Tornado Detection using an Incremental Revised Support Vector Machine with Filters

DDDAS 06

Hyung-Jin Son & Theodore B. Trafalis

June 29, 2006

Organization

- Support Vector Machines (SVMs)
- Incremental SVM
- Revised SVM
- Incremental Revised SVM

Objective of Research

- Modify the standard SVM to reduce # of support vectors
- Construct incremental learning procedure
- On-line setting possible

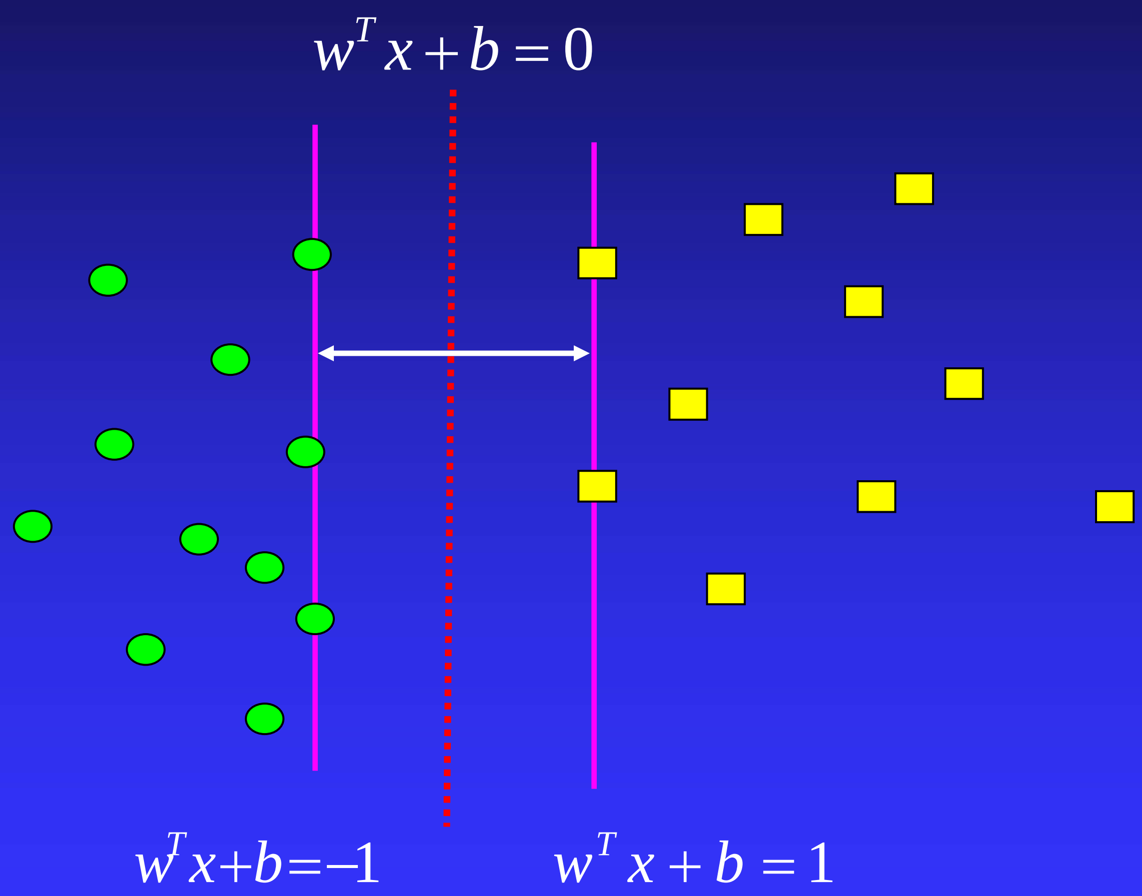
Support Vector Machines (SVMs)

