CSC 2545, Spring 2017  
Kernel Methods and Support Vector Machines  

Assignment 2

This assignment is due at the start of class, at 2:10pm, Thurs March 23.  
No late assignments will be accepted.

The material you hand in should be legible, well-organized and easy to mark, including the use of good English. Short, simple answers and proofs will receive more marks than long, complicated ones. Up to 20% of your mark will be for presentation. Unless stated otherwise, you must justify your answers.

All computer problems are to be done in Python with Scikit-learn and should be properly commented and easy to understand. These programs should all be stored in a single file called `source.py`, which should be submitted electronically as described on the course web site. We should be able to import this file as a Python module and execute your programs.

More questions will be added shortly

1. Prove the following statements in the context of $\nu$-SVC:

   (a) if $y_i(\omega \cdot x_i + b) > \rho$ then $\alpha_i = 0$

   (b) if $y_i(\omega \cdot x_i + b) < \rho$ then $\alpha_i = 1/m$

2. Question 9.2 on page 274 of the text.

3. Question 9.16 on page 276 of the text.

4. Show that the following optimization problem is equivalent to a soft-margin SVM:

   $$ \min_{\omega, b} \frac{\|\omega\|^2}{2} + C \sum_i h[y_i(\omega \cdot x_i + b)] $$

   where $h(z) = \max(0, 1 - z)$ is the hinge loss function.

5. In this problem, we will formulate a soft-margin SVM without using a mapping to a higher-dimensional feature space. Instead, we will search directly for a decision function of the form $f(x) = \sum_i \beta_i k(x, x_i) + b$. In particular, consider the following optimization problem:

   $$ \min_{\beta, b} \sum_{i,j} \beta_i \beta_j k(x_i, x_j)/2 + C \sum_i h[y_i f(x_i)] $$

   where $f(x) = \sum_i \beta_i k(x, x_i) + b$ and $h$ is the hinge loss function (and $C > 0$).
(a) Assuming that $k(x,y)$ is a similarity measure, give an intuitive description of the term $\sum_{i,j} \beta_i \beta_j k(x_i, x_j)$ without referring to a feature space or maximizing a margin. How is this reflected in the properties of an SVM?

(b) Introduce slack variables into the problem, derive the dual, and show that it is the same as the dual of a soft-margin SVM.

(c) Under what conditions are the $\beta_i$ uniquely determined by the solution to the dual problem? In this case, give a simple formula for $\beta_i$.

(d) Do we need to assume that $k(x,y)$ is positive definite as in the standard formulation of an SVM? If so, how is it assumed?

(e) Do we need to assume that $k(x,y)$ is symmetric as in the standard formulation of an SVM? If so, how is it assumed?

(f) Show that the solution to the dual problem (as given on the bottom of page 205 in the text) does not depend on $k$ being symmetric. That is, if $k$ is not symmetric, then there is a symmetric $k'$ that gives exactly the same solution for the dual variables.