Support Vector Machines for Wildfire Identification

Jakub Kruzik, Marek Pecha, David Horak
jakub.kruzik@ugn.cas.cz

January 14, 2025

Institute of Geonics, Ostrava

PermonSVM

Support Vector Machines (SVMs)

- Machine learning
    - Supervised learning models

Support Vector Machines (SVMs)

- Machine learning
  - Supervised learning models
- Non-probabilistic binary linear classifier
  - Divides dataset into two classes/categories
  - Non-linear classifier with the kernel trick (mapping dataset into a higher dimension)

Support Vector Machines (SVMs)

- Machine learning
    - Supervised learning models
- Non-probabilistic binary linear classifier
    - Divides dataset into two classes/categories
    - Non-linear classifier with the kernel trick (mapping dataset into a higher dimension)
- Requires a training set

SVM – Learning Phase

Problem specification

Given a training dataset,

$$\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)\} \qquad \boldsymbol{x}_i \in \mathbb{R}^p, y_i \in \{-1, +1\},$$

find the **maximum–margin hyperplane** separating the training dataset into two classes,

$$(\boldsymbol{x}_i, +1) \qquad \text{and} \qquad (\boldsymbol{x}_j, -1).$$

Any hyperplane can be written as

$$\{\boldsymbol{x} \in \mathbb{R}^p : \boldsymbol{w}^T \boldsymbol{x} - b = 0\},$$

where $\boldsymbol{w}$ is a normal vector to the hyperplane and $\frac{b}{||\boldsymbol{w}||}$ is the distance of the hyperplane from the origin along the normal vector $\boldsymbol{w}$.

SVM – Classification Phase

Problem specification

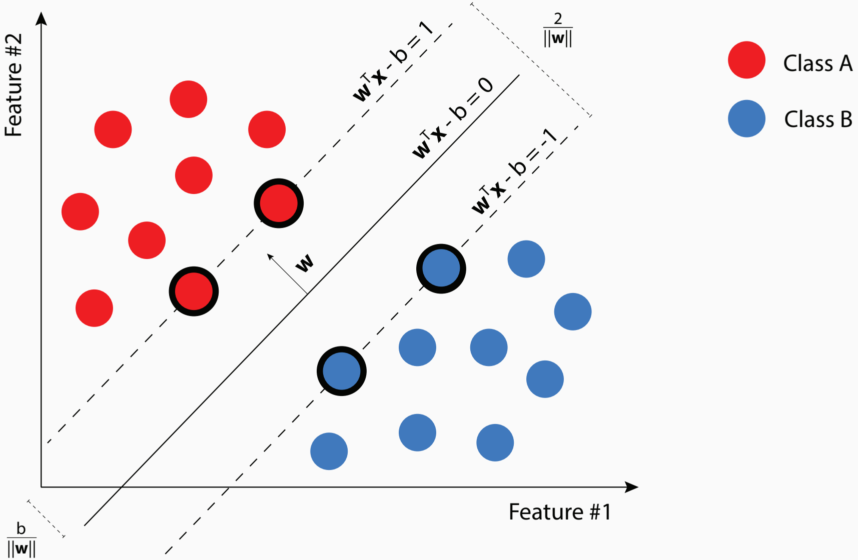
Given a dataset,

$$\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n\} \qquad \boldsymbol{z}_i \in \mathbb{R}^p,$$

and a **model** (determined by $\boldsymbol{w}$ and b), classify each dataset point into one of the two classes.

$$\begin{aligned} \boldsymbol{w}^T\boldsymbol{z}_i - b > 0 \quad &\Longrightarrow \quad \boldsymbol{z}_i \in \text{class } y = +1 \\ \boldsymbol{w}^T\boldsymbol{z}_i - b \leq 0 \quad &\Longrightarrow \quad \boldsymbol{z}_i \in \text{class } y = -1 \end{aligned}$$

SVM – Example with Linearly Separable Training Set

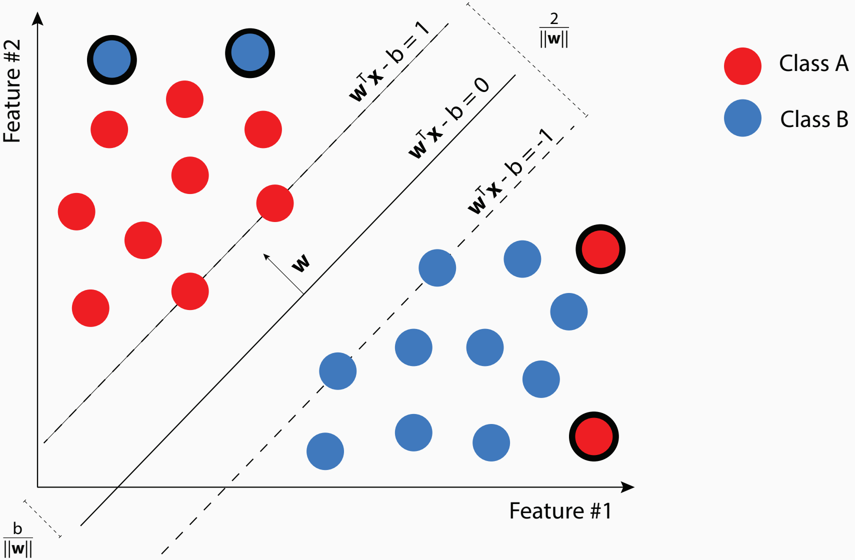


SVM – Formulations for Finding Maximum–margin Hyperplane (L1 loss)

