STAT 538 Homework 7
Out ThursdayMarch 1, 2012
Due Thursday, March 8 , 2012

mmp@stat.washington.edu

Problem 1 – The Markov chain as maximum entropy model

Problem 3 from Homework 6.

Problem 2 - SVM with logarithmic penalty

Consider the following variant of the SVM classifier:

$$(\mathcal{P}) \quad \min_{w,b,\gamma_{1:m}} \quad \frac{1}{2}||w||^2 + \sum_{i=1}^{m} \ln \frac{1+e^{\gamma_i}}{2}$$

$$\text{s.t.} \quad y_i(w^T x_i + b) \geq 1 - \gamma_i, \text{ for all } i \qquad (1)$$

$$\gamma_i \geq 0 \text{ for all } i \qquad (2)$$

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

1.a Explain in 1 sentence what is the role of the γ_i variables.

1.b 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 Lagrangean for the optimization problem $(\mathcal{P})$.

1.c Write the partial derivatives of L w.r.t the primal variables, then derive the dual objective $g(\lambda, \alpha)$.

Hints: It will help to denote $\alpha_i + \lambda_i = \beta_i$ and to express the dual g as a function of λ, β instead of λ, α. Note also the following identity: if $b = 1/(1 + e^{-c})$, then $1 + e^c = 1/(1 + b)$. Recall also that $\ln(1 + e^\gamma)$ is the conjugate function of the entropy, so look for possible entropy terms in the function g. As in the previous homework, to get full credit you must simplify as much as possible the expressions you obtain.

1.d Write $(\mathcal{D})$ the dual optimization problem of $(\mathcal{P})$. It is recommended that you formulate $(\mathcal{D})$ as an optimization over (λ, β). Note that for fixed λ_i the relation between α_i and β_i is very simple.

1.e Is $(\mathcal{D})$ a quadratic problem?

The next questions are OPTIONAL, not graded

1.f Write the expression of w^* and explain how to obtain b^* as functions of (λ^*, α^*) the solution of $(\mathcal{D})$.

1.g 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 $> 0, = 0, < 0$?

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

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