....

STAT 538 Final Exam

Friday March 5, 2010, 3:30-5:20

Student name: ....

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

- *Do Well!*

Problem 1 – Maxima of convex functions

Let f be a strictly convex function, $C \subseteq \operatorname{relint} \operatorname{dom} f$ be a closed convex set, and let $x^* \in C$ be a point such that $f(x^*) \geq f(x)$ for all $x \in C$.

2 points

1.1 Show that x^* is an extreme point of C.

1 point

1.2 Is x^* unique? Prove or give a counterexample.

1 point

1.3 Is x^* isolated? Prove or give a counterexample. [x^* Isolated $\Leftrightarrow$ there is a neighborhood of x^* that contains no other maximum of f.]

Problem 2 – The rate of convergence of gradient descent with line minimization

Let $x \in \mathbb{R}^2$ and $f = \frac{1}{2}x^T Q x$ with Q symmetric, $Q \succ 0$. f is optimized by gradient descent $x^{k+1} = x^k - \alpha g^k$, with $g^k = \nabla f(x^k)$

2 points

2.1 Denote $x^k = x$, $g^k = g$. What is the optimal step size α ?

2 points

2.2 What is $f(x - \alpha g)/f(x)$ as a function of g, Q.

2 points

2.3 Let

$$Q = \begin{bmatrix} 2 & a \\ a & 2 \end{bmatrix}$$

where $a = 2 - \epsilon$ for $\epsilon << 1$ a small number. Use Kantorovich's inequality

$$\frac{(y^T y)^2}{(y^T Q y)(y^T Q^{-1} y)} \geq \frac{4mM}{(M+m)^2} \quad \text{for } y \in \mathbb{R}^n,\ Q \succ 0,\ M = \max \lambda(Q), m = \min \lambda(Q)$$

to determine the rate of convergence of gradient descent with exact line minimization for this problem, as a function of ϵ. ($\max, \min \lambda(Q)$ denote respectively the largest and smallest eigenvalues of the symmetric matrix Q).

Problem 3 – SVM with logarithmic penalty

Consider the following variant of the SVM classifier:

$$
\begin{aligned}
(\mathcal{P}) \quad \min_{w,b,\gamma_{1:m}} \quad & \tfrac{1}{2}||w||^2 + \textstyle\sum_{i=1}^{m} \ln \frac{1+e_i^{\gamma}}{2} & \\
\text{s.t.} \quad & y_i(w^T x_i + b) \geq 1 - \gamma_i, \ \text{for all}\, i & (1) \\
& \gamma_i \geq 0 \ \text{for all}\, i & (2)
\end{aligned}
$$

where $\{(x_i, y_i),\ i = 1 : m\}$ is a data set, $x_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$.

1 point

3.1 Explain in 1 sentence what is the role of the γ_i variables. What is the optimal value of the γ_i variables if the data can be correctly classified by a hyperplane.

2 points

3.2 Denote the dual variables associated with (1) by λ_i, and the dual variables associated with (2) by α_i, $i = 1 : m$. Write the expression of the Laplacian for the optimization problem $(\mathcal{P})$.

Write the partial derivatives of L w.r.t the primal variables.

2 points

3.3 Derive the dual objective $g(\lambda, \alpha)$.

1 point

3.4 Write $(\mathcal{D})$ the dual optimization problem of $(\mathcal{P})$. Verify that it is a concave maximization problem.

1 point

3.5 Is $(\mathcal{D})$ a quadratic problem?

1 point

3.6 Write the expressions of w^*, b^* and of the SV classifier as functions of (λ^*, α^*) the solution of $(\mathcal{D})$.

2 points

3.7 Let (x_i, y_i) be a data point for which (w^*, b^*) has margin < 1. Which of $\gamma_i, \lambda_i, \alpha_i$ will be $> 0, = 0, < 0$?

2 points

3.8 Let (x_i, y_i) be a data point for which (w^*, b^*) has margin > 1. What can you say about λ_i, γ_i, α_i in this case?

2 points

3.9 Let (x_i, y_i) be a data point for which (w^*, b^*) has margin $= 1$. What can you say about λ_i, γ_i, α_i in this case? Find λ_i as a function of the other λ's.

The end!