- Linearly separable training examples (hard–margin)

$$\underset{\substack{\boldsymbol{w}\in\mathbb{R}^p \\ b\in\mathbb{R}}}{\arg\min}\ \frac{1}{2}||\boldsymbol{w}||^2 \qquad \text{s.t.} \qquad y_i(\boldsymbol{w}^T\boldsymbol{x}_i - b) \geq 1$$

SVM – Example with Non–separable Training Set



SVM – Formulations for Finding Maximum–margin Hyperplane (L1 loss)

- Linearly separable training examples (hard–margin)

$$\underset{\substack{\boldsymbol{w}\in\mathbb{R}^p \\ b\in\mathbb{R}}}{\arg\min} \frac{1}{2}||\boldsymbol{w}||^2 \qquad \text{s.t.} \qquad y_i(\boldsymbol{w}^T\boldsymbol{x}_i - b) \geq 1$$

- Non–separable training examples (soft–margin)

$$\underset{\substack{\boldsymbol{w}\in\mathbb{R}^p \\ b\in\mathbb{R} \\ \boldsymbol{\xi}\in\mathbb{R}^n}}{\arg\min} \left[\frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{n}\xi_i\right] \qquad \text{s.t.} \qquad \begin{array}{c} y_i(\boldsymbol{w}^T\boldsymbol{x}_i - b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{array}$$

  - $C > 0$ is a prescribed constant
  - $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots \xi_n]^T$ are slack variables
  - $$\xi_i = \max\{0, 1 - y_i\left(\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle - b\right)\}$$

    is also know as hinge loss function

SVM – Dual Soft–margin QP Formulation (L1 loss)

$$\boldsymbol{X}^T = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$$

$$\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]^T, \quad \boldsymbol{Y} = diag\,[y_1, y_2, \ldots, y_n], \quad \boldsymbol{e} = [1, 1, \ldots, 1]^T$$

$$\arg\min_{\boldsymbol{\lambda} \in \mathbb{R}^n} \frac{1}{2} \boldsymbol{\lambda}^T \boldsymbol{Y}^T \boldsymbol{X} \boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{\lambda} - \boldsymbol{\lambda}^T \boldsymbol{e} \qquad \text{s.t.} \qquad \begin{array}{c} \boldsymbol{y}^T \boldsymbol{\lambda} = 0 \\ \boldsymbol{o} \leq \boldsymbol{\lambda} \leq C\boldsymbol{e} \end{array}$$

Reconstruction:

$$\boldsymbol{w} = \boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{\lambda}$$

$$b = \frac{1}{|I^{SV}|} \sum_{i \in I^{SV}} \left(y_i - \boldsymbol{w}^T \boldsymbol{x}_i \right), \qquad I^{SV} = \{ i \in \{1, \ldots, n\} : 0 < \lambda_i < C \}$$

SVM – Soft–margin QP Formulation (L2 loss)

Primal formulation:

$$\underset{\substack{\boldsymbol{w}\in\mathbb{R}^p\\ b\in\mathbb{R}\\ \boldsymbol{\xi}\in\mathbb{R}^n}}{\arg\min} \left[\frac{1}{2}||\boldsymbol{w}||^2 + \frac{C}{2}\sum_{i=1}^{n}\xi_i^2\right] \qquad \text{s.t.} \qquad y_i(\boldsymbol{w}^T\boldsymbol{x}_i - b) \geq 1 - \xi_i$$

Dual QP formulation:

$$\underset{\boldsymbol{\lambda}\in\mathbb{R}^n}{\arg\min}\, \frac{1}{2}\boldsymbol{\lambda}^T\left(\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y} + C^{-1}\boldsymbol{I}\right)\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\boldsymbol{e} \qquad \text{s.t.} \qquad \begin{aligned} \boldsymbol{y}^T\boldsymbol{\lambda} &= 0 \\ \boldsymbol{o} &\leq \boldsymbol{\lambda} \end{aligned}$$

SVM – No–bias Formulation

Incorporate bias $\boldsymbol{b}$ into $\boldsymbol{w}$ by augmenting samples by one dimension:

$$\widehat{\boldsymbol{w}} = \begin{pmatrix} \boldsymbol{w} \\ b \end{pmatrix}, \quad \widehat{\boldsymbol{x}}_i = \begin{pmatrix} \boldsymbol{x}_i \\ -1 \end{pmatrix}.$$

Primal formulation for $p = 1, 2, \ldots$:

$$\operatorname*{arg\,min}_{\widehat{\boldsymbol{w}},\, b,\, \boldsymbol{\xi}} \left[\frac{1}{2} ||\widehat{\boldsymbol{w}}||^2 + \frac{C}{p} \sum_{i=1}^{n} \xi_i^p \right] \qquad \text{s.t.} \qquad y_i(\widehat{\boldsymbol{w}}^T \boldsymbol{x}_i) \geq 1 - \xi_i$$

