CSE 5523. Homework 2. Due April 8.

Problem 1. Run a linear SVM on the two class dataset given online (you can use a standard toolbox). Compare its performance to that of the least squares linear classifier.

Instructions: download *79.mat*, which contains images of digits. Each image is given as a 28x28 matrix of grayscale pixel values. It is stored as a 784 ($= 28 \times 28$) array. You are given 1000 images of 7 and 1000 images of 9. These are stored as a single 2000×784 matrix in the file 79.mat. The first 1000 digits are sevens, the rest are nines. Download that file and type "load 79.mat" in Matlab. The matrix d79 contains the data. You can visualize the digits by typing, e.g., the following:

colormap(gray);

x = reshape (d79(1234,:),28,28);

y = x(:,28:-1:1);

pcolor(y)

This bit of code shows you the digit number 1234 (which is a 9).

Problem 2.

Implement (do not use standard toolboxes) the Least Squares Kernel classifier with a Gaussian Kernel. Learn a kernel classifier (choosing a proper value for the kernel width and λ using cross-validation). After it is done (no cheating!), download *test79.mat* and test your code on the test set. Compare your testing and training results. Compare results to linear SVM. What happens, when the kernel bandwidth σ is small? Large?

Problem 3. Use Fourier features to approximate the kernel. Try different numbers of features. Compare your results to Problem 2.

Problem 4. Why is the Hilbert space of square integrable functions on $[0, 1]$, i.e. functions f, such that $\int_0^1 |f(x)|^2 dx$, not an RKHS?

Problem 5. Read about smoothing splines. What is the connections to the material discussed in class?