Homework Set Four
ECE 271B - Winter 2021
Department of Electrical and Computer Engineering
University of California, San Diego

By submitting your HW, you agree with the following.

1. HW should be treated as a **take-home test** and be an **INDIVIDUAL** effort. **NO collaboration is allowed**. The submitted work must be yours and must be original.
2. We expect the work that you turn-in to be your own, using the resources that are available to **all** students in the class: TAs, myself, Piazza, books, papers, internet.
3. You are not allowed to consult or use resources provided by tutors, previous sludents in the class, or any websites that provide solutions or help in solving assignments and exams.
4. 0 points will be assigned to any problem that seems to violate these rules and, if recurrent, the incident(s) will be reported to the Academic Integrity Office.

Problem 1. In this problem, we consider some minimization problems with inequality constraints.

a) Consider a vector $\mathbf{x} \in \mathbb{R}^n$. Find the optimal solution to the problem

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{x}||^2$$

$$\text{subject to} \sum_i x_i \leq -3.$$

b) Prove the inequality

$$(\mathbf{x}^T\mathbf{y})^2 \leq (\mathbf{x}^T\mathbf{Q}\mathbf{x})(\mathbf{y}^T\mathbf{Q}^{-1}\mathbf{y}),$$

where $\mathbf{Q}$ is a positive definite symmetric matrix. (Hint: you may want to consider the problem

$$\max_{\mathbf{x}} \mathbf{y}^T\mathbf{x}$$

$$\text{subject to } \mathbf{x}^T\mathbf{Q}\mathbf{x} \leq 1.)$$

Problem 2. In this problem, we study the topic of duality by considering some simple examples. In all cases, the minimization problems are of the form

$$\min_{x \in X} f(x)$$

$$\text{subject to } g(x) \leq 0.$$

We consider six problems, described by the table below.

Problem	$f(x)$	$g(x)$	X
1	$x_1 - x_2$	$x_1 + x_2 - 1$	$\{(x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0\}$
2	x	x^2	$\mathbb{R}$
3	$-x$	$x - \frac{1}{2}$	$\{0,1\}$
4	$-x$	$x - \frac{1}{2}$	$[0,1]$
5	$\frac{1}{2}(x_1^2 + x_2^2)$	$x_1 - 1$	$\mathbb{R}^2$
6	$\lvert x_1 \rvert + x_2$	x_1	$\{(x_1, x_2) \mid x_2 \geq 0\}$

For each of the six problem answer the following.

1. Sketch the set of feasible solutions $S = \{g(x), f(x) \mid x \in X\}$ in (g, f)-space.
2. Determine if there exists a Lagrange multiplier. If so, what is the set of Lagrange multipliers? If not, why is there no multiplier?
3. State the dual problem and sketch a plot of $q(\mu)$ as a function of μ. Determine if there is a duality gap.

Problem 3. In class, we have studied Vapnik's SVM formulation, i.e. the search for the hyperplane that solves the following optimization problem

$$\min_{\mathbf{w},\boldsymbol{\xi}} \left(\frac{1}{2}||\mathbf{w}||^2 + \frac{C}{n}\sum_{i=1}^{n} \xi_i \right)$$

subject to

$$\begin{array}{rcl} y_i(< \mathbf{x}_i, \mathbf{w} > +b) & \geq & 1 - \xi_i, \; i = 1, \ldots, n \\ \xi_i & \geq & 0, \; i = 1, \ldots, n. \end{array}$$

One limitation of this formulation is that there is no intuition for what the parameter C means and it can therefore be difficult to find good values for it in practice. In this problem, we consider a slightly different, but more intuitive formulation, based on the solution of the following problem

$$\min_{\mathbf{w},\boldsymbol{\xi},\rho,b} \left(\frac{1}{2}||\mathbf{w}||^2 - \nu\rho + \frac{1}{n}\sum_{i=1}^{n} \xi_i \right)$$

subject to

$$\begin{array}{rcl} y_i(< \mathbf{x}_i, \mathbf{w} > +b) & \geq & \rho - \xi_i, \; i = 1, \ldots, n \\ \xi_i & \geq & 0, \; i = 1, \ldots, n \\ \rho & \geq & 0. \end{array}$$

a) For this case, determine the dual problem and the form of the resulting decision function.

b) Given the dual solution, how would you determine the values of b and ρ?

c) Define the fraction of margin errors as

$$\epsilon_\rho = \frac{1}{n} \left|\{ i \,|\, y_i g(\mathbf{x}_i) < \rho \}\right|$$

and suppose that we solve the optimization problem on a dataset with the result that $\rho > 0$. Show that

1. ν is an upper bound on ϵ_ρ;
2. ν is a lower bound on the fraction of vectors that are support vectors.

d) Show that, if the solution of the second problem leads to $\rho > 0$, then the first problem with C set a priori to $\frac{1}{\rho}$ leads to the same decision function.

Problem 4. In this problem, we will use the SVM to, once again, classify digits. In all questions, you will use the training set contained in the directory `MNISTtrain` (60,000 examples) and the test set in the directory `MNISTtest` (10,000). To reduce training time, **we will only use 20,000 training examples**. To read the data, you should use the script `readMNIST.m` (use readDigits=20,000 or readDigits=10,000 respectively, and offset=0). This returns two matrices. The first (`imgs`) is a matrix with $n \in \{10000, 20000\}$ rows, where each row is a 28×28 image of a digit, vectorized into a 784-dimensional row vector. The second (`labels`) is a matrix with $n \in \{10000, 20000\}$ rows, where each row contains the class label for the corresponding image in `imgs`. Since there are 10 digit classes, we will learn 10 binary classifiers. Each classifier classifies one class against all others. For example, classifier 1 assigns label $Y = 1$ to the images of class 1 and label $Y = -1$ to images of all other classes. Download and instal the `libsvm` package to learn the SVM classifiers.

a) In this problem, we will learn linear SVMs. Using `libsvm`, learn three SVMs with values of the regularization constant $C \in \{2, 4, 8\}$. For each classifier and digit, 1) report the test error, 2) report the number of support vectors, and 3) plot the three support vectors of largest Lagrange multiplier on each side of the boundary. For each classifier, report the overall classification error. Comment on the results.

b) For each binary classifier, make a plot of cumulative distribution function (cdf) of the margins $y_i(\mathbf{w}^T\mathbf{x}_i + b)$ of all training examples. Comment on the results.

c) Repeat **a)** and **b)** for an SVM with radial basis function kernel. Use the script `grid.py`, included in the `libsvm` package, to cross-validate the values of C and γ, and then use the two parameters for the classification. Compare the results with those of **a)** and **b)**.