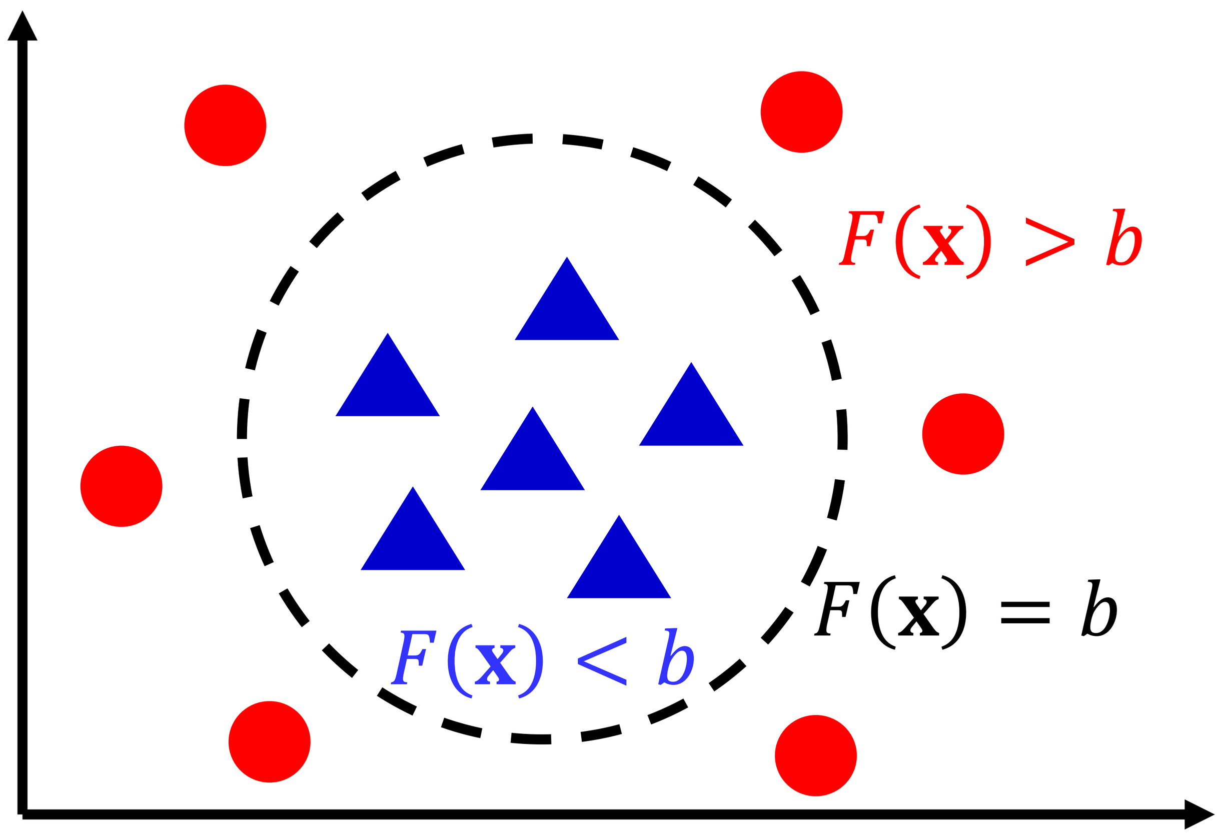
# PLAME: Piecewise-Linear Approximate Measure for Additive Kernel SVM


Tsz Nam Chan, Shenzhen University, edisonchan@szu.edu.cn
Zhe Li, Alibaba Cloud, huoju.lz@alibaba-inc.com
Leong Hou U, University of Macau, ryanlhu@um.edu.mo
Reynold Cheng, The University of Hong Kong, ckcheng@cs.hku.hk


## Overview of Kernel SVM




$F(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i)$ is the kernel SVM classifier, where:

- $(\mathbf{x}_i, y_i)$ denotes the $i^{\text{th}}$ training data point ($y_i$ can be +1 or -1).
- $n$ denotes the number of training data points.
- $d$ denotes the dimensionality of training data points.

Goal: Obtain $\alpha_i$ and $b$.

## Overview of Additive Kernels

| Kernel name | $K(\mathbf{x}_i, \mathbf{x})$ |
|---|---|
| $\chi^2$ | $\sum_{\ell=1}^{d} \frac{2x_i^{(\ell)} x^{(\ell)}}{x_i^{(\ell)} + x^{(\ell)}}$ |
| JS | $\sum_{\ell=1}^{d} \frac{1}{2} x_i^{(\ell)} \log_2\left(\frac{x_i^{(\ell)} + x^{(\ell)}}{x_i^{(\ell)}}\right) + \frac{1}{2} x^{(\ell)} \log_2\left(\frac{x_i^{(\ell)} + x^{(\ell)}}{x^{(\ell)}}\right)$ |
| Intersection | $\sum_{\ell=1}^{d} \min(x_i^{(\ell)}, x^{(\ell)})$ |
| Hellinger | $\sum_{\ell=1}^{d} \sqrt{x_i^{(\ell)} x^{(\ell)}}$ |

Additive kernels are demonstrated to be useful in many application domains, including human activity detection and pedestrian detection.

