Homework 3

Convex Optimization 10-725/36-725

Due Wednesday October 21 at 4:00pm
submitted to Mallory Deptola in GHC 8001
(Remember to submit each problem on a separate sheet of paper, with your name on at the top)

1 Projection onto the ℓ_1-ball [Matt]

We project a point $v \in \mathbb{R}^n$ onto the unit ℓ_1-ball by solving the optimization problem

$$\min_x \ \frac{1}{2}\|x - v\|_2^2$$
$$\text{subject to} \ \ \|x\|_1 \leq 1.$$

(a) Write down the Lagrangian and form the dual problem. Explain how a solution to the dual problem can be used to find a solution to the primal problem.

(b) Prove that the optimal primal-dual pair $(x^\star, \lambda^\star)$ satisfies the following. First, if $\|v\|_1 \leq 1$, show that $(v, 0)$ is primal-dual optimal. Otherwise, show that $\lambda^\star$ must satisfy

$$\sum_{i=1}^n (|v_i| - \lambda^\star)_+ = 1$$

where $a_+ = \max\{0, a\}$.

(c) Using the sufficient conditions for optimality derived in part (b), give an efficient, *noniterative* algorithm for finding $(\lambda^\star, x^\star)$ that computes the exact solution up to machine precision. Provide the computational complexity for your solution in big O notation.

2 Practice with conjugate functions [Hanzhang]

(a) Let $f(x) = \|x\|$ be a norm of x, and it has a dual norm $\|x\|_*$. Prove that if $\|x\|_* > 1$, then the conjugate f^* of f is infinite at x. Prove that if $\|x\|_* \leq 1$, then the conjugate f^* is zero at x. Thus you have established, as asserted in class, that

$$f^*(x) = 1\{\|x\|_* \leq 1\}.$$

(b) Compute the conjugate function of the following functions:

1. $f(x) = \log(1 + e^{b^T x})$, for $x \in \mathbb{R}^n$.
2. $f(x) = \max_{i=1,\ldots n} \ x_i$, for $x \in \mathbb{R}^n$.
3. $f(x) = \|x\|_p$, for $x \in \mathbb{R}^n$ and $p \geq 1$.

3 Practice with duality [Shashank]

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a closed and convex function, whose conjugate function is f^*. Consider the following primal problem:

$$\min_x \ f(x) \qquad \text{subject to} \qquad Ax = b.$$

   (a) Prove that the Lagrange dual of this problem is:

$$\min_v \ f^*(-A^T v) + v^T b.$$

   (b) Prove that the dual of the dual problem in part (a) matches the primal problem. (Hint: set a new variable $u = -A^T v$ and add a new linear constraint to the dual problem.)

2. The Kanotorovich inequality (*BV Additional Exercise 4.14*).

   (a) Suppose $a \in \mathbb{R}^n$ with $a_1 \geq a_2 \geq a_3 > ... \geq a_n > 0$, and $b \in \mathbb{R}^n$ with $b_k = 1/a_k$. Derive the KKT conditions for the convex optimization:

$$\begin{aligned} &\min \quad -\log(a^T x) - \log(b^T x) \\ &\text{subject to} \quad x \in \mathbb{R}^n_+, \qquad \mathbf{1}^T x = 1 \end{aligned}$$

   Show that $x = (1/2, 0, ..., 0, 1/2)$ is the optimal solution.

   (b) Suppose $A \in \mathbb{S}^n{}_{++}$ with eigenvalues λ_k sorted in decreasing order. Apply the result of part (a), with $a_k = \lambda_k$, to prove the Kantorovich inequality:

$$2(u^T A u)^{1/2}(u^T A^{-1} u)^{1/2} \leq \sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}$$

   for all u with $\|u\|_2 = 1$.

4 Solving SVMs using a QP solver [Dallas]

In the lecture, we covered the SVM in both its primal and dual formulations. For $X \in \mathbb{R}^{n \times p}$ and $y \in \{-1, 1\}^n$ the primal problem is:

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \ & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{s.t.} \ & \xi_i \geq 0, \forall i \\ & y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, \forall i \end{aligned}$$

The dual problem is:

$$\begin{aligned} \max_w \ & -\frac{1}{2} w^T \tilde{X} \tilde{X}^T w + 1^T w \\ \text{s.t.} \ & 0 \leq w_i \leq C, \forall i \\ & w^T y = 0 \end{aligned}$$

where $\tilde{X} = \text{diag}(y)X$.

For this problem, we ask you to use a QP solver, such as quadprog in Matlab or R to solve both the primal and the dual formulations. The website has a dataset created from a subset of the 20 newsgroups dataset, in news.mat. (Note to R users: to read Matlab files into R, you can use the package `R.matlab` on CRAN.) Specifically, the data correspond to posts from the talk.politics.misc and talk.religion.misc forums. The goal is to compare the primal and dual SVM solutions for classifying documents as belonging to one or the other. `y_train` is a vector of labels (-1 or 1, corresponding to the two categories). `X_train` is a sparse matrix of training data, where each row represents one document, and each column represents one feature. The features here are binary, with $X_{ij} = 1$ indicating that word j is present in document i and 0 indicating it is not. `y_test` and `X_test` follow the same pattern. The files words.mat contain the words that correspond to the features in X. Headers, footers, and quotes have been removed from the text in advance to make the classification slightly more difficult.

1. Construct two optimization problems to solve the SVM problem in both primal and dual formulations, with $C = 1$ for the training data (`X_train` and `y_train`), and solve them using a QP solver. For "quadprog" in Matlab, specify the "interior-point-convex" algorithm (see http://www.mathworks.com/help/optim/ug/quadprog.html for more details). For "quadprog" in R, you may need to add a small positive amount to the diagonal of the quadratic term in the objective. **Please submit your code as an appendix to this problem**.

2. Report the final objective values for both the primal and dual problems.

3. Extract β from the primal solution, and call this β_{primal}. Computer β from the dual solution, and call this β_{dual}. Plot of the elements of β_{primal} vs the elements of β_{dual}. What do you see?

4. Calculate the misclassification rate on the training and test data.

5. Extract the slack variables ξ from the primal solution, and plot $y_i(x_i^T\beta + \beta_0) - 1 + \xi_i$ vs ξ_i, over $i = 1, \ldots n$. Does this conform to your expectations?

6. How many items among the training data are a) misclassified; b) within the margin? Report the value of w_i for all of these points.

7. For those items in the training data that fall within the margin, use the information in words.mat to reconstruct the original text of the corresponding documents.