EE6550 Machine Learning, Spring 2016

Homework Assignment #5

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 30 Thursday. 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 sequential minimal optimization (SMO) algorithm for kernel-based support vector regression (SVR). Please read the supplement of this homework which gives a detailed description of how to implement the SMO algorithm for kernel based SVR. 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 based SMO-SVR algorithm which will return a hypothesis h_S^{SVR} 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^{SVR} from the SMO-SVR algorithm based on the labeled training sample S.
3. Set the values of the insensitivity parameter ϵ and the tolerance τ (see the supplement of this homework).
   - It is reasonable to preset the tolerance τ to be 1%-10% of ϵ.
   - It is suggestive to set ϵ to be 0.1 or 1%-10% of the range of the labels of the items in the sample S.
4. Choice of the kernel function $K(\mathbf{x}, \mathbf{x}')$. There are two choices:
   (a) The polynomial kernel of degree d: $K(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x} \cdot \mathbf{x}')^d$. (See page 12 of Lecture 5 on Kernel with $c = 1$.)
   (b) The Gaussian kernel: $K(\mathbf{x}, \mathbf{x}') = \exp\left\{\frac{-\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right\}$.
5. Choice of n-fold cross-validation, where $n = 5$ or $n = 10$.
   The free model parameter vector $\boldsymbol{\theta}$ is (C, d) if the chosen kernel is the polynomial kernel of degree d and (C, σ) if the chosen kernel is the Gaussian kernel. 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, $1 \leq i \leq n$.

- Usually $m_i = \frac{m}{n}$ for all i.

- For any $i \in [1, n]$, the SMO-SVR 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 SMO-SVR 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^{SVR} from the SMO-SVR 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^{SVR} returned by the SMO-SVR algorithm for $n = 5$ and for $n = 10$ for the n-fold cross-validation.
   - If the chosen kernel is the polynomial kernel, then

     $$h_S^{SVR}(\mathbf{x}) = \mathbf{w}^{SVR} \cdot \Phi(\mathbf{x}) + b^{SVR}$$

     where Φ is the feature mapping associated with the polynomial kernel of degree d and

     $$\mathbf{w}^{SVR} = \sum_{k=1}^{m} \beta_k^{SVR} \Phi(\mathbf{x}_k)$$

     is the weight vector returned by the SMO-SVR algorithm. For the computation of the offset b^{SVR} returned by the SMO-SVR algorithm, please see the supplement of HW#5.
   - If the chosen kernel is the Gaussian kernel K, then

     $$h_S^{SVR}(\mathbf{x}) = \sum_{k=1}^{m} \beta_k^{SVR} K(\mathbf{x}_k, \mathbf{x}) + b^{SVR}.$$

3. Performance evaluation of the returned hypothesis h_S^{SVR} 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 SMO-SVR 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 training and testing data files provided later as the training data set and the testing data set respectively.

(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^{SVR} returned by the SMO-SVR 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.