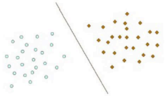
University of Chinese Academy of Sciences

Extreme weather prediction by Support Vector Machine

Statistical Analysis and Application in Climate Research

庄逸

中国科学院大气物理研究所



Nov. 2021, UCAS

Contents

University of Chinese Academy of Sciences

Extreme weather

- Extreme weather is **disastrous** and tends to occur **more frequently**.
- Heat Waves, Drought, Heavy Downpours, Floods, Hurricanes, ...
- By making **better prediction**, we can reduce its loss effectively.

Fig: Extreme weathers

中国科学院大学
University of Chinese Academy of Sciences

Method for predicting extreme weather

- There are many ways to predict extreme weather.

Numerical weather prediction (NWP)

- NWP rely on basic **physical laws** and current **weather state**.
- Generally, NWP works fine; But it fails to predict certain **extreme weather** well, e.g. heavy rainfall.
- This may results from **complicated processes** and **multiscale** property.

University of Chinese Academy of Sciences

Method for predicting extreme weather

- There are many ways to predict extreme weather.

Numerical weather prediction (NWP)

- NWP rely on basic **physical laws** and current **weather state**.
- Generally, NWP works fine; But it fails to predict certain **extreme weather** well, e.g. heavy rainfall.
- This may results from **complicated processes** and **multiscale** property.

Analog method

- Analog method is a **statistical** and **probabilistic** model.
- Based on **similarity of atmospheric conditions** on extreme days.

University of Chinese Academy of Sciences

Method for predicting extreme weather

- There are many ways to predict extreme weather.

Numerical weather prediction (NWP)

- NWP rely on basic **physical laws** and current **weather state**.
- Generally, NWP works fine; But it fails to predict certain **extreme weather** well, e.g. heavy rainfall.
- This may results from **complicated processes** and **multiscale** property.

Analog method

- Analog method is a **statistical** and **probabilistic** model.
- Based on **similarity of atmospheric conditions** on extreme days.

The key point is how to define "Similarity"?

中国科学院大学 University of Chinese Academy of Sciences

To be more specific...

- Assume we have the following knowledge[1].

Date	Temperature at noon (°C)	Weather in the afternoon
2021/8/16	33	Heavy rain
2021/8/17	35	Heavy rain
2021/8/18	28	Sunny
2021/8/19	31	Heavy rain
2021/8/20	26	Sunny

Table: Example data

[1] Fake examples, just for explanation.

University of Chinese Academy of Sciences

To be more specific...

- Assume we have the following knowledge[1].

Date	Temperature at noon (°C)	Weather in the afternoon
2021/8/16	33	Heavy rain
2021/8/17	35	Heavy rain
2021/8/18	28	Sunny
2021/8/19	31	Heavy rain
2021/8/20	26	Sunny

Table: Example data

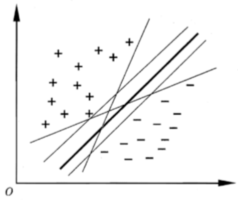
- We may conclude that an ≥ 30 °C Temp. at noon leads to heavy rain in the afternoon. And we can use this **criterion** to predict heavy rainfall in the afternoon.
- Now we have **large amount** of atmospheric data before extreme weather, how can we develop a **criterion** for prediction?

[1] Fake examples, just for explanation.

中国科学院大学
University of Chinese Academy of Sciences

What is SVM?

- Support Vector Machine(SVM), is a **binary classifier**.
- We have labelled data $D = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}, y_i = \pm 1$.
  - Vector $\boldsymbol{x}_i$ represents **atmospheric conditions**(Temp., Wind, etc.).
  - $y_i = +1, -1$ stands for **extreme** weather and **non-extreme** weather respectively.
- We seek for a hyperplane for **separation** by the sign of y_i.



- For **generalization** purpose, the "center" one is the best.

中国科学院大学 University of Chinese Academy of Sciences

How to compute?

We define **Canonical Separating Hyperplane** $\mathcal{H}$, that

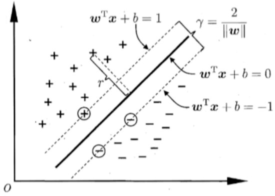
$$\mathcal{H} : \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = 0 \tag{1}$$

For $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ which are two **closet** points from each side, they satisfy

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_1 + b = 1, \qquad \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_2 + b = -1 \tag{2}$$

And the **margin width** γ can be computed as

$$\gamma = \frac{\boldsymbol{w}^{\mathrm{T}}}{\|\boldsymbol{w}\|}(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \frac{2}{\|\boldsymbol{w}\|} \tag{3}$$



中国科学院大学
University of Chinese Academy of Sciences

How to compute? The optimization problem.

