R}^n \). In particular, we would like to find \( w \in \mathbb{R}^n \) such that \( w^\top x_i \) is closest to the response \( y_i \). and we quantify this by minimizing \( \phi(y_i - w^\top x_i) \), which after adding an \( \ell_2 \) regularization term corresponds to the following unconstrained convex optimization problem:

\[
\minimize_{w \in \mathbb{R}^n} \sum_{i=1}^{m} \phi(y_i - w^\top x_i) + \frac{\lambda}{2} \|w\|_2^2
\]  

(1)

In order to be able to derive a meaningful dual, e.g. in order to be able to obtain certificates of suboptimality, we rewrite Equation (1) as:

\[
\minimize_{w \in \mathbb{R}^n, z \in \mathbb{R}^m} \sum_{i=1}^{m} \phi(z_i) + \frac{\lambda}{2} \|w\|_2^2
\]

s.t. \( z_i = y_i - w^\top x_i \quad i = 1 \ldots m \).  

(2)

(a) Use the general form for duals of a problem with linear equalities and an arbitrary objective to write down the dual of Equation (2) in terms of the conjugate of the loss. Simplify the dual by eliminating unnecessary variables.
(b) Write down the KKT conditions for a pair of primal and dual optimal solutions of Equation (2). Explain how to use the KKT conditions to easily obtain a primal optimal solution if you are given a dual optimal solution.

Answer the remaining questions for each of the following loss functions:

\( \phi(z) = z^2 \)
\( \phi(z) = [|z| - 1]_+ \)

(c) Derive the Fenchel conjugate of \( \phi(z) \) as a function of \( z \).

(d) Write down the Fenchel conjugate of the objective \( f_0(z, w) \) of Equation (2) (Hint: express the objective as an independent sum of functions of each of the optimization variables).

(e) Write down the dual.

(f) Suppose you are given \( w \), and a set of dual variables \( \alpha_i \). How can this be used to obtain a bound on the suboptimality of \( w \)? Write down the bound.

(g) [Optional] Suppose \( \alpha \) is \( \epsilon \)-suboptimal for the dual problem. You pretend its optimal and derive \( w \) based on the formula from the KKT conditions. What can you say about the suboptimality of \( w \)? Can it be worse then \( \epsilon \)-suboptimal? How much worse? Construct the worst example you can and/or obtain the tightest upper bound you can.

## 2 Semi-Definite Program

In this problem, we consider a variant of the binary rating reconstruction problem we looked at in class. Consider \( n \) “users,” \( m \) “movies,” and a sparse set of ratings \( y_{ij} \in \{\pm 1\} \) over a (small) subset of possible user-movie pairs \( (i,j) \in S \). We again want to find small-norm vectors \( u_i \in \mathbb{R}^k \) and \( v_j \in \mathbb{R}^k \), where \( k \geq n + m \). User \( i \) is associated with \( u_i \) and movie \( j \) is associated with \( v_j \). We want the \( u_i \) and \( v_j \) to explain the ratings in the sense that:

\[ y_{ij}\langle u_i, v_j \rangle \geq 1 \]

for each \((i,j) \in S \). While satisfying these conditions, we would like to minimize the maximum norm, i.e.

\[
\begin{align*}
\text{minimize} & \quad \max_i \|u_i\|_2, \max_j \|v_j\|_2 \\
\text{s.t.} & \quad y_{ij}\langle u_i, v_j \rangle \geq 1 \quad \forall (i,j) \in S.
\end{align*}
\]

(3)

(a) Express (3) as a semi-definite program.

(b) Given a primal solution for the SDP, describe how you could determine the optimal \( u_i \)'s and \( v_i \)'s.

(c) Derive the dual of this SDP.

(d) Write down the KKT conditions for the problem, and simplify them as much as you can.
3 Programming

In “maxnorm.py”, you are provided with a function GEN_RATINGS(n, m, p) to generate \( p \) ratings \( y_{ij} \in \{-1, +1\} \) from \( n \) users and \( m \) movies.

(a) Create a set of ratings \( y_{i,j} \) with \( n = 10 \), \( m = 15 \), and \( p = 12 \), and use CVXPY to solve the SDP you formulated in 2(a) for these ratings. What is the optimum value \( p^* \)?

(b) Now, experiment with different values of \( m = n \) and \( p \) and see how large of a problem you can solve within \( \approx 1 \) minute (this will depend on your computer). Report roughly how large of an \( m, p \) you can solve. Which dimension seems to affect the computational complexity more?

Submit all of the code you used to solve this problem. In addition to attaching a .py file, copy-paste into your writeup any code that you wrote, i.e., do not include the code that is provided to you.