- Introduced by Vapnik in 1995

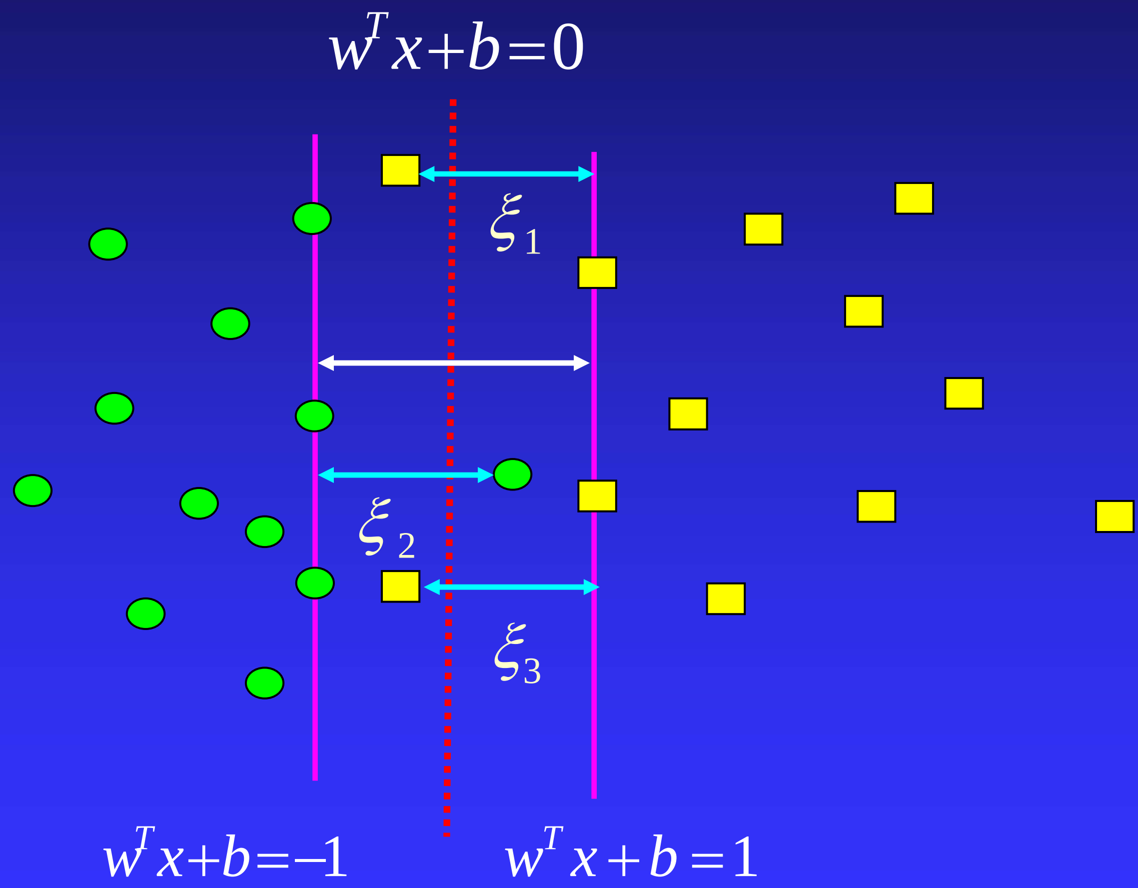
- Basic idea:

  Construct a decision hyperplane to separate samples, maximizing the margin of separation

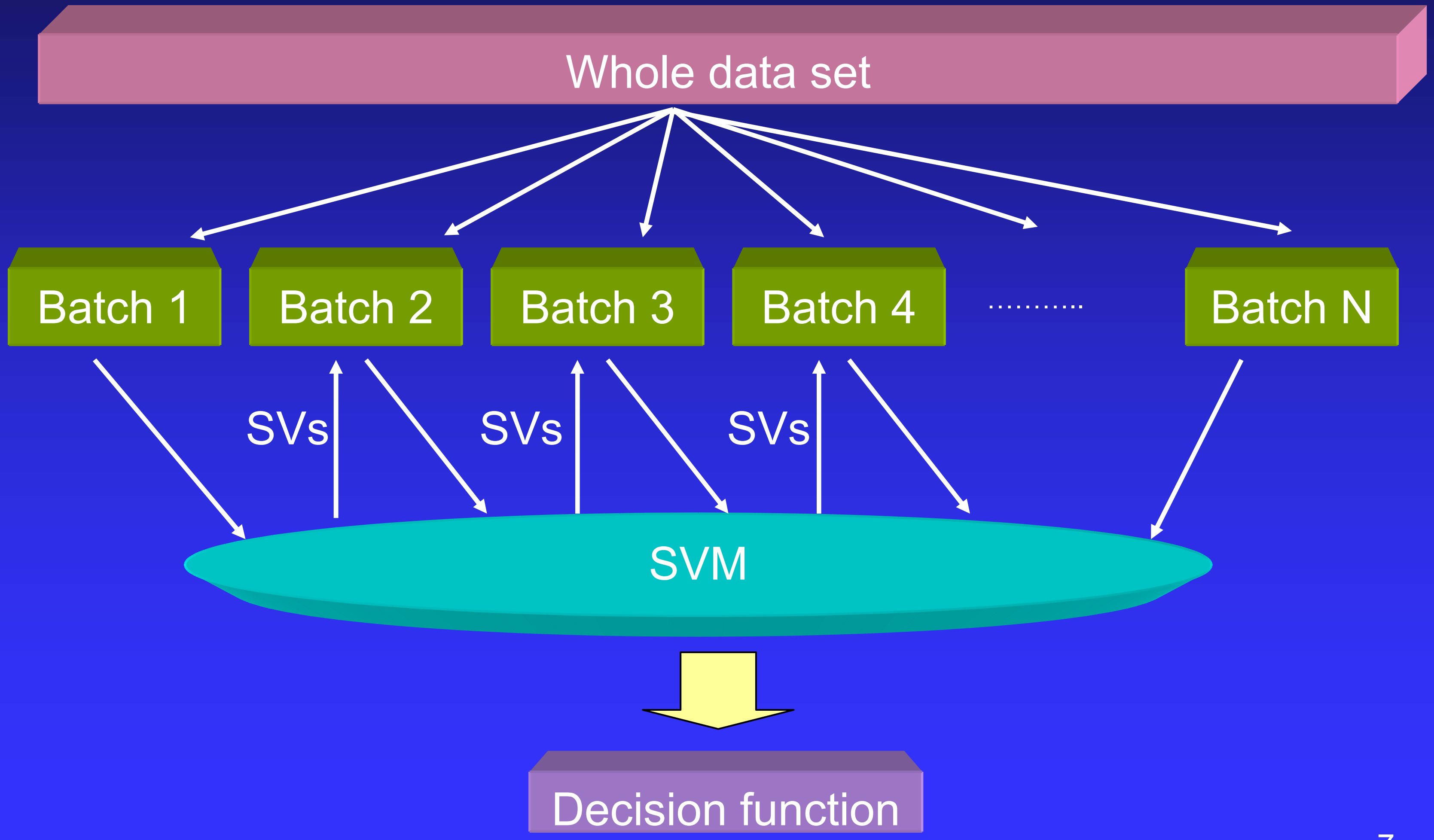
SVMs (separable)