Dual QP for $p = 1$:

$$\operatorname*{arg\,min}_{\boldsymbol{\lambda}} \frac{1}{2} \boldsymbol{\lambda}^T \boldsymbol{Y}^T \widehat{\boldsymbol{X}} \widehat{\boldsymbol{X}}^T \boldsymbol{Y} \boldsymbol{\lambda} - \boldsymbol{\lambda}^T \boldsymbol{e} \qquad \text{s.t.} \qquad \boldsymbol{o} \leq \boldsymbol{\lambda} \leq C \boldsymbol{e}$$

and for $p = 2$:

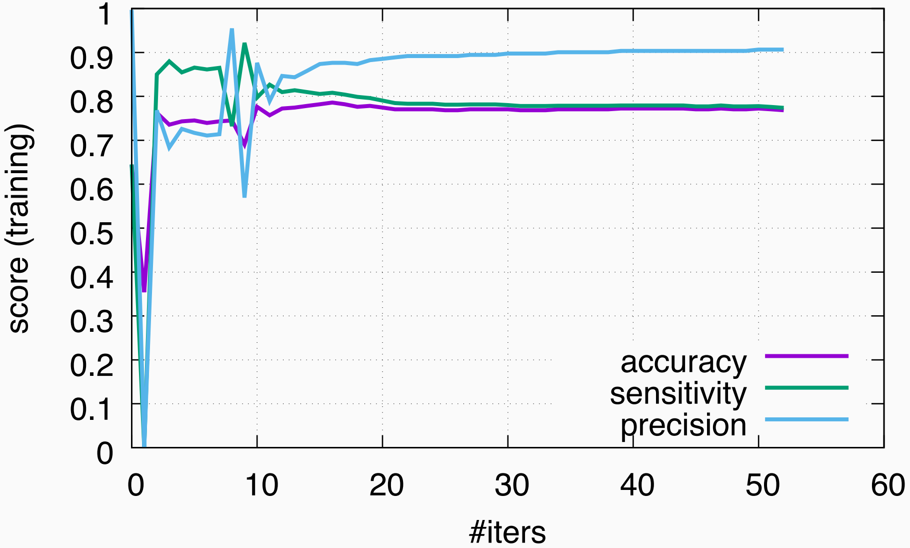
$$\operatorname*{arg\,min}_{\boldsymbol{\lambda}} \frac{1}{2} \boldsymbol{\lambda}^T \left(\boldsymbol{Y}^T \widehat{\boldsymbol{X}} \widehat{\boldsymbol{X}}^T \boldsymbol{Y} + C^{-1} \boldsymbol{I} \right) \boldsymbol{\lambda} - \boldsymbol{\lambda}^T \boldsymbol{e} \qquad \text{s.t.} \qquad \boldsymbol{o} \leq \boldsymbol{\lambda}$$

PermonSVM - Features

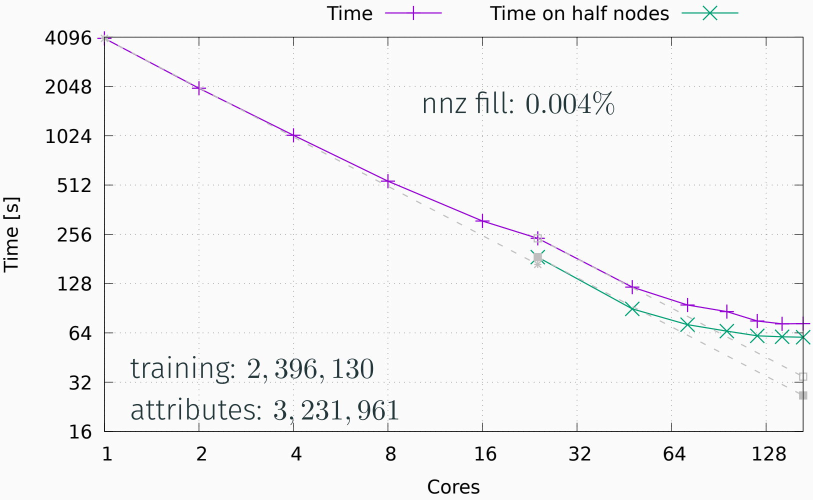
- Solves bias and no bias dual formulation of linear SVM
- Grid search
- Monitors for accuracy, recall, specifity, F1,...
- Parallel loading of datasets (LIBSVM (plain text) or HDF5)
- Future: Probability SVM, Multiclass, Multilabel
- Use as a C library (#include <permonsvm.h>) or bin:

```
mpirun -n 24 permonsvmfile -f_train train \
-f_test test \
-svm_view_score \
-smalxe_qps_type tao -smalxe_qps_tao_type blmvm
```

PermonSVM - Monitors



PermonSVM – MPGP Scalability, URL Dataset – Salomon (IT4I)



Data source: LIBSVM Data – J. Ma et al., Identifying suspicious URLs: An application of large-scale online learning (2014).

