

Homework 7
ECE 175
Electrical and Computer Engineering
University of California San Diego

Nuno Vasconcelos

Winter 2014

Due: 3/19 (hand in with your exam)

In this final assignment we will perform SVM classification on the digits dataset. We will use LIBSVM – a Library for Support Vector Machines to solve the optimization problem. The instructions to install and use the library are provided on the course web-page. It will also be useful to read the README file that comes along with the library and the FAQ section on LIBSVM webpage.

1. Two class SVM using linear kernels $K(x, y) = x^T y$.
 - (a) Train a SVM model with the training images from classes digit6 and digit8. With a linear kernel the SVM has a single 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 containing the information regarding the support vectors (refer to README: Returned Model Structure). For each class (digit6 and digit8), plot the five support vectors closest to the classification boundary (as 28×28 images).
 - (c) What is the normal vector to the separating hyperplane? Plot as 28×28 image.
 - (d) For each test example, compute the absolute value of its distance to the classification boundary. Plot an histogram for all distances using 10 bins.
 - (e) Plot, as images, the five test examples from each class that are farthest from the boundary.
 - (f) Plot, as images, the five test examples from each class that are closest to the boundary. What can you say about the differences between these and the digits of (e)?
2. Ten digit class SVM using Gaussian kernel $K(x, y) = \exp(-\gamma \|x - y\|^2)$, $\gamma > 0$.
 - (a) In this case, there are two model parameters (C, γ). γ controls the scale of the kernel, C controls the penalty on classification errors. Use LIBSVM to determine these model parameters using a grid search and 2-fold cross-validation. For your grid search, use the C values $[2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}]$ and the γ values $[2^{-11}, 2^{-9}, 2^{-7}, 2^{-5}, 2^{-3}, 2^{-1}]$. What is the best (C, γ) pair?
 - (b) Using the (C, γ) obtained above perform a SVM classification on the test set. What is the classification error?
3. Comparison between the various supervised classification techniques on the 10 digit dataset: Nearest Neighbor, Gaussian Classification, Gaussian classification on PCA space, SVM classification.
 - (a) Compare all techniques with respect to their classification accuracy on the test set.
 - (b) What are the advantages and disadvantages of various methods?