SVMs (inseparable)



Incremental SVM

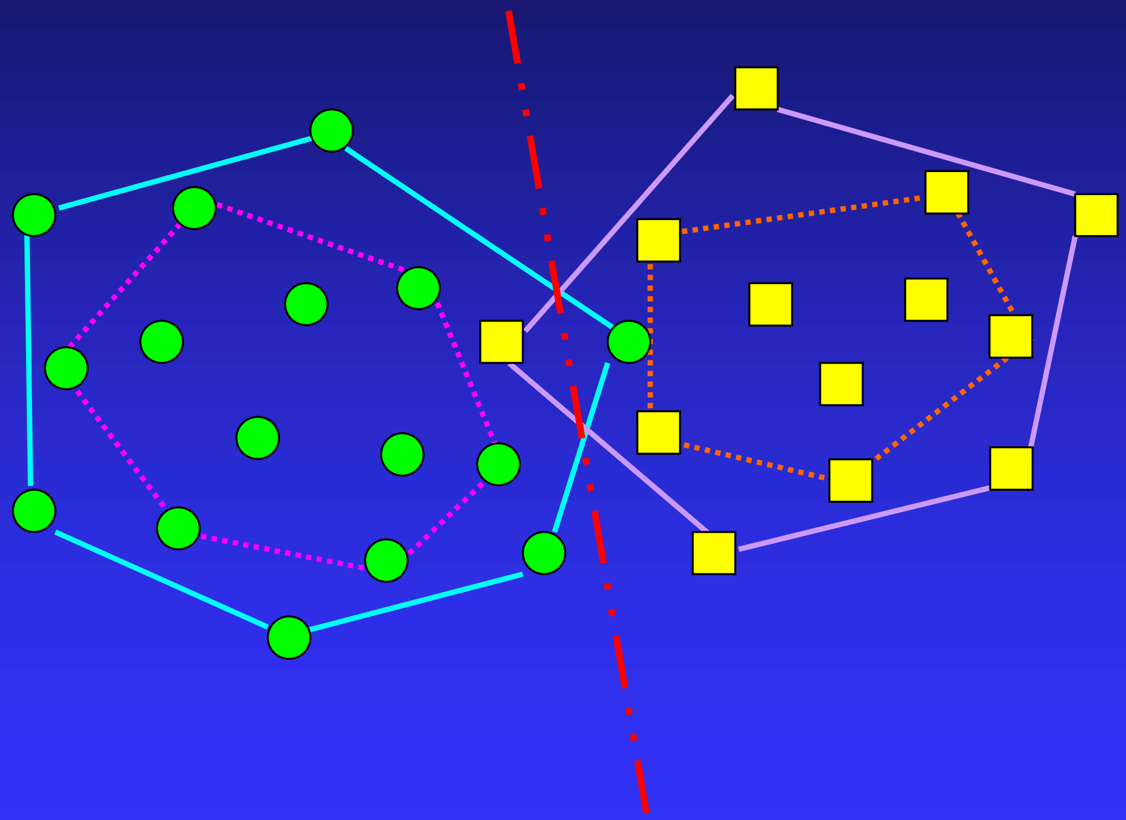


Incremental SVM

- Advantages
  - Can be used with huge training data sets
  - Data available at periodic intervals
- Disadvantages
  - support vectors might be accumulated in worst case
  - Inefficient for unbalanced problems

Reduced SVM

$w^T x + b = 1$



- Bennett (2000)

