Moving Least Squares Support Vector Machines for weather temperature prediction

Z. Karevan, Y. Feng, J. A. K. Suykens

Department of Electrical Engineering (ESAT/STADIUS), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium.

Abstract

Local learning methods have been investigated by many researchers. While 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. In this paper, we propose Moving Least Squares Support Vector Machines (M-LSSVM) in which each training sample is involved in the model fitting depending on the similarity between its feature vector and the one of the test point. The experimental results on an application of weather forecasting indicate that the proposed method can improve the prediction performance.

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

  → each training point has a weight based on the distance between the training sample and the test point

$$\sum_{i=1}^{N}(f(x_i)-y_i)^2 S_i(x,x_i)$$

Moving Least Squares Support Vector Machines (M-LSSVM)

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

$$\hat{y}_x(x) = \hat{w}_x^T \varphi(x) + \hat{b}_x, \tag{1}$$

where $\hat{b}_x \in \mathbb{R}$ and $\hat{w}_x \in \mathbb{R}^h$ are estimated for a given x.
Given $s_i(x)$ as a non-negative similarity measure between the ith training data feature vector and any fixed x, the optimization problem for training M-LSSVM model in primal space is as follows:

$$(\hat{w}_x, \hat{b}_x, \hat{e}_x) = \min_{w,b,e} \tfrac{1}{2} w^T w + \tfrac{\gamma}{2}\sum_{i=1}^{N} s_i(x) e_i^2, \quad \text{s.t. } y_i = w^T\varphi(x_i) + b + e_i, i = 1, \dots, N. \tag{2}$$

The optimality conditions can be expressed as below.

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i), \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0, \\ \frac{\partial \mathcal{L}}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma s_i(x) e_i, i = 1, \dots, N, \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \rightarrow y_i = w^T\varphi(x_i) + b + e_i, i = 1, \dots, N, \end{cases} \tag{3}$$

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

$$\begin{pmatrix} 0 & 1_N^T \\ 1_N & \Omega + S_\gamma(x) \end{pmatrix} \begin{pmatrix} b \\ \alpha_i \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}. \tag{4}$$

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

$$\hat{y}_x(x) = \sum_{i=1}^{N} \hat{\alpha}_{i,x} K(x, x_i) + \hat{b}_x. \tag{5}$$

Here we investigate two similarity criterion:

1. Gaussian similarity: $s_i(x) = \exp(-||x - x_i||_2^2 / h^2)$
2. Cosine-based similarity: $s_i(x) = \frac{x^T x_i}{||x|| \times ||x_i||} + 1$

Tuning the parameters

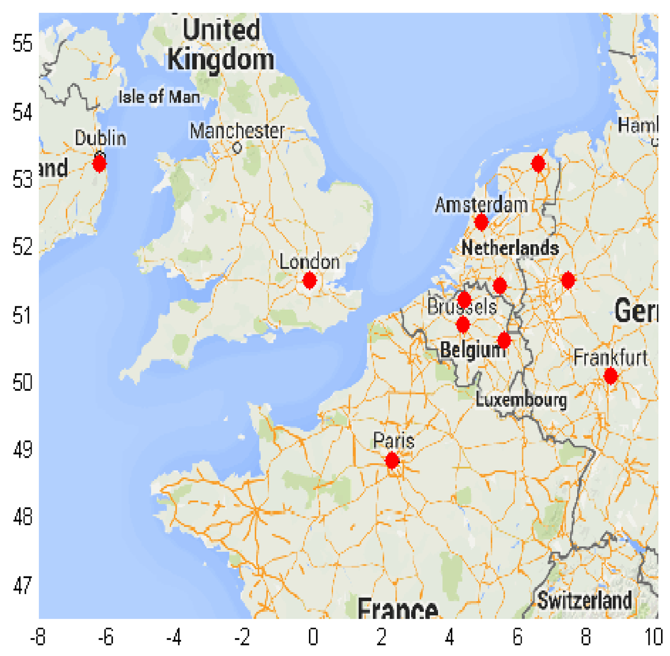
k-fold Moving Cross Validation (M-CV):

$$Error_{M-CV}(x) = \frac{\sum_{v=1}^{k}\sum_{i=1}^{N_v} s_i(x)\, err_i}{\sum_{i=1}^{N} s_i(x)}, \tag{6}$$

where err_i is a performance evaluation criterion.

Dataset

The data have been collected from the Weather Underground website and include real measurements for 11 cities. It cover 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 min-May 2014 (Apr/May).