Localization of Wildfires from Satellite Images

M. Pecha[1], Z. Langford[2], D. Horák[1,3], R. T. Mills[4], J. Kružík[1,3]

[1] Institute of Geonics of the Czech Academy of Sciences
[2] Oak Ridge National Laboratory
[3] VSB - Technical University of Ostrava
[4] Argonne National Laboratory

Motivation – Monitoring Trends in Burn Severity (MTBS)

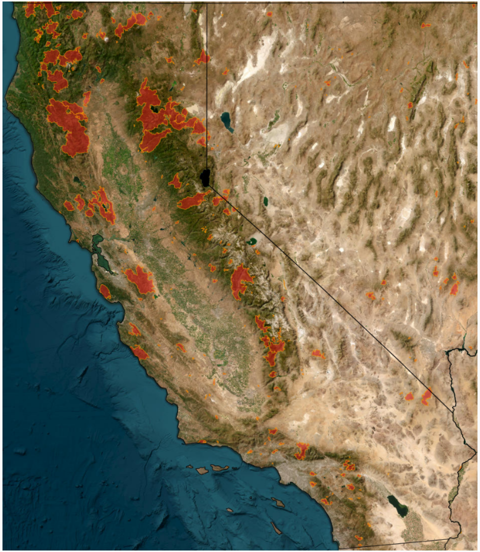
Mapping severity and extent of wildfires

Usage examples:

- Evacuation routes
- Ecological restoration
  - 84+ % of wildfires are caused by humans
- Climate modelling

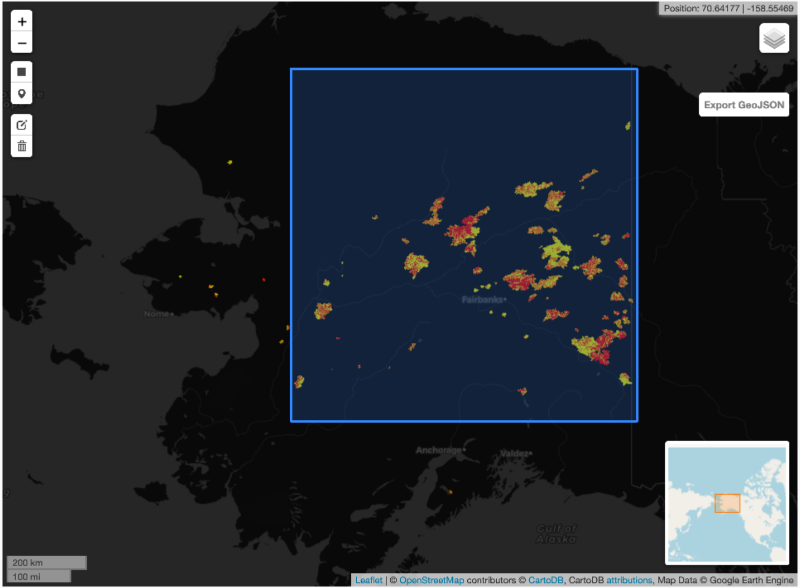
Limitations:

- Fires over 4 km^2 (eastern USA)
- Errors in localization
- Many hours of human labour
- About 2 years behind
- (Government shutdown)

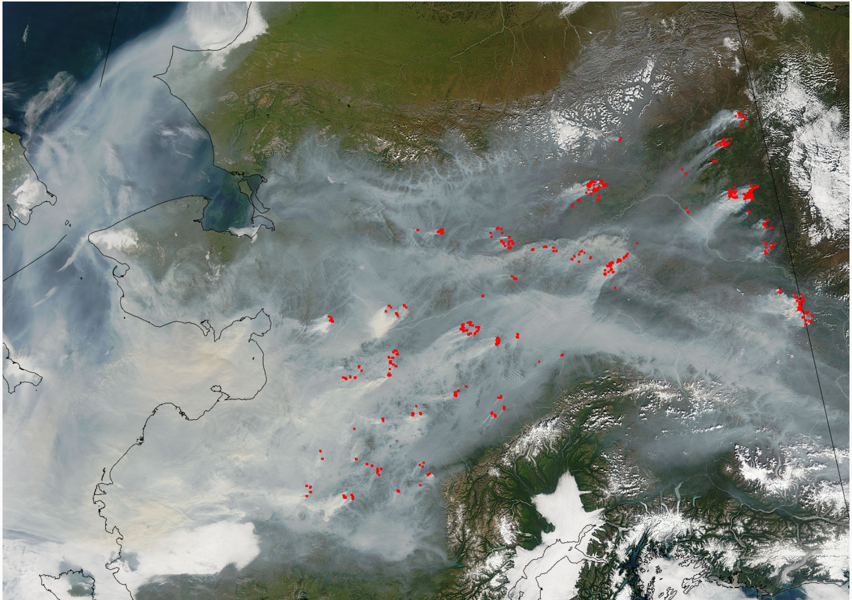


Wildfires in California 2019–2023

Dataset: 2004 Alaska Wildfire Season (Selected Area: 722 500 km^2)



Satellite Images – Alaska Wildfires 2005-08-14



Data – Moderate Resolution Imaging Spectroradiometer (MODIS)

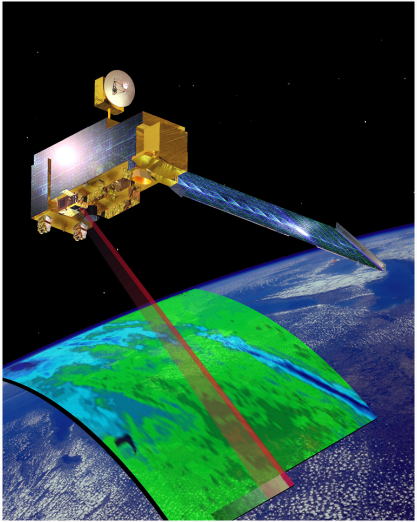
Spectroradiometer carried by Terra (EOS AM-1) a Aqua (EOS PM-1) satellites