Motives

- Incremental learning approach
- \# of support vectors is accumulated

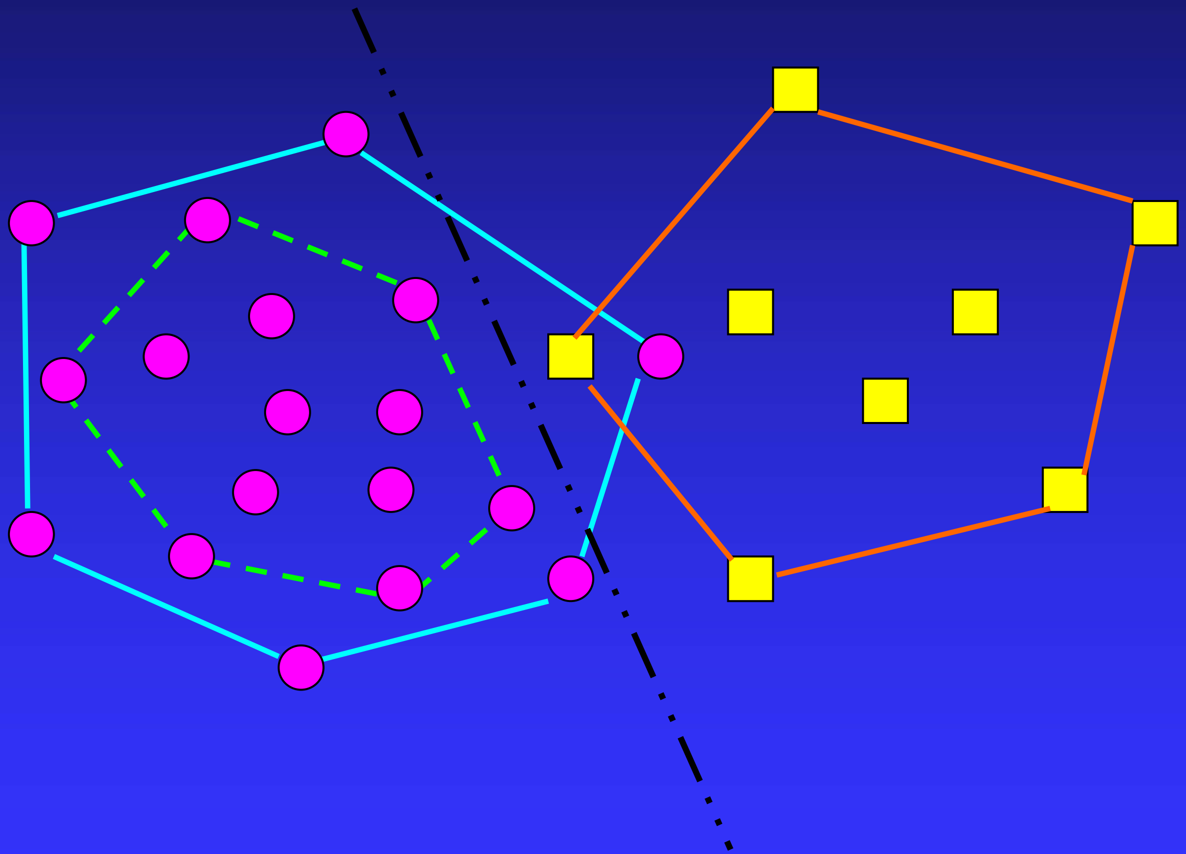
- Reduced SVM
- Classification error ↑

- Unbalanced problem
- Important (few) and unimportant classes

Revised SVM

- To reduce # of support vectors
- To reduce effect of noisy data
- To solve unbalanced, asymmetric problems

Revised SVM



Incremental Revised SVM

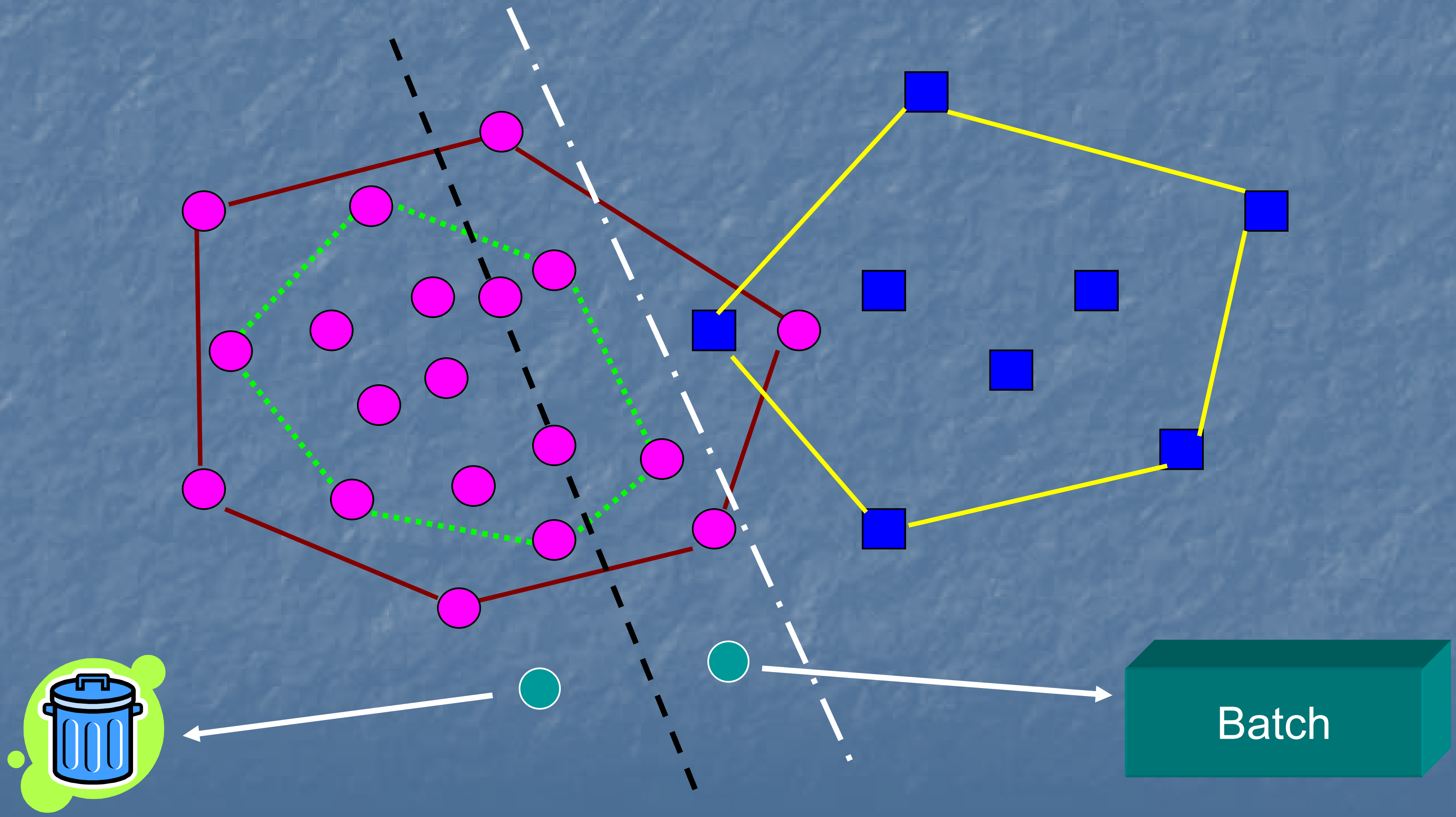
- Incremental learning procedure
- Standard SVM