- Now, as we want to **maximize** margin and the margin directly depends on $\|\boldsymbol{w}\|$, we reach the following optimization problem.

Optimization problem for solving SVM

$$\begin{aligned} &\min \frac{1}{2}\|\boldsymbol{w}\|^2 \\ &s.t. \quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 \end{aligned} \tag{4}$$

- There are many developed **optimization methods** to solve it.

University of Chinese Academy of Sciences

How to compute? The optimization problem.

- Now, as we want to **maximize** margin and the margin directly depends on $\|\boldsymbol{w}\|$, we reach the following optimization problem.

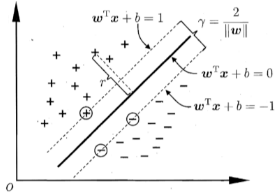
Optimization problem for solving SVM

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 \quad s.t. \quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 \tag{4}$$

- There are many developed **optimization methods** to solve it.

What is support vector?

- It is obvious that, only closet points (e.g. $\boldsymbol{x}_1, \boldsymbol{x}_2$) will affect the result.
- They are called **Support Vectors**, and that is where **Support Vector Machine** comes from.



University of Chinese Academy of Sciences

Application and Discussion

- Face recognition, text classification, OCR, bioinformatics, ...
- Based on analog methods and SVM, Nayak(2013) developed a **classifier** which predicts **extreme rainfall** in Mumbai 6-48 h ahead, according to corresponding atmosphere conditions.
- They collected extreme rainfall data of Mumbai from 1969 to 2008.
  - The **training set** contains data from 1969 to 1999.
  - The **validation set** contains data from 2000 to 2008.

中国科学院大学 University of Chinese Academy of Sciences

Application and Discussion

- Face recognition, text classification, OCR, bioinformatics, ...
- Based on analog methods and SVM, Nayak(2013) developed a **classifier** which predicts **extreme rainfall** in Mumbai 6-48 h ahead, according to corresponding atmosphere conditions.
- They collected extreme rainfall data of Mumbai from 1969 to 2008.
  - The **training set** contains data from 1969 to 1999.
  - The **validation set** contains data from 2000 to 2008.
- For better performance, **day** events and **night** events are separately trained.
- Both SVM1 and SVM2 are used for prediction.

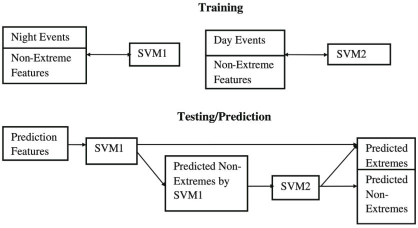


Fig. 4 Flowchart of the two-phase SVM model

University of Chinese Academy of Sciences

Application and Discussion

- Result:
  - Besides **16 correct** extreme predictions, there are **133 false alarms**. 0 miss.
  - Much better than previous fingerprinting method (**900+ false alarms**).
- Limitations:
  - Region choice: small → **exclude** important factors; large → **less weight**.
  - Lack of data: only 40 yrs and extremes are **rare**.
  - Detailed data: **high-resolution** weather pattern, **Doppler radar data**.

Table 8 Best SVM architecture

SVM1		SVM2	
Kernel function	RBF	Kernel function	Quadratic
Kernel function argument (sigma)	0.8900	Bias	0.9489
Bias	0.3999	Support vectors	45×4
Support vectors	48×32	Optimization method	SMO

University of Chinese Academy of Sciences

Application and Discussion

- An advantage of SVM is that we know **how predictor works**.
  - E.g. if we find $\boldsymbol{w} = (w^{(1)}, \ldots, w^{(m)}, \ldots, w^{(n)})$ have $w^{(m)} \approx 0$, then it indicates the corresponding variable $x_i^{(m)}$ may not be important. (Why?)
  - The article does not provide it though, which may results from **kernel function** and other difficulties.