MOD09A1 dataset: best values over 8-day period

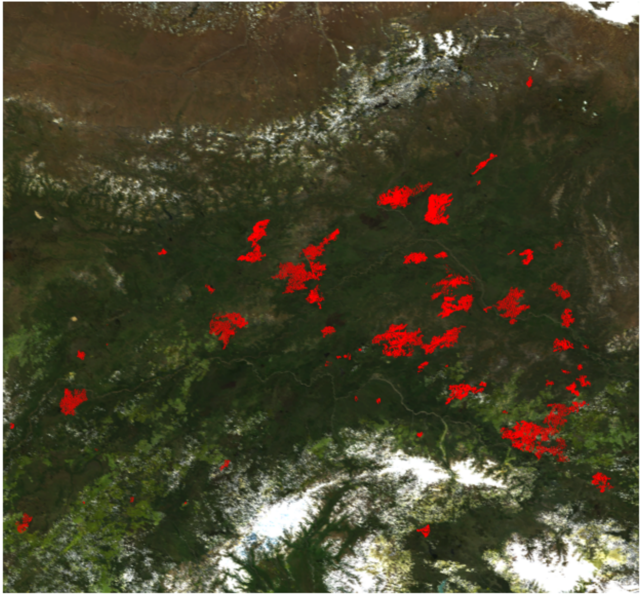
Resolution $500 \times 500\ \text{m}^2 = 0.25\ \text{km}^2$

Surface reflectance:

Band	Wave Length [nm]	Description
1	620–670	red
2	841–876	NIR
3	459–479	blue
4	545–565	green
5	1230–1250	SWIR1
6	1628–1652	SWIR2
7	2105–2155	SWIR3



Example of Data – MODIS visible spectrum + MTBS (1918 $\times$ 1780 px)

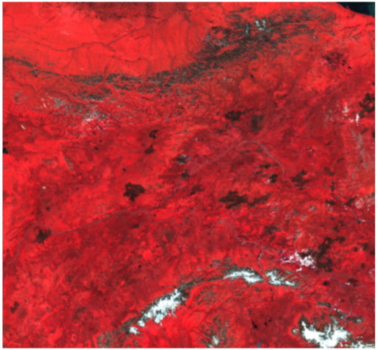


Additional Data

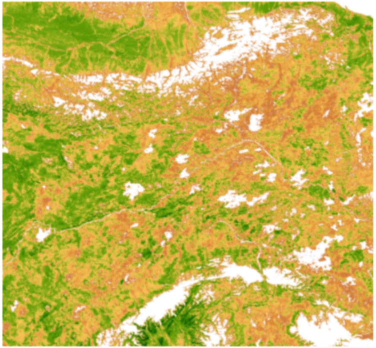
Vegetation indices computed from reflectance:

- Enhanced vegetation index 1, 2 (EVI1, EVI2)
- Normalized difference vegetation index (NDVI)

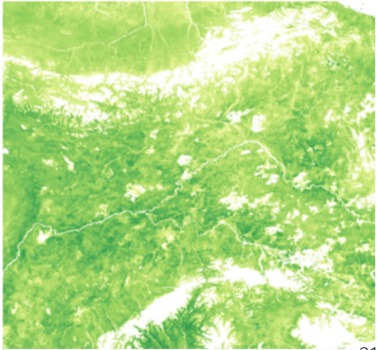
Temperature



Near-infrared (NIR)



EVI1



NDVI

Data

46 measurements per year (time-series of images)

Each measurement k for each pixel i adds sample:

$$\boldsymbol{X}_{i,:} = \begin{bmatrix} \boldsymbol{m}_{1,i} & \boldsymbol{m}_{2,i} & \dots & \boldsymbol{m}_{46,i} \end{bmatrix},$$

where

$$\boldsymbol{m}_{k,i} = \begin{bmatrix} \text{7 reflectance bands} & \text{3 vegetation indices} & \text{temperature} \end{bmatrix}_{k,i}$$

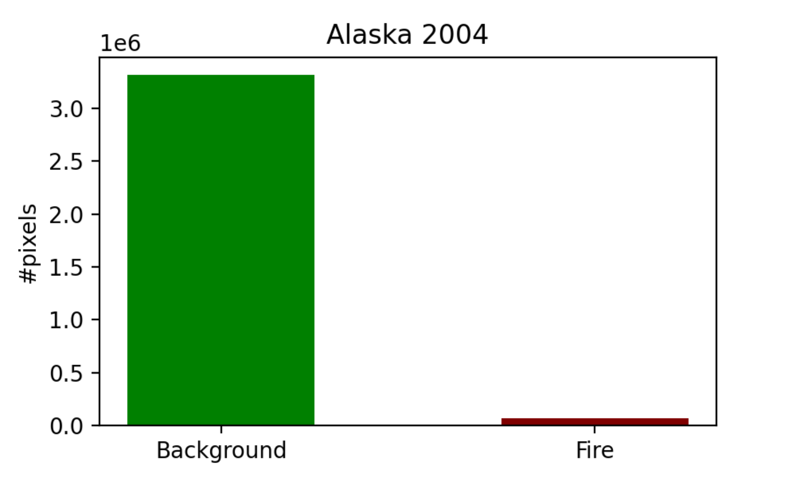
$$\#\text{samples} = \#\text{pixels} = 1918 \times 1780 = 3,414,040$$

$$\#\text{features} = \#\text{measurements} \times \text{length } \boldsymbol{m_k} = 46 \times 11 = 506$$

