STOR 565 Homework

1. Let $D_n = (x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \{\pm 1\}$ be sequence of labeled pairs. Show that the constraint set

$$C := \{w, b : y_i(x_i^t w - b) \geq 1 \text{ for } i = 1, \ldots, n\}$$

appearing in the primal SVM optimization problem is convex. To make things a bit more formal, treat the elements of C as vectors $v = (w_1, \ldots, w_p, b)^t \in \mathbb{R}^{p+1}$. Hint: Show that C is the intersection of n sets, one for each i, and then show that each of these sets is convex.

In the next two questions you are asked to fill in some of the details from the SVM lecture concerning how one finds the maximum margin classifier for linearly separable data.

2. Write down the primal problem, with optimal value $p*$, and argue using the previous question and results from a previous homework that the primal problem is a convex program. Now consider the Lagrangian $L : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^n_+$, which is defined by

$$L(w, b, \lambda) \ := \ \frac{1}{2}||w||^2 \ - \ \sum_{i=1}^n \lambda_i \{y_i(w^t x_i - b) - 1\}$$

Establish that

$$\max_{\lambda \geq 0} L(w, b, \lambda) \ = \ \begin{cases} ||w||^2 & \text{if } y_i(x_i^t w - b) \geq 1 \text{ for } i = 1, \ldots, n \\ +\infty & \text{otherwise} \end{cases}$$

To see why this is true, note that if one of the constraints $y_i(x_i^t w - b) \geq 1$ is *not* satisfied, then one can increase the corresponding λ_i to make the Lagrangian arbitrarily large. Using the last display above, argue informally that the primal problem can be written as

$$p^* \ = \ \min_{w,b} \max_{\lambda \geq 0} \ L(w, b, \lambda)$$

3. Show that the Lagrange dual function, defined by

$$\tilde{L}(\lambda) \ = \ \min_{w,b} \ L(w, b, \lambda)$$

is concave. Hint: Argue that the dual function is the minimum of linear (hence concave) functions, and is therefore concave. The SVM dual problem is given by the program

$$\max \tilde{L}(\lambda) \quad \text{s.t.} \quad \sum_{i=1}^n \lambda_i \, y_i = 0 \ \text{ and } \ \lambda_1, \ldots, \lambda_n \geq 0$$

Carefully define the constraint set for λ in this problem and argue that this set is convex. (Note that there are $n+1$ constraints.) Thus the dual problem seeks to maximize a concave function over a convex set.

4. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be any real valued function. Show that

$$\max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} f(x, y) \;\leq\; \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y)$$

This inequality shows that the value d^* of the SVM dual problem is less than or equal to the value p^* of the SVM primal problem.

5. Let $X_1, \ldots X_n$ be independent and identically distributed random variables. Calculate $\mathbb{E}[X_1 | X_1 + \ldots + X_n = x]$. (Hint: Consider $\mathbb{E}[S_n | X_1 + \ldots + X_n = x]$ where $S_n = X_1 + \ldots + X_n$ and use symmetry.)

6. Let $\phi(x)$ and $\Phi(x)$ be the density function and cumulative distribution function, respectively, of the standard normal distribution. Here we will find a useful upper bound on $1 - \Phi(x)$, which is the probability that a standard normal random variable exceeds x.

(a) Write down the formula for the density $\phi(t)$, and compute the derivative $\phi'(t)$.

(b) Justify the following sequence of equalities: For $x > 0$,

$$1 - \Phi(x) \;=\; \Phi(-x) \;=\; \int_{-\infty}^{-x} \phi(t)\, dt \;=\; \int_{-\infty}^{-x} \frac{1}{t} \cdot t\, \phi(t)\, dt.$$

(c) Integrate the last term above by parts to establish the useful inequality $1 - \Phi(x) \leq x^{-1}\, \phi(x)$ for $x > 0$.