Experiments

Test set	Days ahead	Temp.	Mean Absolute Error LSSVM	Mean Absolute Error M-LSSVM RBF	Mean Absolute Error M-LSSVM Cosine	Mean Square Error LSSVM	Mean Square Error M-LSSVM RBF	Mean Square Error M-LSSVM Cosine
Nov/Dec	1	Min	1.44	1.44	**1.35**	3.70	3.65	**3.53**
		Max	1.19	**1.03**	1.13	2.88	**2.14**	2.53
	2	Min	**1.55**	**1.55**	1.57	**4.47**	**4.47**	4.50
		Max	1.23	1.26	**1.20**	3	3.11	**2.90**
	3	Min	1.65	1.67	**1.62**	**5.26**	**5.26**	5.29
		Max	1.48	**1.30**	1.48	4.19	**3.76**	3.98
	4	Min	**1.61**	**1.61**	1.64	4.53	**4.50**	4.61
		Max	1.37	1.37	**1.32**	**2.82**	**2.82**	**2.82**
	5	Min	**1.50**	1.53	1.53	**3.96**	4	4
		Max	**1.12**	1.17	**1.12**	1.94	2.10	**1.90**
	6	Min	1.69	1.63	**1.62**	4.69	**4.55**	**4.55**
		Max	1.44	1.41	**1.41**	**3.96**	**3.96**	4.08
Apr/May	1	Min	1.46	**1.29**	1.39	3.22	**3.03**	3.24
		Max	2.25	2.17	**2.09**	7.46	7.28	**7.04**
	2	Min	1.85	**1.80**	1.82	6.39	6.25	**6.19**
		Max	2.16	2.11	**2.08**	7.54	**7.38**	**7.38**
	3	Min	1.70	**1.65**	1.70	5.33	**5.14**	5.33
		Max	**2.43**	**2.43**	2.48	**8.11**	8.17	8.68
	4	Min	1.74	**1.62**	1.75	4.85	**4.65**	4.93
		Max	**2.44**	2.49	2.48	**8.54**	8.62	9.38
	5	Min	**1.96**	**1.96**	**1.96**	**6.40**	6.48	6.60
		Max	2.33	2.23	**2.20**	7.43	7.57	**7.34**
	6	Min	2.18	**2.08**	2.18	8.03	**7.80**	8.03
		Max	**2.50**	**2.50**	2.55	**8.43**	**8.43**	9.31

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/Nov.*

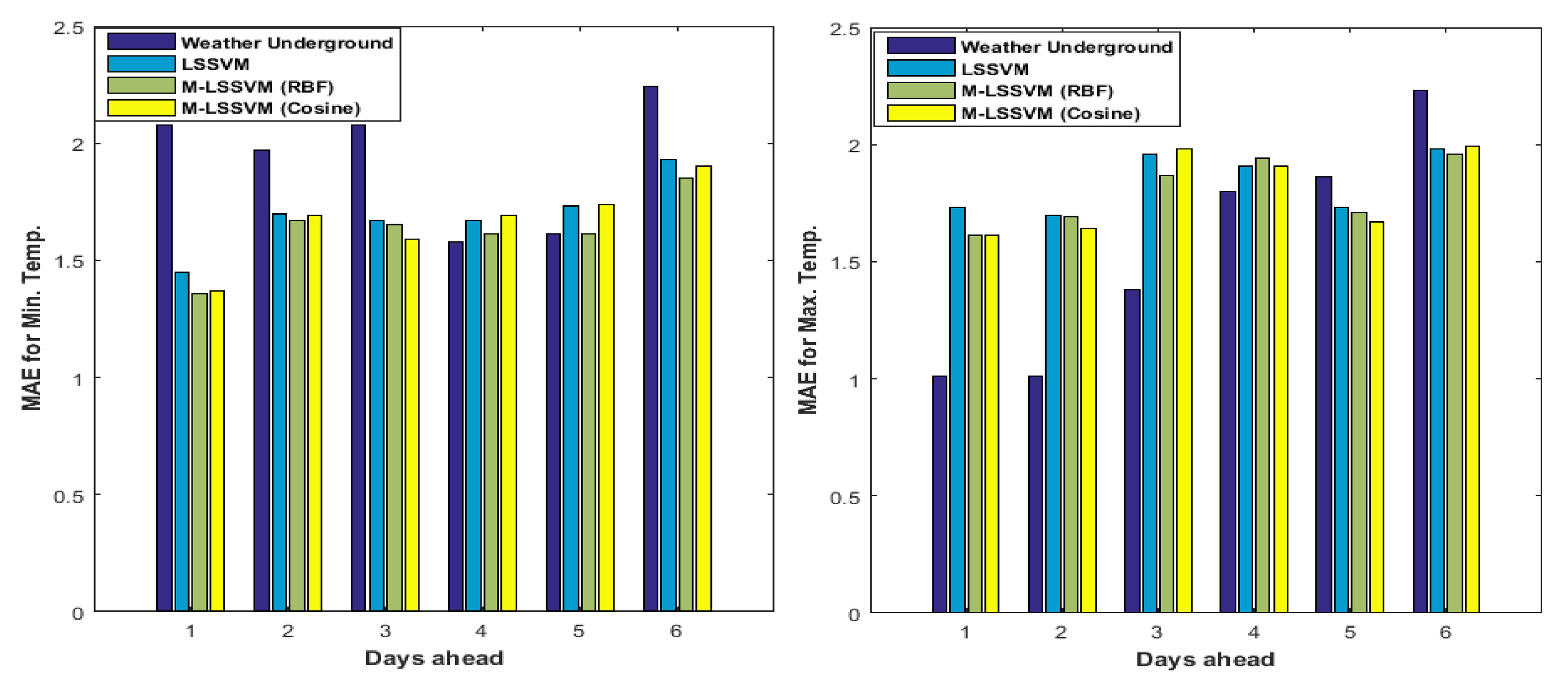


Figure 1: *MAE of the predictions for Weather Underground, LSSVM and M-LSSVM with RBF and cosine based similarity for Max. and Min. temperature*

References

[1] Z. Karevan, Y. Feng, J. A. K. Suykens "Moving Least Squares Support Vector Machines for weather temperature prediction", Accepted for publication in Proc. of the ESANN 2017.

[2] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle "Least squares support vector machines", World Scientific, 2002.

[3] Clive Loader "Local regression and likelihood", Springer Science & Business Media, 2006.

[4] Weather Underground Co. (www.wunderground.com)