Each pixel i is marked based on MTBS if there was a wildfire:

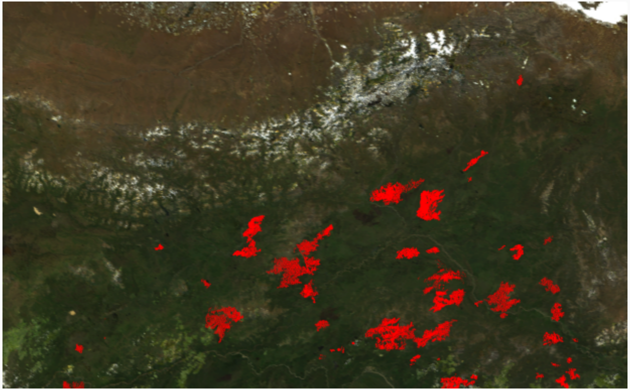
$$\boldsymbol{y}_i = +1 \text{ or } -1$$

Highly Unbalanced Dataset

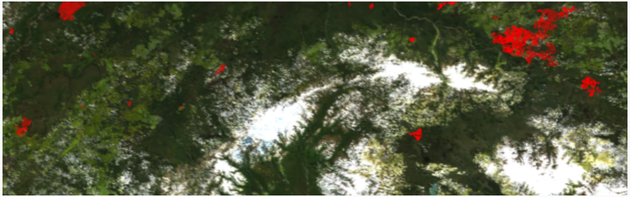


$\mathbf{3,317,870}$ (97.92%) of background pixels and $\mathbf{70,631}$ (2.08%) of wildfire pixels

Training and testing data (2:1)



Training data



Testing data

Data Preprocessing

Savitzky–Golay filter:

- Smoothing the time series (for each band)
- Quadratic polynomial
- Window length = 9

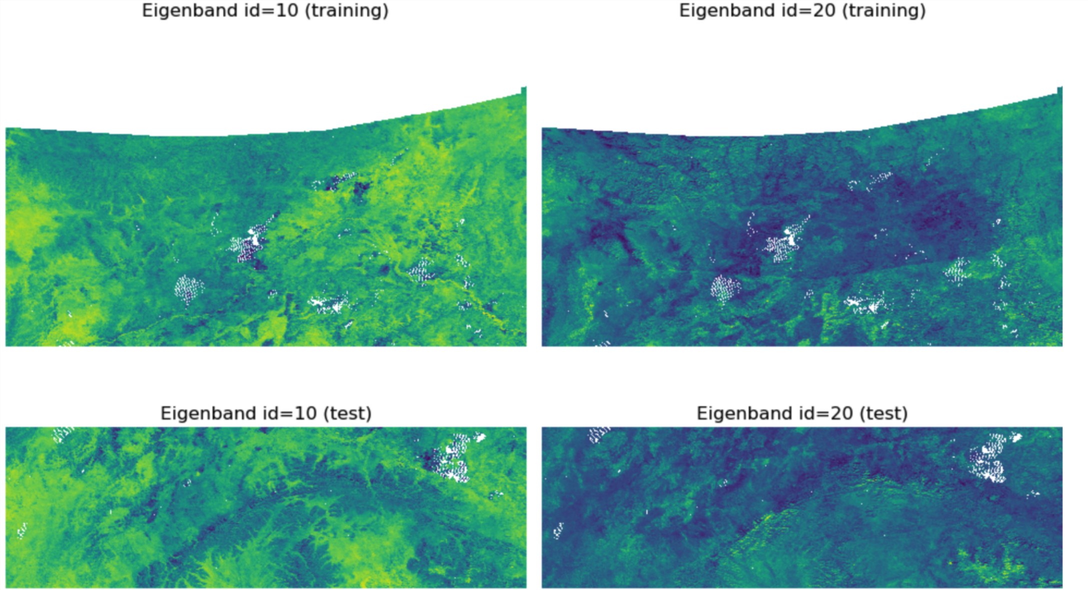
Standarization (z-score):