Guo et al. [a] "... we train a SVM classifier with chi-square kernel for multi-class recognition task, which is beneficial for classifying the histogram features."


[a] Y. Guo, Y. Li, and Z. Shao. DSRF: A flexible trajectory descriptor for articulated human action recognition. Pattern Recognition, 76: 137-148, 2018.


## Existing Methods of Additive Kernel SVM and PLAME

| Method | Classification error | Memory space | Training time |
|---|---|---|---|
| Kernel SVM solver [13] | low | low | high |
| Linear SVM solver [23], [28] | high | low | low |
| Feature approximation [35], [47], [59], [60] [7], [17], [33], [64] | high | high | low |
| Function approximation [69], [71], [75] | high | low | low |
| PLAME (ours) | low | low | low |

Existing SVM training methods (those references in the above table can be found in our TKDE paper.) cannot simultaneously fulfill these three conditions.

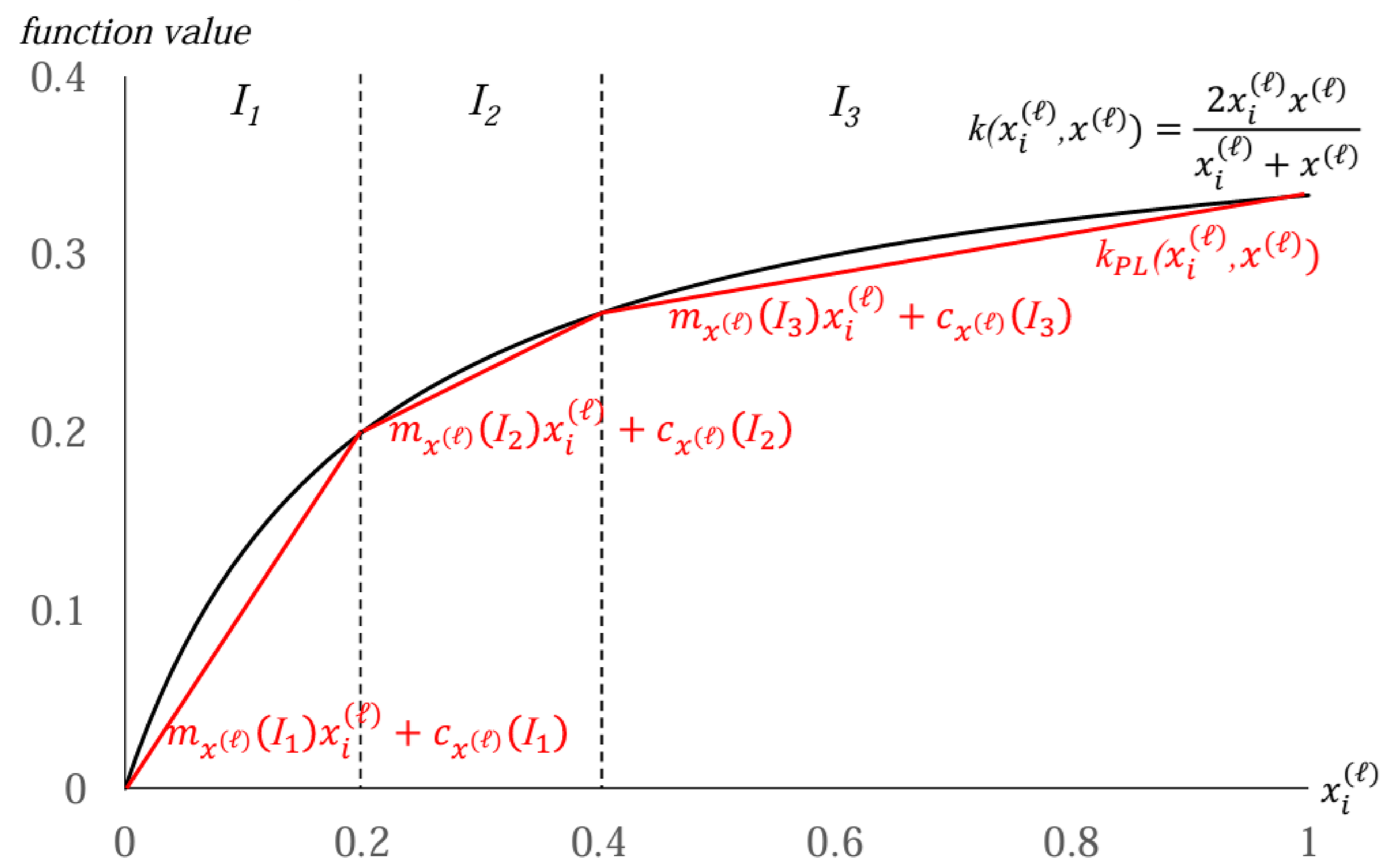
- Low classification error
- Low memory space
- Low training time

PLAME is the **first approach** that can achieve these three conditions.

