Homework Set Seven
ECE 175
Department of Computer and Electrical Engineering
University of California, San Diego
Nuno Vasconcelos

This HW set contains several problems. Only the problem labeled **Quiz** must be handed in and will be graded. The remaining problems are for practice. You should not submit them for grade. By submitting your Quiz solution, you agree to comply with the following.

1. The Quiz is treated as a **take-home test** and is an **INDIVIDUAL** effort. **NO collaboration is allowed**. The submitted work must be yours and must be original.
2. The work that you turn-in is your own, using the resources that are available to **all** students in the class.
3. You can use the help of **GENERAL** resources on programming, such as MATLAB tutorials, or related activities.
4. You are not allowed to consult or use resources provided by tutors, previous students in the class, or any websites that provide solutions or help in solving assignments and exams.
5. You will not upload your solutions or any other course materials to any web-sites or in some other way distribute them outside the class.
6. 0 points will be assigned if your work seems to violate these rules and, if recurrent, the incident(s) will be reported to the Academic Integrity Office.

Quiz (Computer) In this final assignment we will perform SVM classification on the digits dataset. We will use the LIBSVM – A Library for Support Vector Machines to solve the optimization problem. The instructions to install and use the library are provided on the homework site. It will also be useful to read the `README` file that comes along with the library.

1. Classification using SVM: We shall use the gaussian kernel $K(x,y) = \exp(-\gamma||x-y||^2), \gamma > 0$ which maps the image vectors to a higher dimensional space. Apart from the selection of the kernel, we also need find what are the model parameters (C, γ). The γ parameter controls the scale of the kernel, and C controls the harshness of the penalty on classification errors. We will learn these model parameters using 2-fold cross-validation.
   (a) Perform a grid search on the model parameters to find the best classification accuracy using cross validation. Choose the C values from $[2^{-3},\ 2^{-1},\ 2^{1},\ 2^{3},\ 2^{5},\ 2^{7},\ 2^{9},\ 2^{11}]$ and the γ parameter from $[2^{-11},\ 2^{-9},\ 2^{-7},\ 2^{-5},\ 2^{-3},\ 2^{-1}]$. What is the best (C, γ) pair and the corresponding classification accuracy?
   (b) Now using the (C, γ) obtained above train the SVM classifier on the whole training set. Using this model perform a SVM classification on the test set. What is the final classification error.
2. Two class SVM using linear kernels $K(x,y) = x^T y$.

(a) Train a SVM model with images from digit6 and digit8 as two classes. Linear Kernel have one model parameter - C. Fix C equal to 2^{-4}. Perform the classification on the test data (also containing the images only from digit6 and digit8). What is the classification accuracy?

(b) The SVM model is a structure which has the information regarding the support vectors (refer to README: Returned Model Structure). `model.SVs` contain the support vectors and `model.sv_coef`. Plot the top 5 support vectors (as 28×28 images) each for digit6 and digit8. What do you observe from the support vectors?

(c) What is the normal vector to the separating hyperplane. Plot as 28×28 image.

3. Comparison between the various supervised classification techniques: Nearest Neighbor, Gaussian Classification, Gaussian classification on PCA space, SVM classification.

(a) Compare the various techniques mentioned above with respect to their overall classification accuracy on the complete dataset.

(b) What are the advantages and disadvantages of various methods.