  → Revised SVM
- For dynamic data driven application system (DDDAS), filter concept is introduced

Filtering

- To speed up the learning process
- To discard potential unimportant data
- Supporting hyperplane is used as a filter

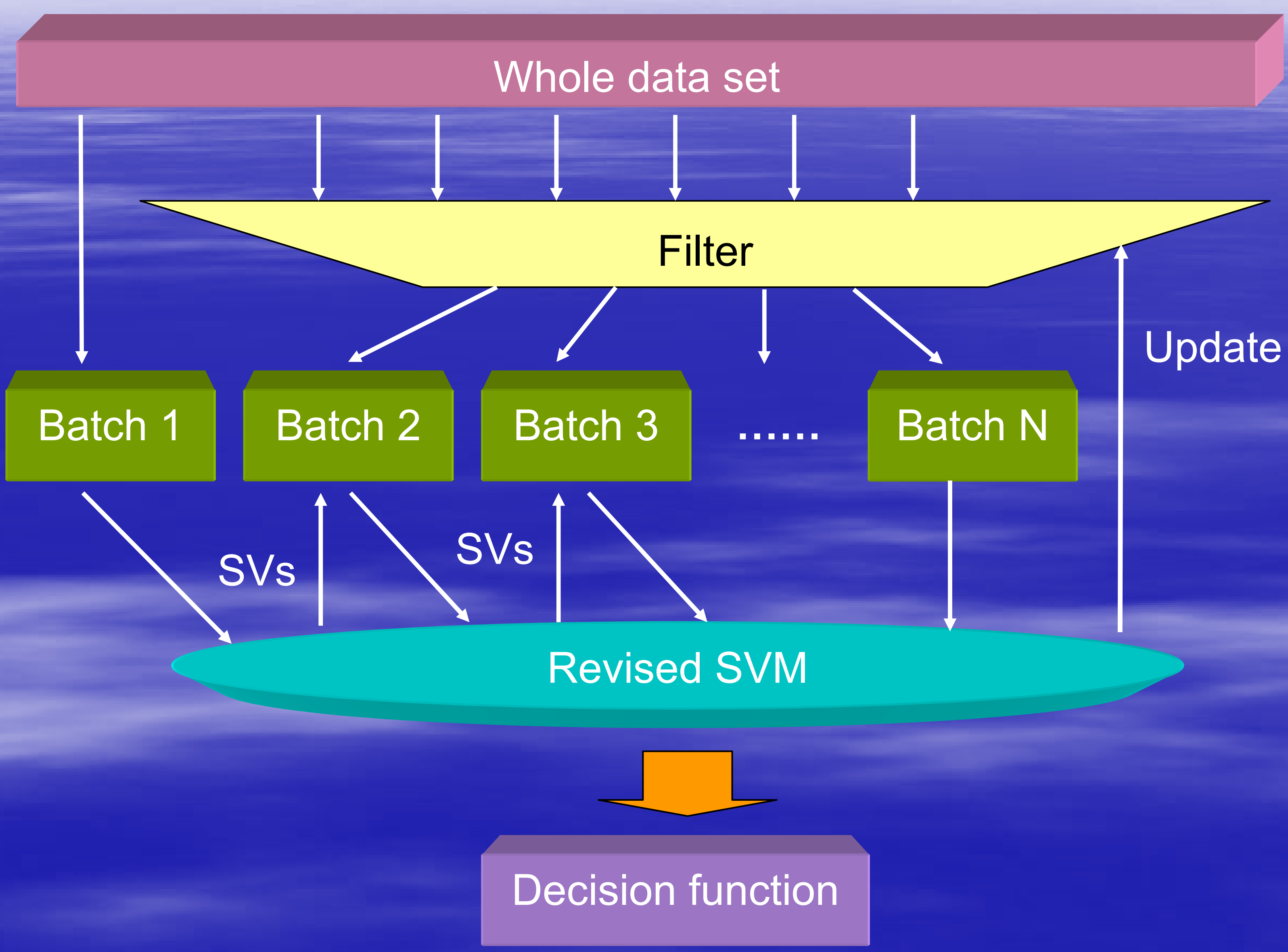
Filtering



Revised SVM with Filters

- Reduce the support vectors
- Appropriate for DDDAS
- Decrease “miss” rate (false negative)