Our observation: Every additive kernel can be represented by $d$ one-dimensional additive kernel functions.

$$K(\mathbf{x}, \mathbf{x}_i) = \sum_{\ell=1}^{d} k(x_i^{(\ell)}, x^{(\ell)})$$

Core idea of PLAME: Use a **piecewise-linear function** to approximate $k(x_i^{(\ell)}, x^{(\ell)})$ and modify the linear SVM solver.



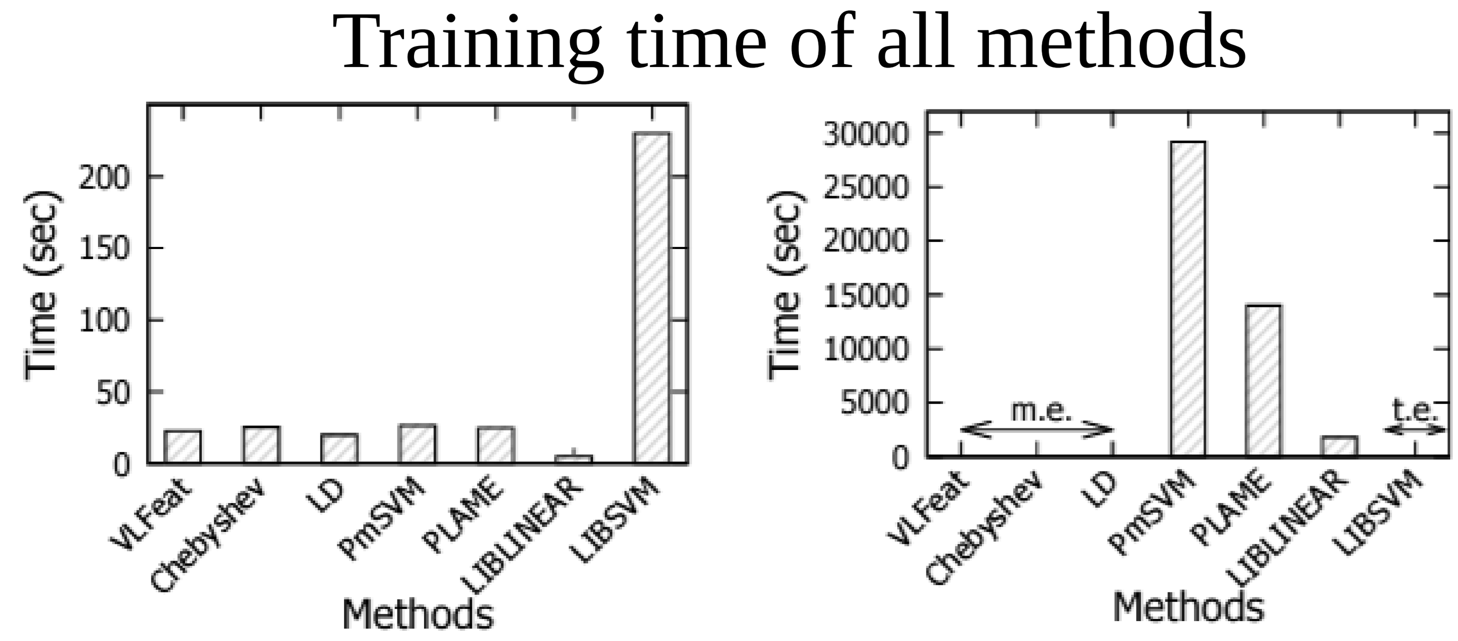

## Experimental Evaluation

Accuracy of all methods

| Method | LIBSVM | LIBLINEAR | VLFeat | Chebyshev | LD | PmSVM | PLAME |
|---|---|---|---|---|---|---|---|
| skin | **0.992** | 0.895 | 0.927 | 0.942 | 0.949 | 0.919 | **0.988** |
| casas | t.e. | 0.71 | m.e. | m.e. | m.e. | **0.728** | **0.78** |

Remark:
t.e.: more than three days for training
m.e.: more than 16 GB for training

Training time of all methods




(a) skin (b) casas