....

STAT 538 Final Exam

Wednesday March 16, 2011, 4:30-6:20

Student name: ....

- notes and books are allowed
- electronic devices are not allowed

- *Do Well!*

Problem – Convex sets

2 points

1.1 If A, B are sets, denote by $A + B = \{a + b \,|\, a \in A,\, b \in B\}$. Show that $A + B$ is convex whenever A and B are convex.

If S is a set and $a > 0$, denote by $S_a = \{x \,|\, inf_{s-\in S}||x - s|| \leq a\}$. Show that S_a is convex if S is convex.

Can you find a counterexample, i.e a set S which is not convex, and an s, so that S_a is convex?

If S is a set and $a > 0$, denote by $S_{-a} = \{x \,|\, x' \in S \,\text{if}\, ||x - x'|| \leq a\}$. In other words, $x \in S_{-a}$ if x is contained in S together with the radius a ball around it.

Show that S_a is convex if S is convex.

Let p be a probability distribution, and $a > 0$. Is the set $H(p) \geq a$ convex?

Let p, q be a probability distributions over the same domain Ω, and $a > 0$.

Let $\phi(x)$ be a strictly convex function with domain in $\setminus^n$ and let $d_\phi(y, x)$ be the Bregman divergence defined by ϕ, i.e

$$d_\phi(y, x) \;=\; \phi(y) - \phi(x) - \nabla f(x)^T(y - x)$$

Show that the Bregman ball $B(y, a) = \{d_\phi(y, x) \geq a\}$ is a convex set.

Problem – Linearly Separable Support Vector Machine

Prove that for the linear SVM in the linearly separable case, the solution satisfies

$$||w||^2 \;=\; \sum_i \alpha_i$$

(Hint: a simple solution exists.)

Problem – Linear SVM

Consider the following optimization problem, that corresponds to the linear SVM for data set $\{(x^i, y_i), i = 1 : m\}$, $x^{1:m} \in \setminus^n$. It is assumed that the data are linearly separable, so that the problem is feasible.

$$(\mathcal{P}) \quad \min_{w,b} \;\; \frac{1}{2}||w||_1 \tag{1}$$

$$\text{s.t.} \quad y_i(w^T x^i + b) \;\geq\; 1 \tag{2}$$

This problem can be transformed into the differentiable problem

$$
\begin{array}{lllll}
(\mathcal{P}') & \min_{t,w,b} & \mathbf{1}^T t & & \\
& \text{s.t.} & y_i(w^T x^i + b) \geq 1 & \lambda_i,\ i = 1:m & \\
& & w \prec t & \alpha^+_{1:n} & \\
& & -w \prec t & \alpha^-_{1:n} & \\
& & t \succ 0 & \mu_{1:n} &
\end{array} \tag{3}
$$

In the above, on the right, were introduced the Lagrange multipliers corresponding to each of the constraints. Note that $(\mathcal{P}')$ is a linear program. Denote the solution of these dual problems by $w^*, b^*, t^*, \lambda^*, (\alpha^{\pm})^*, \mu^*$.

.1 Show that $t^*_j = |w^*_j|$ for $j = 1 : n$, i.e. the variable t represents the magnitude of w.

.1 Write the expression of the Lagrangean of $(\mathcal{P}')$.

.2 Write the expression of the partial derivatives of the Lagrangean w.r.t the primal variables. Find the dual function $g(\lambda, \alpha^+, \alpha^-, \mu)$.

.3 Write now the dual problem $(\mathcal{D}')$ corresponding to $(\mathcal{P}')$. Show that this problem can be expressed only in terms of $\lambda, \alpha^{\pm}$ (i.e that μ can be eliminated). Is $(\mathcal{D}')$ a linear program?

.5 Assume that for some $j \in 1 : n$ the solution $\mu^*_j > 0$. What can you say about w^*_j in this case?

.4 Assume that for some $j \in 1 : n$ the solution $w^*_j \neq 0$. Show that for this j, one has either $\alpha^+_j = 0, \alpha^-_j = 1$ or $\alpha^+_j = 1, \alpha^-_j = 0$. What does $\alpha^+_j - \alpha^-_j$ represent in this case?

FYI: The Linear SVM, by penalizing the 1-norm of w, encourages the appearance of zeros among the elements of w. This is a different kind of sparsity than that encountered for the "quadratic" (standard) SVM, where the sparsity consisted on having λ_i's equal to 0. Note that it is in general difficult to ensure both.

Another remark is that, while LP's are somewhat easier to solve, the Linear formulation is not kernelizable, and we don't have the "representer theorem" $w = \sum \lambda_i y_i x^i$ that we had for quadratic SVM's.

Problem – General barrier function

Let $h : (-\infty, 0) \to \searrow$ be a twice differentiable, closed, increasing convex function, with $\lim_{u \to 0} h(u) = \infty$. Now consider the *convex* optimization problem

$$\min_x \quad f_0(x) \tag{4}$$
$$\text{s.t.} \quad f_i(x) \leq 0, \;\; i = 1:m \tag{5}$$

where $f_0, f_{1:m}$ are twice differentiable. We define the h-barrier for this problem as $\phi_h(x) = \sum_{i=1}^m h(f_i(x))$ with domain $\{x \,|\, f_i(x) < 0\}$.

Explain why $tf_0(x) + \phi_h(x)$ is convex in x for every $t > 0$.

Let $x^*(t) = \min_x tf_0(x) + \phi_h(x)$ (the h-central path). We assume that the minimizer exists and is unique for every $t > 0$. Show how to construct a dual feasible λ from $x^*(t)$ and find the associated duality gap.

For what functions h does the duality gap found in depend only of t and m (and no other problem data)?

Problem – Boosting as Minimum Relative Entropy

In this problem we will recover the DiscreteAdaBoost algorithm parameters as the solution of a Minimum Relative Entropy problem.

Denote the data set by $\{(x^i, y_i), i = 1:n\}$, $x^{1:m} \in \searrow^n, y_i = \pm 1$, and the weights at step k by $w_i^k,, i = 1:n$. The classifier at step k is $f_k(x)$ taking values ± 1. For simplicity, we denote

$$z_i = f_k(x_i)y_i = \pm 1$$

We need to determine how to update the weights. Let $x \equiv w^{k+1}$ be the unknown weights. We will show that the new weights can be obtain as the solution to the MRE problem below.

$$\min_x \quad KL(x||w^k) \tag{6}$$
$$\text{s.t.} \quad \sum_i z_i x_i = 0 \tag{7}$$

The above problem says the new weights should be “orthogonal” to the previous classifier ($z^T x = 0$) w.r.t classification, but should be as close as possible to the previous weight distribution w^k. We do not enforce $x_i > 0$ because the domain of the KL divergence enforces it, and do not enforce normalization either (hence the solution may be an unnormalized distribution).

.1 Is this a convex optimization problem? **1.** Let c be the Lagrange multiplier of the linear constraint. Write the expression of the Lagrangean of this optimization problem.

2. Take the partial derivative of L w.r.t each primal variable x_i and by equating them to 0 find the general form of the solution x_i as a function of c, z_i.

.3 Replace the solution found in **.** in (7) and solve for c. You should recover the c_k of the DiscreteAdaBoost algorithm.

The end!