Revised SVM with Filters



Tornado Detection

- Two-class classification problem
- Tornado & Non-tornado

- Unbalanced problem
- Tornado (few) & Non-tornado (many)

- Asymmetric problem
- Tornado (important) & non-tornado(none)

Tornado Detection (confusion matrix)

		Tornado Observed		
Forecast Torado		Yes	No	
	Yes	Hit (a)	False alarm (b)	“Yes” Forecasts
	No	Miss (c)	Correct (d)	“No” Forecasts
		“Yes” Observation	“No” Observation	Total # of observations

Probability of Detection (POD) = $\frac{a}{a + c}$

Tornado detection

- WSR-88D radar
- Mesocyclone Detection Algorithm (MDA) data - 23 attributes

- 10 training sets (1500 data/set)
- 10 testing sets (1500 data/set)

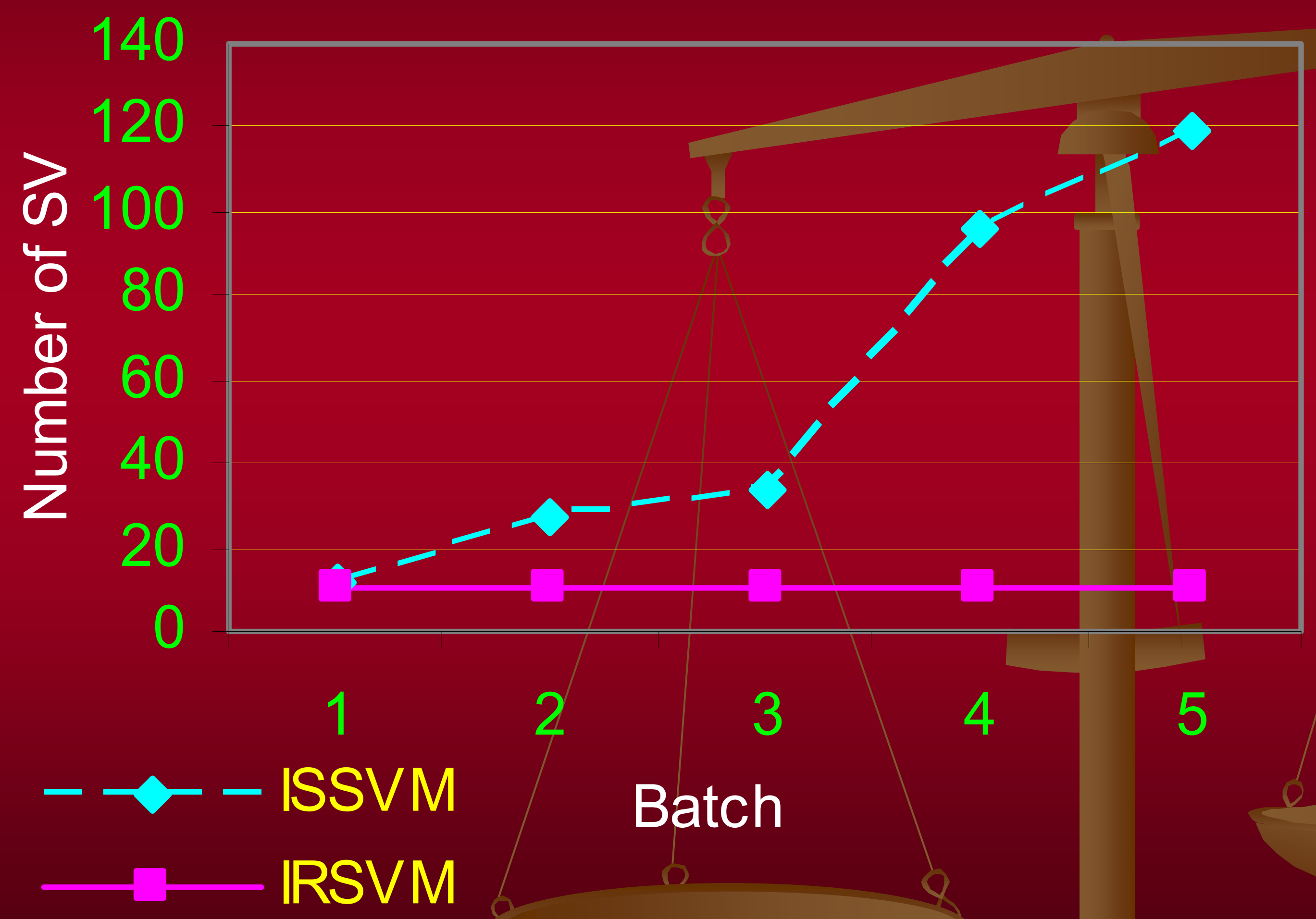
Tornado data (samples)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	2 0	21	22	23	c
10299	1	102	443	0	3335	3335	102	18	18	102	10	41	358	20	20	102	102	443	0	3772	3	5	13	0	-1
10299	1	143	520	0	4961	4961	143	13	15	407	5	19	407	6	20	407	143	520	0	3653	3	6	10	0	-1
10299	1	129	539	0	686	2233	434	14	14	129	40	40	129	2	11	434	129	539	0	3351	3	6	10	0	-1
10299	1	3388	6041	1	3008	5939	6451	24	24	3388	16	16	3388	47	47	3388	3388	6041	0	3610	4	36	0	44	-1
10299	1	121	531	0	2609	4105	432	15	15	121	12	12	121	20	20	121	121	531	0	3799	3	6	6	0	-1
10299	1	4655	10346	3	3636	8659	8166	8	10	8166	5	5	4655	16	16	4655	4655	5165	12	1541	2	58	0	0	1
10299	1	4429	10333	2	7265	7265	4429	8	10	7940	2	3	7940	12	15	7940	4429	5334	18	1190	1	36	0	0	1
10299	1	4330	6782	3	6908	10433	7626	10	10	4330	3	3	4330	14	14	4330	4330	3486	23	1418	1	38	0	0	1
10299	1	3995	6352	2	3253	9453	7033	8	9	7033	5	5	3995	10	10	7033	3995	6352	0	1163	1	33	0	9	1
10299	1	3794	9269	6	6265	9339	6829	15	15	3794	5	5	3794	23	23	3794	3794	3204	6	2903	3	62	0	13	1

