Machine Learning	Dr. Mario Fritz and Dr. Björn Andres
Exercise Sheet 7 - 15.12.2014	due: 05.01.2015

Exercise 18 - Multiclass Schemes for Classification of Handwritten Digits

In this exercise we are doing handwritten digit classification using multi-class SVM with a Gaussian kernel. In order to solve the optimization problem for the SVM, we are using the MATLAB interface to the LIBSVM package (http://www.csie.ntu.edu.tw/ cjlin/libsvm/). Download `DataEx7.zip` from the course webpage. Note: You do not need to download anything from the LIBSVM webpage - everything you need is contained in the zip-file.

- Extract the files in libsvm-3.20.zip somewhere in your home directory.
- Start Matlab and use addpath to add the directory where you have extracted LIBSVM to the MATLAB search path (use savepath in order to add it permanently).
- Go to the subfolder matlab and type make in the Matlab prompt. (Pre-built binary files for Windows 64bit already contained in the folder).
- The matlab function `getKernelSVMSolution` provides a nice interface to the LIBSVM package (use `help getKernelSVMSolution` to see how it works).
- **(4 Points)** The problem deals with the classification of handwritten digits (10 classes). You are supposed to use the SVM with the Gaussian kernel:

  $$k(x, y) = e^{-\lambda \, \|x-y\|^2}.$$

  The training and test data is in `USPSTrain.mat` and `USPSTest.mat`. The 16×16-images of the digits are represented as 256-dimensional column vectors. Write two matlab scripts:

  - one which solves the multi-class problem using **one-versus-all** (save it in `OneVersusAll.m`),
  - one which solves the multi-class problem using **one-versus-one** (save it in `OneVersusOne.m`),

  In both cases use $C = 100$ and $\lambda = \frac{3}{\gamma}$, where γ is the median of all squared distances between **training** points, as parameters for the binary SVM.

  Visually inspect the digits which have been misclassified. How do you judge the result? Compare the quality of the classification obtained by the two multi-class schemes. How do the two multiclass schemes compare in terms of runtime?

  Save your prediction on the test set in a file `USPSResults.mat` as `PredOneVersusOne` and `PredOneVersusAll` and report the test error for both cases (written on paper). Also generate for both cases a figure (`ErrorsOneVersusOne.png` and `ErrorsOneVersusAll.png`) containing the misclassified images in the test set.
- **(2 Points)** Suppose the computation of a binary classifier has complexity $O(n^m)$, where n is the number of training points. Suppose we have k classes and the training set contains $\frac{n}{k}$ points of each class. What is the computational complexity of the one-versus-all and one-versus-one scheme ? Which multi-class scheme is better in terms of complexity ?

Hints:

- Use `dist_euclidean.m` to compute the squared Euclidean distances between two sets of points (warning: your own code will probably be too slow), `getKernelSVMSolution.m` to obtain the dual variables α and the offset b of the SVM and `VecToImage.m` to plot the images of the digits which have been wrongly classified.
- Given a matrix D, with $D_{ij} = \|X_i - Z_j\|^2$, $i = 1, \ldots, n, j = 1, \ldots, m$, of the squared distances between two point sets $\{X_1, \ldots, X_n\}$ and $\{Z_1, \ldots, Z_m\}$ with $X_i \in \mathbb{R}^d$ and $Z_j \in \mathbb{R}^d$, you can compute the kernel matrix `K` in Matlab as: `K=exp(-lambda*D);`.

Submission:

- Create **one** zip/rar/tar.gz/tgz-file containing the m-files, your plots as .png files and the matlab data files and send the file to your respective tutor. The filename has to follow the following convention:
  `[group:A,B,C]_[matrikel numbers separated by underscore]_ex[nr].[extension]`
  e.g. if you are in group B and your team members have matrikelnumbers 3503239, 3028258 and the current exercise number is 10 then the filename reads: `B_3503239_3028258_ex10.zip`.

Exercise 19 - Construction of Kernels

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite kernel. Prove that the following functions $k'(x, y)$ are again positive definite kernels:

a. **(2 points)** $k'(x, y) = \sum_{s=0}^{\infty} \alpha_s k(x, y)^s$ for $\alpha_s \geq 0, s \geq 0$.

b. **(2 points)** $k'(x, y) = e^{-\beta(d_k^2(x,y))}$, where $\beta > 0$ and $d_k(x, y)$ is the (semi-)metric induced by the kernel k, $d_k^2(x, y) = k(x, x) + k(y, y) - 2k(x, y)$.

c. **(1 point)** $k'(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$.

d. **(2 Bonus points)** $\mathcal{X} = \{x \in \mathbb{R}^d \mid \|x\|_2 < 1\}$ and $k'(x, y) = \frac{1}{1-\langle x,y \rangle}$