Local learning vs. Global learning

- Global learning methods consider the same weight for all training points in model fitting.
- Local learning methods assume that the training samples in the test point region are more influential.
- Moving Least Squares assumes the training samples may not have similar importance in function estimation.

Moving Least Squares Support Vector Machines (M-LSSVM)

Considering $x \in \mathbb{R}^d$, $y \in \mathbb{R}$ and $\phi(\cdot)$ being the feature map, the M-LSSVM model in primal space is formulated as below:

$$\hat{y}_i(x) = \hat{w}^T \phi(x) + \hat{b},$$

where $\hat{w}_i, \hat{b}_i \in \mathbb{R}$ and $\hat{w}_i \in \mathbb{R}^d$ are estimated for a given $x$.

Given $s_j(x)$ as a non-negative similarity measure between the $i$th training data feature vector and any fixed $x$, the optimization problem for training M-LSSVM model in primal space is as follows:

$$\begin{align*}
\min_{\hat{w}, \hat{b}} & \quad \frac{1}{N} \sum_{i=1}^{N} s_i(x)^2 e_i^2, \\
\text{s.t.} & \quad y_i = \phi^T \phi(x) + b + e_i, i = 1, \ldots, N.
\end{align*}$$

The optimality conditions can be expressed as below.

$$\begin{align*}
\frac{\partial \mathcal{L}}{\partial w} &= 0 = \sum_{i=1}^{N} \alpha_i s_i(x) e_i^2, \\
\frac{\partial \mathcal{L}}{\partial b} &= 0 = \sum_{i=1}^{N} \alpha_i e_i, \\
\frac{\partial \mathcal{L}}{\partial \alpha_i} &= 0 = \gamma s_i(x) e_i + b + e_i, i = 1, \ldots, N,
\end{align*}$$

where $\alpha_i \in \mathbb{R}$ are the Lagrange multipliers. The dual problem is written as follows:

$$\begin{pmatrix} 0 \\ \frac{1}{N} \sum_{i=1}^{N} s_i(x) \end{pmatrix} \begin{pmatrix} \hat{b} \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}.$$  (4)

After solving the linear equation, the M-LSSVM function estimator is written as below:

$$\hat{y}_i(x) = \sum_{i=1}^{N} \hat{\alpha}_i s_i(x) + \hat{b}.$$  (5)

Here we investigate two similarity criterion:

1. Gaussian similarity: $s_i(x) = \exp(-||x - x_i||^2 / \gamma^2)$
2. Cosine-based similarity: $s_i(x) = x_i^T x / ||x_i|| ||x|| + 1$

Tuning the parameters

k-fold Moving Cross Validation (M-CV):

$$\text{Error}_{M-CV}(x) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} s_i(x) \text{err}_j}{\sum_{i=1}^{N} s_i(x)},$$

where $\text{err}_j$ is a performance evaluation criterion.

Dataset

The data have been collected from the Weather Underground website and include real measurements for 11 cities. It covers a time period from beginning 2007 to mid 2014 and comprise 198 measured weather variables for each day. The performance of the proposed method is evaluated on minimum and maximum temperature prediction in two test sets: (i) from mid-November 2013 to mid-December 2013 (Nov-Dec) and (ii) from mid-April 2014 to mid-May 2014 (Apr-May).

Experiments

Table 1: Average MAE and MSE of the predictions by LSSVM and M-LSSVM based on Cosine and RBF similarity $s_i(x)$ for test sets Nov-Dec and Apr-May.

<table>
<thead>
<tr>
<th>Test set</th>
<th>Days ahead</th>
<th>LSSVM</th>
<th>M-LSSVM</th>
<th>RBF</th>
<th>Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov-Dec</td>
<td>Max</td>
<td>2.18</td>
<td>1.97</td>
<td>2.11</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>1.39</td>
<td>1.17</td>
<td>1.85</td>
<td>1.41</td>
</tr>
<tr>
<td>Apr-May</td>
<td>Max</td>
<td>2.43</td>
<td>2.17</td>
<td>2.45</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>1.38</td>
<td>1.17</td>
<td>1.96</td>
<td>1.41</td>
</tr>
</tbody>
</table>

Table 1: Average MAE and MSE of the predictions by LSSVM and M-LSSVM based on Cosine and RBF similarity $s_i(x)$ for test sets Nov-Dec and Apr-May.

<table>
<thead>
<tr>
<th>Test set</th>
<th>Days ahead</th>
<th>LSSVM</th>
<th>M-LSSVM</th>
<th>RBF</th>
<th>Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov-Dec</td>
<td>Max</td>
<td>2.18</td>
<td>1.97</td>
<td>2.11</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>1.39</td>
<td>1.17</td>
<td>1.85</td>
<td>1.41</td>
</tr>
<tr>
<td>Apr-May</td>
<td>Max</td>
<td>2.43</td>
<td>2.17</td>
<td>2.45</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>1.38</td>
<td>1.17</td>
<td>1.96</td>
<td>1.41</td>
</tr>
</tbody>
</table>

Figure 1: MAE of the predictions for Weather Underground, LSSVM and M-LSSVM with RBF and cosine based similarity $s_i(x)$ for Max. and Min. temperature.

References