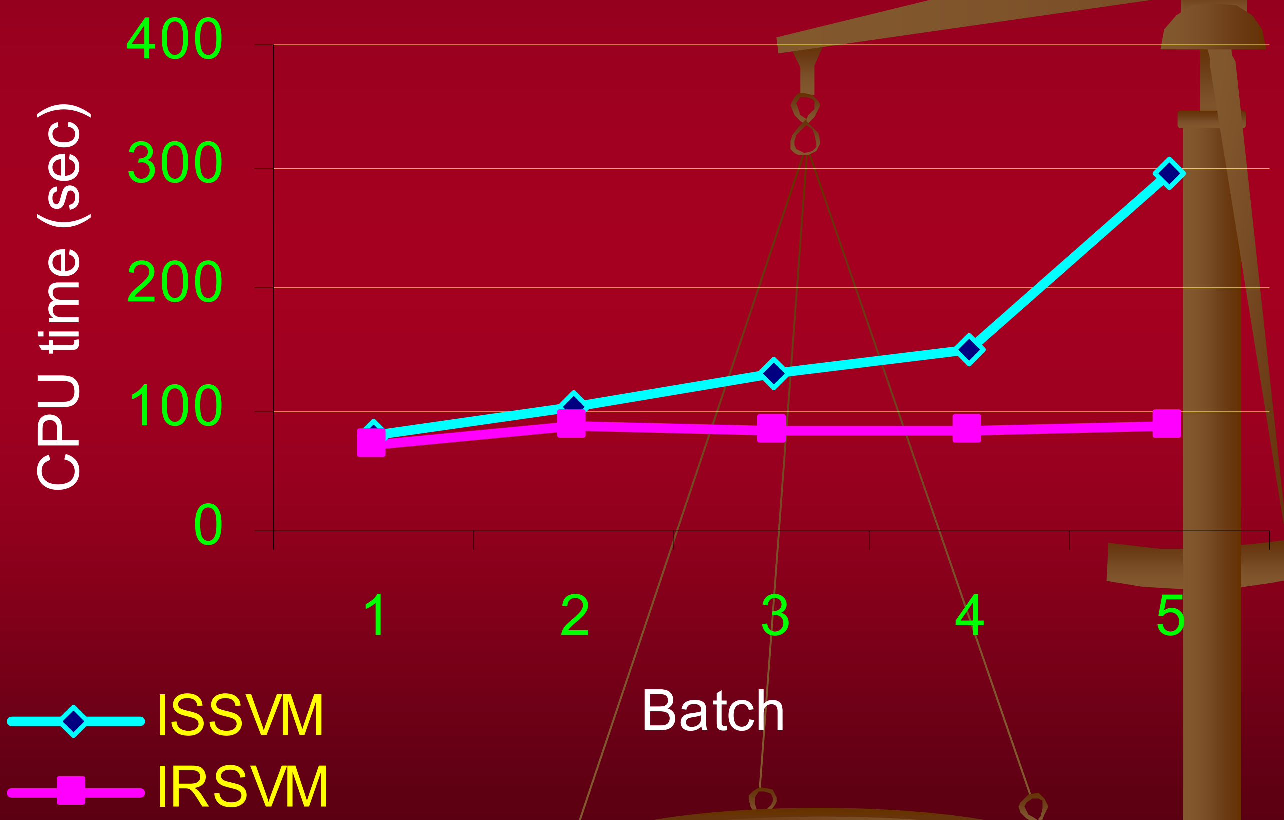
Tornado Detection (results)