中国科学院大学
University of Chinese Academy of Sciences

Application and Discussion

- An advantage of SVM is that we know **how predictor works**.
  - E.g. if we find $\boldsymbol{w} = (w^{(1)}, \ldots, w^{(m)}, \ldots, w^{(n)})$ have $w^{(m)} \approx 0$, then it indicates the corresponding variable $x_i^{(m)}$ may not be important. (Why?)
  - The article does not provide it though, which may results from **kernel function** and other difficulties.

- SVM disadvantages:
  - Cost **great computational effort** for large amount of training data.
  - The selection of kernel function, parameters, etc. is **subjective**.

中国科学院大学
University of Chinese Academy of Sciences

Application and Discussion

- An advantage of SVM is that we know **how predictor works**.
  - E.g. if we find $\boldsymbol{w} = (w^{(1)}, \dots, w^{(m)}, \dots, w^{(n)})$ have $w^{(m)} \approx 0$, then it indicates the corresponding variable $x_i^{(m)}$ may not be important. (Why?)
  - The article does not provide it though, which may results from **kernel function** and other difficulties.

- SVM disadvantages:
  - Cost **great computational effort** for large amount of training data.
  - The selection of kernel function, parameters, etc. is **subjective**.

- Open questions:
  - Is it reliable in the future? How can we take **climate change** into account?
  - Should **other factors** be included, like forest area, pollution level, etc.?
  - Can we turn binary classification into **continous** one, which provides rainfall **probability** and **strength** information?
  - How to adapt the method for **other extreme weather** prediction?

中国科学院大学
University of Chinese Academy of Sciences

Take Home Message

- Support Vector Machine(SVM) is a **binary classifier** and is trained by solving an **optimization** problem.

- Analog method predicts extreme weather by recognizing **similar weather pattern** ahead.

- After training with historical data, SVM is able to predict extreme weather.

University of Chinese Academy of Sciences

Tools for SVM

- LIBSVM
  http://www.csie.ntu.edu.tw/~cjlin/libsvm/
- LIBLINEAR
  http://www.csie.ntu.edu.tw/~cjlin/liblinear/
- SVM-light, SVM-perf, SVM-struct
  http://svmlight.joachims.org/svm_struct.html
- Pegasos
  http://www.cs.huji.ac.il/~shais/code/index.html

中国科学院大学
University of Chinese Academy of Sciences

Reference I

[1] NAYAK M A, GHOSH S. Prediction of Extreme Rainfall Event Using Weather Pattern Recognition and Support Vector Machine Classifier[J/OL]. Theoretical and Applied Climatology, 2013, 114(3): 583-603(2013-11-01). https://doi.org/10.1007/s00704-013-0867-3. DOI: 10.1007/s00704-013-0867-3.

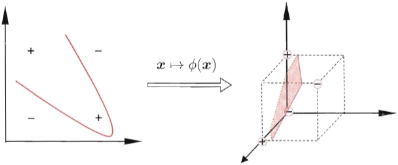
[2] 周志华. 机器学习[M]. 第 1 版. 北京: 清华大学出版社, 2016.

Many thanks to lecture slides from Prof. Lan Yanyan (2019).

THANKS!

Pratical problems and Extensions: Kernel Function

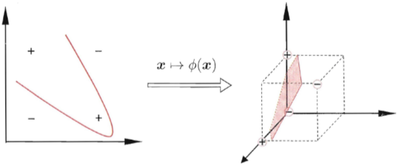
- What if... the data is not **linearly separable**?



University of Chinese Academy of Sciences

Pratical problems and Extensions: Kernel Function

- What if... the data is not **linearly separable**?



- We can introduce a **function**, which maps data into the **feature space**, where they are separable.

In practice, we only need to deal with $\phi(\boldsymbol{x}_i)^{\mathrm{T}}\phi(\boldsymbol{x}_j)$, and we simply define

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^{\mathrm{T}}\phi(\boldsymbol{x}_j) \tag{5}$$

Where K is called **Kernel Function**.

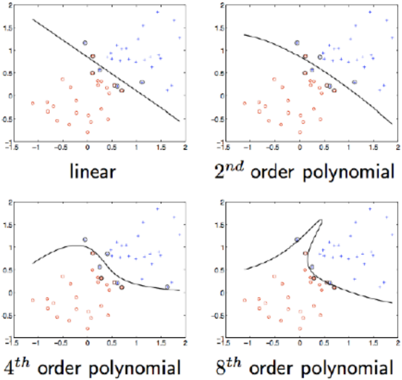
University of Chinese Academy of Sciences

Pratical problems and Extensions: Kernel Function

The choice of K requires experience and attempts.