- centralize (subtract means)
- normalize variance

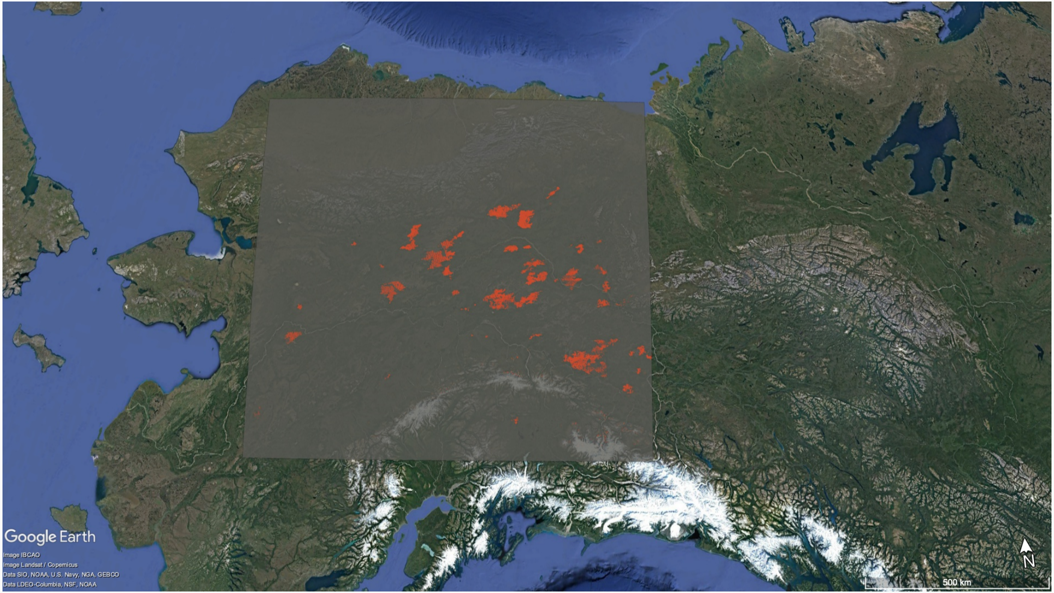
Principal Component Analysis (PCA)

- Reduce the time series dimension
- Target cumulative explained variance

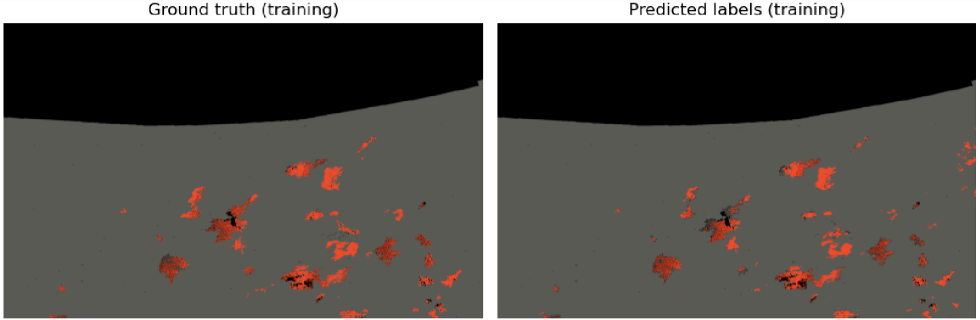
PCA Example (SWIR1)



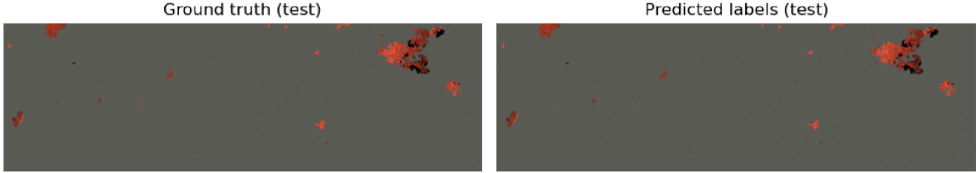
Final Map of 2004 Alaska Wildfire Season



Visual Comparison of the Best Result



(a) Comparison of ground truth and predicted wildfire localization on the training data set.



(b) Comparison of ground truth and predicted wildfire localization on the test data set.

Scores

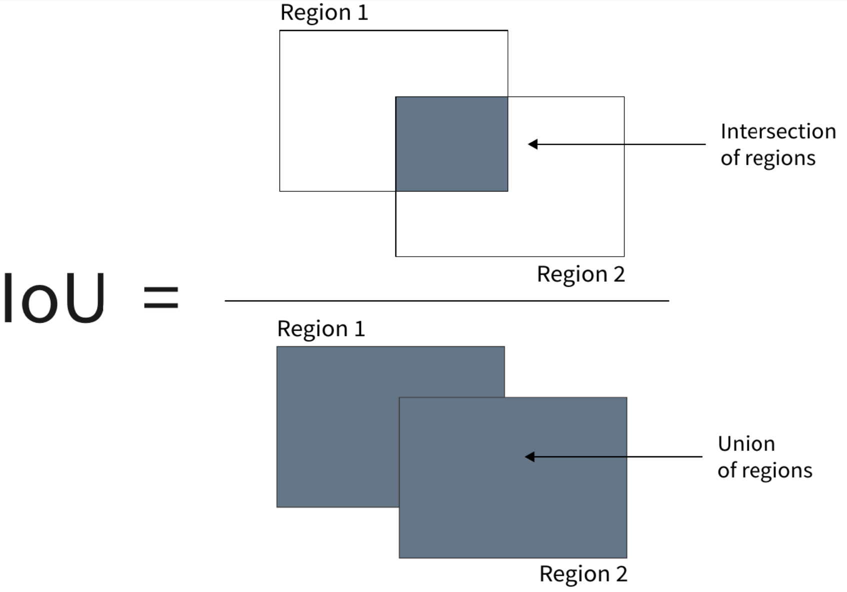
		Predicted	
		Class A (Positive)	Class B (Negative)
Actual	Class A	True Positive (TP)	False Negative (FP)
	Class B	False Positive (FN)	True Negative (TN)

$$\text{precision} = \frac{TP}{TP + FP}$$
$$\text{recall} = \frac{TP}{TP + FN}$$

Harmonic mean of precision and recall:

$$\text{F1} = \frac{2}{\text{precision}^{-1} + \text{recall}^{-1}}$$

Scores - Intersection over Union



Results L1 Loss – Computed on 16 $\times$ NVIDIA V100 (DGX2)