Methods	POD	CPU time (Sec)	# of SV	"Miss" /all (%)
Incremental approach with standard SVM	0.62	754.62	57	3.83
Incremental approach with revised SVM	0.69	406.26	11	3.14
Incremental approach with revised SVM & filter	0.60	314.46	11	3.97

Tornado Detection (results)



Tornado Detection (results)



Future Work

- Parameters (e.g., reducing rate) should be tuned

- Algorithms will be tested in a real operating setting (data stream)

Questions / Comments

Supplementary slides

SVM (separable)

$$Min \quad \Phi(w) = \frac{1}{2} w^T w$$

subject to

$$y_i [w^T x_i + b] \geq +1$$

$$y_i = \pm 1$$

SVM (separable-dual)

$$Max\ F(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$

$$subject\ \ to$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \le \alpha_i \qquad for \quad i = 1, 2, \cdots, N$$

Back to presentation

SVM (inseparable)

$$Min \quad \Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

subject to

$$y_i [w^T x_i + b \ ] \geq 1 - \xi_i$$

$$y_i = \pm 1$$

SVM (inseparable-dual)

$$Max\ F(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$0 \le \alpha_i \le C \qquad for \quad i = 1, 2, \cdots, N$$

Back to presentation

Determine the optimal batch size

- In the sense of a tradeoff between accuracy (classification error) and computing time (CPU time)

Back to presentation

Next

Determine the optimal batch size

Back to presentation

Reduce SVM

$$\min_{a,b} \quad \frac{1}{2}\left\| \sum_{i=1}^{m} a_i u_i - \sum_{j=1}^{n} b_j v_j \right\|^2$$

s. t.

$$\sum_{i=1}^{m} a_i = 1$$

$$\sum_{j=1}^{n} b_j = 1$$

$$0 \le a_i \le \mu \qquad \text{for } i = 1, 2, \cdots, m$$

$$0 \le b_j \le \mu \qquad \text{for } j = 1, 2, \cdots, n$$

Back to presentation

Advantages of Reduced SVM

- Number of Support Vectors is reduced
- Effect of noisy data is reduced

Back to presentation

Definitions of problems

- Unbalanced problem:

  When there are two data sets, the sizes of two data sets are different.

- Asymmetric problem :

  When there are two data sets, the importance of data in one data set is different from the one of data in the other set.

Back to presentation

Generalization

- Def.

  The input-output mapping computed by the learning machine is correct for test data that are never used in training the input data

Back to presentation

Revised SVM

$$Max\ Q(\alpha,\ \beta) = \sum_{i=1}^{N_P}\sum_{j=1}^{N_N}\alpha_i\beta_j K(x_i, x_j) - \frac{1}{2}[\sum_{i=1}^{N_P}\sum_{j=1}^{N_P}\alpha_i\alpha_j K(x_i, x_j) + \sum_{i=1}^{N_N}\sum_{j=1}^{N_N}\beta_i\beta_j K(x_i, x_j)]$$

s.t.

$$\sum_{i=1}^{N_P}\alpha_i = 1$$

$$\sum_{j=1}^{N_N}\beta_j = 1$$

$0 \le \alpha_i \le 1$ for $i = 1, 2, \cdots, N_P$

$0 \le \beta_j \le \mu < 1$ for $j = 1, 2, \cdots, N_N$

Back to presentation

Definition of “Miss”

- A test incorrectly reports that a result is not detected when it is actually present
- Type II error; False negative

Back to presentation

Tornado data attributes

No.	Attributes	No.	Attributes
1	Base (m)	13	Lowe-level gate-to-gate velocity difference (m/s)
2	Depth (m)	14	Maximum gate-to-gate velocity difference (m/s)
3	Strength rank	15	Height of maximum gate-to-gate velocity difference (m)
4	Low-level diameter (m)	16	Core base (m)
5	Maximum diameter (m)	17	Core depth (m)
6	Height of maximum diameter (m)	18	Age (min)
7	Low-level rotational velocity (m/s)	19	Strength index (MSI) weighted by average density of integrated layer
8	Maximum rotational velocity (m/s)	20	Strength index (MSIr) “rank”
9	Height of maximum rotational velocity (m/s)	21	Relative depth (%)
10	Lowe-level shear (m/s/km)	22	Low-level convergence (m/s)
11	Maximum shear (m/s/km)	23	Mid-level convergence (m/s)
12	Height of maximum shear (m)		

Back to presentation

Conditions of experiments

- Polynomial kernel with power of 2
- Reducing rate = 0.1

Back to presentation

Tornado Detection (analysis)

α =0.05

Source of variation	SS	df	MS	F	P-value	F_crit
Between Groups	0.059367	2	0.029683	0.378024	0.688784	3.354131
Within Groups	2.120099	27	0.078522			
Total	2.179466	29			POD	

Back to presentation

Next

Tornado Detection (analysis)

- No significant difference of POD among methods
- Computing time is reduced

Back to presentation

Number of SVs comparison

	Tornado data	
Batch	ISSVM	IRSVM
1	12	11
2	27	11
3	34	11
4	96	11
5	119	11

Back to presentation

CPU time comparison

	Tornado data	
Batch	ISSVM	IRSVM
1	79.36	70.25
2	102.72	84.98
3	128.67	81.23
4	150.29	82.95
5	293.58	86.84

Back to presentation