Homework 4
(due Tuesday, March 8th 2016)

1. **SVM implementation:** Implement and execute a majorization-based two class linear SVM on Fisher's Iris flower dataset (`http://en.wikipedia.org/wiki/Iris_flower_data_set`). The dataset contains 50 samples from each of the three classes (*setosa, virginica, versicolor*). For the choices of 10%, 30% and 50% as training data, run 10 random trials in each case by selecting each pair of two classes. Since there are three classes, ${}^3C_2 = 3$ possibilities. If you choose 10% for training, in each of the 10 trials, randomly pick 10% for training and the remainder for testing. To document performance, evaluate and report the training and test set misclassification errors for each pairwise case (10%, 30% or 50% training data chosen and for each random trial). Try and balance the percentages of data from each class in the training set. Document all choices including percentages of each class in each random trial.

The SVM objective function can be written as

$$\frac{1}{2}\|\Theta\|^2 + C\sum_{n=1}^{N} \max\left[0, 1 - y_n\left(\mathbf{\Theta}^T\mathbf{x}_n + \Theta_0\right)\right]. \tag{1}$$

This can be majorized as follows. The majorized objective function is written as

$$E(\mathbf{\Theta}, \Theta_0, \{z_n\}) = \frac{1}{2}\|\mathbf{\Theta}\|^2 + C\sum_{n=1}^{N} g(\mathbf{\Theta}, \Theta_0, z_n) \tag{2}$$

where the auxiliary function $g(\Theta, \Theta_0, z)$ is written as

$$g(\Theta, \Theta_0, z) = \frac{1}{4z}\left[\left(1 - y\left(\Theta^T\mathbf{x} + \Theta_0\right)\right) + z\right]^2 \tag{3}$$

In (3), $\mathbf{x}$ and y are generic placeholders for specific patterns $\mathbf{x}_n$ and y_n respectively. The minimum of $g(\Theta, \Theta_0, z)$ can be obtained as

$$z = \max\left\{\epsilon, \left|1 - y\left(\Theta^T\mathbf{x} + \Theta_0\right)\right|\right\} \tag{4}$$

where ϵ is a very small positive value chosen to avoid potential numerical problems in (3).

Implement a majorization scheme wherein you alternate between solutions for $\{\Theta, \Theta_0\}$ given $\{z_n\}$ and vice versa. The least-squares solution for $\{\Theta, \Theta_0\}$ was derived in class. Document your choices of C and ϵ.

2. Derive a majorization scheme for the multiclass kernel SVM. Use the one-versus-all objective function below with each hinge loss term containing the discrimination $\langle(\Theta_k - \Theta_l), \Phi(\mathbf{x}_n)\rangle + (\Theta_{0k} - \Theta_{0l})$ where k, l are class indices. That is, use the objective function

$$\frac{1}{2}\sum_{k=1}^{K}\langle\Theta_k, \Theta_k\rangle + C\sum_{k=1}^{K}\sum_{l\neq k}\sum_{n=1}^{N} y_{nk}\,\max\left[0, 1 - \left(\langle\Theta_k - \Theta_l, \Phi(\mathbf{x}_n)\rangle + \Theta_{k0} - \Theta_{l0}\right)\right]$$

where $y_{nk} = 1$ for $\mathbf{x}_n \in C_k$ and 0 otherwise. Explain how this objective function generalizes the two class objective function.

Use the expansion

$$\Theta_k = \sum_{n=1}^{N} \alpha_{nk}\Phi(\mathbf{x}_n)$$

for feature vector $\mathbf{x}_n$ where $\Phi(\mathbf{x}_n) \in \mathcal{H}$ with the kernel picking the particular RKHS. You do not need to write a program to test this majorization approach.