Transformation			Features	Model scores (test)				Train time [s]
SG	z-score	PCA		recall	precision	F1	mIoU	
			46×11	DIVERGED				
	$\times$		46×11	DIVERGED				
$\times$			46×11	DIVERGED				
$\times$	$\times$		46×11	0.87	0.55	0.56	0.49	$3,280$
	$\times$	$\times$ / 0.80	17×11	0.89	0.55	0.55	0.49	$2,978$
$\times$	$\times$	$\times$ / 0.80	5×11	0.89	0.82	0.85	0.77	$1,915$
	$\times$	$\times$ / 0.90	23×11	0.86	0.55	0.56	0.49	$3,567$
$\times$	$\times$	$\times$ / 0.90	8×11	0.90	0.66	0.72	0.63	$2,157$
	$\times$	$\times$ / 0.95	28×11	0.73	0.86	0.78	0.69	$3,950$
$\times$	$\times$	$\times$ / 0.95	11×11	0.93	0.59	0.64	0.56	$2,822$

Training time includes hyperoptimization of C over $\{10^i \mid i \in \{-3.0, -2.9, \cdots - 2.0\}\}$

Results L2 Loss – Computed on 16 $\times$ NVIDIA V100 (DGX2)

Transformation			Features	Model scores (test)				Train time [s]
SG	z-score	PCA		recall	precision	F1	mIoU	
			46×11	DIVERGED				
	$\times$		46×11	0.74	0.56	0.59	0.53	$7,003$
$\times$			46×11	DIVERGED				
$\times$	$\times$		46×11	0.87	0.55	0.56	0.49	$2,740$
	$\times$	$\times$ / 0.80	17×11	0.88	0.70	0.76	0.67	872
$\times$	$\times$	$\times$ / 0.80	5×11	0.87	0.86	0.86	0.78	481
	$\times$	$\times$ / 0.90	23×11	0.91	0.61	0.67	0.59	$1,056$
$\times$	$\times$	$\times$ / 0.90	8×11	0.93	0.69	0.76	0.67	772
	$\times$	$\times$ / 0.95	28×11	0.91	0.76	0.82	0.73	$1,267$
$\times$	$\times$	$\times$ / 0.95	11×11	0.91	0.68	0.75	0.66	$1,023$

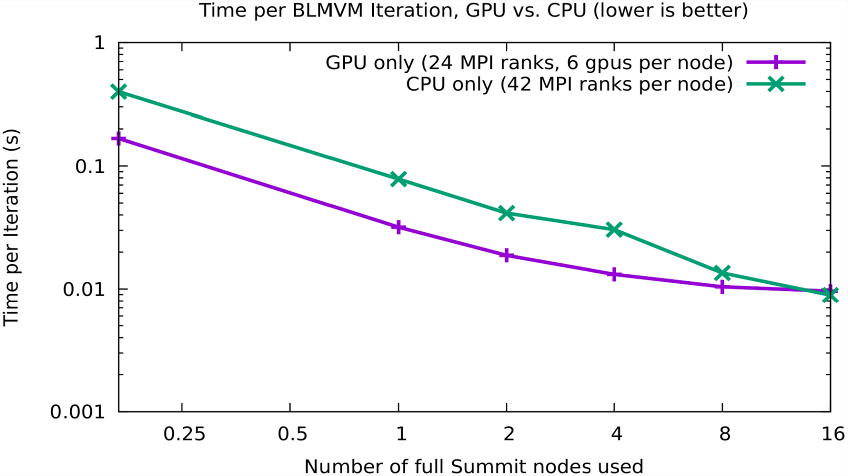
Training time includes hyperoptimization of C over $\{10^i \mid i \in \{-3.0, -2.9, \cdots - 2.0\}\}$

Comparison with scikit-learn SVM – California 2020

- Only the 7 MODIS bands
- June to October
- Randomly shuffled data split into 70 % training and 30 % testing

	PermonSVM	scikit-learn
TP	18,199	18,028
TN	519,253	519,287
FP	6,069	1,132
FN	1,515	6,589
Accuracy	0.98	0.98
Precision	0.75	0.94
Recall	0.94	0.73
F1	0.83	0.82

Strong Scalability California 2016–2020 @ Summit (Oak Ridge)



leftmost 1/6 of a node (7 CPU cores or 1 GPU)

Software

Open-source framework for wildfire localization from satellite images:

https://github.com/natural-hazards/wildfires

- Download data from Google Earth engine
- Data preprocessing
- Generation of training/testing datasets for PermonSVM
- Visualization

Thank you for your attention!

Any questions?

Jakub Kruzik, Marek Pecha, David Horak

jakub.kruzik@ugn.cas.cz

http://permon.vsb.cz