Type	Formula
Linear	$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$
Polynomial	$(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)^q$
Gaussian	$\exp(-\lVert\boldsymbol{x}_i-\boldsymbol{x}_j\rVert^2/2\sigma^2)$
Laplace	$\exp(-\lVert\boldsymbol{x}_i-\boldsymbol{x}_j\rVert/\sigma)$
Sigmoid	$\tanh(\beta\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+\theta)$

Table: Common Kernel Functions

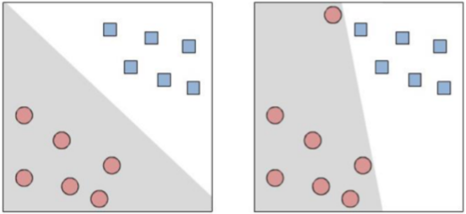


From Tommi Jaakkola, MIT CSAIL

University of Chinese Academy of Sciences

Pratical problems and Extensions: Soft margin

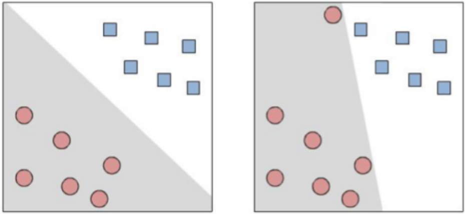
- What if... there is noise or **outliers** in the data?



中国科学院大学 University of Chinese Academy of Sciences

Pratical problems and Extensions: Soft margin

- What if... there is noise or **outliers** in the data?



- For **generalization** purpose, we may want a separation that is not so **strict**.
- So we can relax the constraint a little.

$$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 \quad \rightarrow \quad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - \xi_i \tag{6}$$

- Where $\xi_i > 0$ represents the **error**.

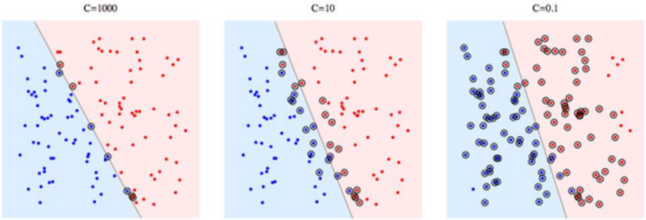
University of Chinese Academy of Sciences

Pratical problems and Extensions: Soft margin

- On the other hand, we don't want the error to be **too large**, thus the goal is reformulated as

$$\min \frac{1}{2}\|w\|^2 \quad \rightarrow \quad \min \left(\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i \right) \tag{7}$$

- Where parameter C measures the tradeoff between **margin maximization** and **training error minimization**.
- Now we can solve the new **optimization problem**.



University of Chinese Academy of Sciences

Backup: AFM method

- Anomaly frequency method(AFM) is an efficient technique in extracting the **features** which discriminate extreme events and non-extreme events.
- For a variable, those grid points are selected as feature grid points which have a very **high frequency** of extreme anomalies.

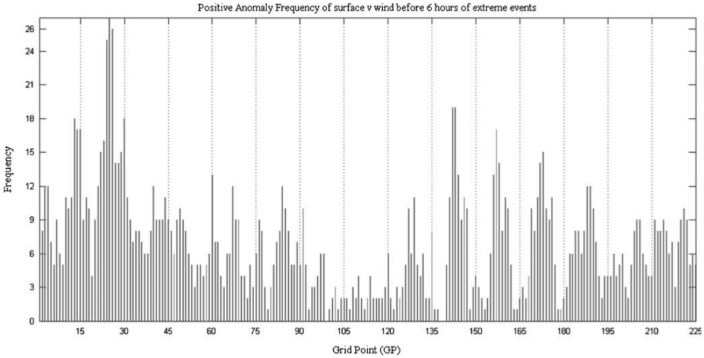


Fig. 3 Frequency of high positive anomaly of V-wind velocity at the surface level, at different grid points, 6 h before the extreme events. Fifty extreme events are considered for this

中国科学院大学
University of Chinese Academy of Sciences

Backup: Fingerprinting approach drawbacks

1. The fingerprints identified by the approach may also be present on a **non-extreme** day, which may result in false alarms.
2. There may be **multiple numbers of weather patterns**, which may result in extreme events; however, the fingerprinting approach considers only one fingerprint.