EE6550 Machine Learning, Spring 2016

Homework Assignment #4

In this homework assignment, there is no word problems but one programming problem. Please submit your programs, including source code, report and user manual, to iLMS by 13:00 on June 13 Monday. You are encouraged to consult or collaborate with other students while solving the problems but you will have to turn in your own solutions and programs with your own words. Copying will not be tolerated. If you consult or collaborate, you must indicate all of your consultants or collaborators with their contributions.

The Programming problem: Implementation of the kernel perceptron algorithm for binary classification. You have to use the MatLab functions in the toolboxes provided by the university version of MatLab to write your code. This means that you have to successfully run your MatLab program in the environment of the university version.

Input:

1. A data file which contains a labeled training sample S. This labeled training sample is used to train the kernel perceptron algorithm which will return a hypothesis h_S^{Perctn} after n-fold cross-validation.
2. A data file which contains a labeled testing sample $\tilde{S}$. This labeled testing sample is used to evaluate the performance of the returned hypothesis h_S^{Perctn} from the kernel perceptron algorithm based on the labeled training sample S.
3. Choice of the kernel function $K(\mathbf{x}, \mathbf{x}')$. There are two choices:
   (a) The standard inner product kernel: $K(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$.
   (b) The Gaussian kernel: $K(\mathbf{x}, \mathbf{x}') = \exp\left\{\frac{-\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right\}$.
4. Choice of n-fold cross-validation, where $n = 5$ or $n = 10$.
   The free model parameter vector $\boldsymbol{\theta}$ is (T, η) if the chosen kernel is the inner product kernel and (T, η, σ) if the chosen kernel is the Gaussian kernel.
   - η is the learning rate of the kernel perceptron algorithm at each tth round:
   $$\hat{y}_t \leftarrow \operatorname{sgn}\left(\sum_{s=1}^{T} \alpha_s \eta c(\mathbf{x}_s) K(\mathbf{x}_s, \mathbf{x}_t)\right).$$
   - The number of round T is usually a multiple of the size m of the training sample used. A stoping rule should apply to the case of a linearly separable training data set since the kernel perceptron algorithm will converge.

   As discussed in Lecture 1, we use n-fold cross-validation to determine the best value of the free parameter vector $\boldsymbol{\theta}$.
   - Randomly partition a given training sample S of m labeled items into n subsamples or folds.
   - $((\omega_{i1}, c(\omega_{i1})), ..., (\omega_{im_i}, c(\omega_{im_i})))$: the ith fold of size m_i, $1 \le i \le n$.
     - Usually $m_i = \frac{m}{n}$ for all i.

- For any $i \in [1, n]$, the kernel perceptron algorithm is trained on all but the ith fold to generate a hypothesis h_i, and the performance of h_i is tested on the ith fold.
- $\hat{R}_{CV}(\boldsymbol{\theta})$: the cross-validation error under the model parameter $\boldsymbol{\theta}$.

$$\hat{R}_{CV}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i} 1_{h_i(\omega_{ij})\neq c(\omega_{ij})} = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i} 1_{h_i(\omega_{ij})\neq c(\omega_{ij})}.$$

- Choose the parameter vector $\boldsymbol{\theta}^*$ which minimizes the cross-validation error $\hat{R}_{CV}(\boldsymbol{\theta})$.
- Train the kernel perceptron algorithm with the best parameter setting $\boldsymbol{\theta}^*$ over the full training sample S of size m. The resulted hypothesis will be the returned hypothesis h_S^{Perctn} from the kernel perceptron algorithm.

Output:

1. The optimal value of the free model parameter vector $\boldsymbol{\theta}$ for $n = 5$ and for $n = 10$ for the n-fold cross-validation.
2. The hypothesis h_S^{Perctn} returned by the kernel perceptron algorithm for $n = 5$ and for $n = 10$ for the n-fold cross-validation.
   - If the chosen kernel is the inner product kernel, then

   $$h_S^{Perctn}(\mathbf{x}) = \mathrm{sgn}(\mathbf{w}^{Perctn} \cdot \mathbf{x})$$

   where

   $$\mathbf{w}^{Perctn} = \sum_{s=1}^{T} \alpha_s \eta c(\mathbf{x}_s)\mathbf{x}_s$$

   is the weight vector returned by the kernel perceptron algorithm.
   - If the chosen kernel is the Gaussian kernel, then

   $$h_S^{Perctn}(\mathbf{x}) = \mathrm{sgn}\left(\sum_{s=1}^{T} \alpha_s \eta c(\mathbf{x}_s) K(\mathbf{x}_s, \mathbf{x})\right).$$

3. Performance evaluation of the returned hypothesis h_S^{Perctn} on the labeled testing sample $\tilde{S}$ for $n = 5$ and for $n = 10$ for the n-fold cross-validation.

What to submit? You should submit the following items:

1. The electronic source code of your kernel perceptron algorithm. (It is recommended for you to use MatLab to write your programs.)
2. A printed report consisting of at least:
   (a) You should use the two sets of training and testing data files given in HW#2 as the training data set and the testing data set respectively.
      - Set 1: adult_training.mat and adult_testing.mat.
        This is to predict the annual income of a person in US to be over 50K or not. Each row corresponds to an item with label indicated in the last column.

- Set 2: iris_set_ver_training.mat and iris_set_ver_testing.mat.
  This is to predict one of two types of iris plant. Each row corresponds to an item with label indicated in the last column.

(b) A table of the cross-validation error $\hat{R}_{CV}(\boldsymbol{\theta})$ as a function of the parameter vector $\boldsymbol{\theta}$ for $n = 5$ and for $n = 10$ for n-fold cross-validation. Discuss how do you determine the optimal value $\boldsymbol{\theta}^*$ from such a table.

(c) The hypothesis h_S^{Perctn} returned by the kernel perceptron algorithm with the best parameter vector $\boldsymbol{\theta}^*$ and its performance evaluation on the labeled testing sample $\tilde{S}$ for $n = 5$ and for $n = 10$ for the n-fold cross-validation.

3. A user manual which should include instructions of

   (a) how to compile the source code;

   (b) how to run the algorithm.