EARLY EVENT CLASSIFICATION IN SPATIO-TEMPORAL DATA STREAMS

Sevvandi Kandanaarachchi
Rob J. Hyndman
Kate Smith-Miles

ISF 2019

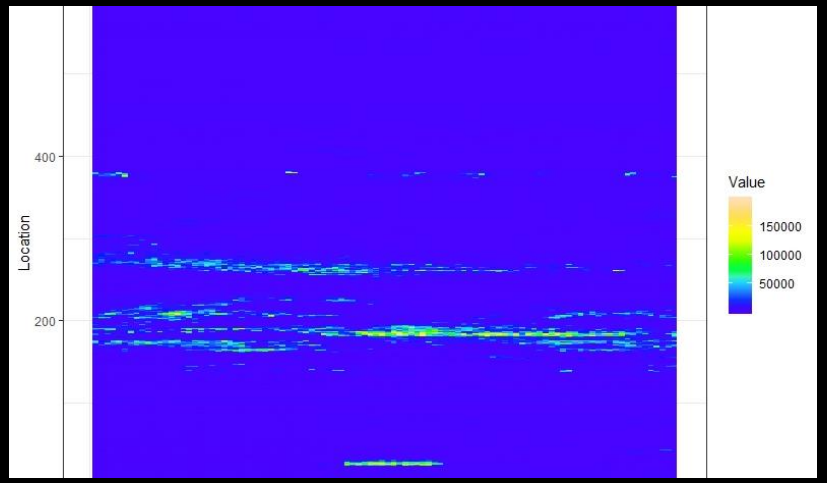
PREDICTING USING PARTIAL INFORMATION

- Information gets available gradually
- Do NOT have full information at once
- BUT required to make a prediction using partial information
- Early detection is important
- A growing prediction

PREDICTING USING PARTIAL INFORMATION

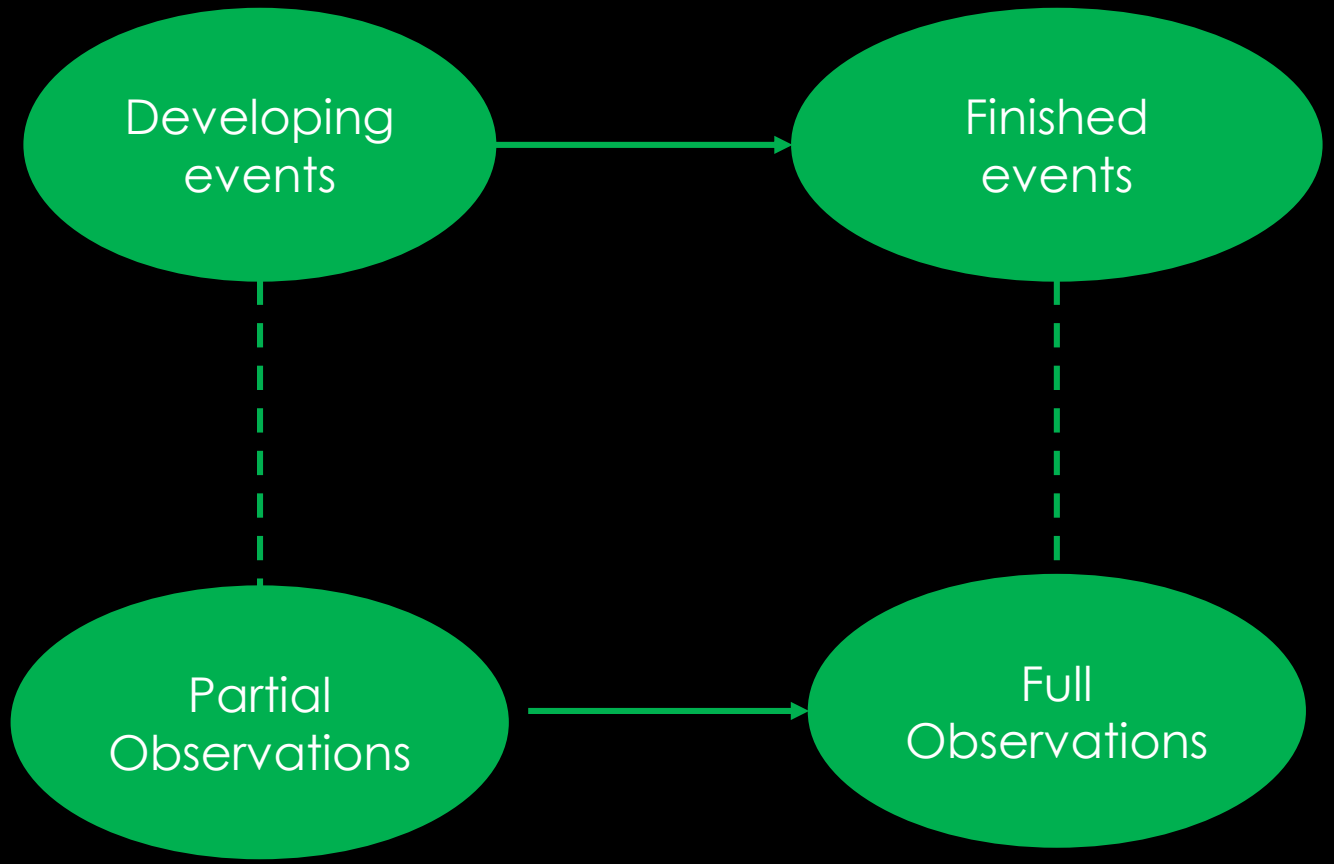
- Information gets available gradually
- Do NOT have full information at once
- BUT required to make a prediction using partial information
- Early detection is important
- A growing prediction

EXAMPLES OF PARTIAL INFORMATION

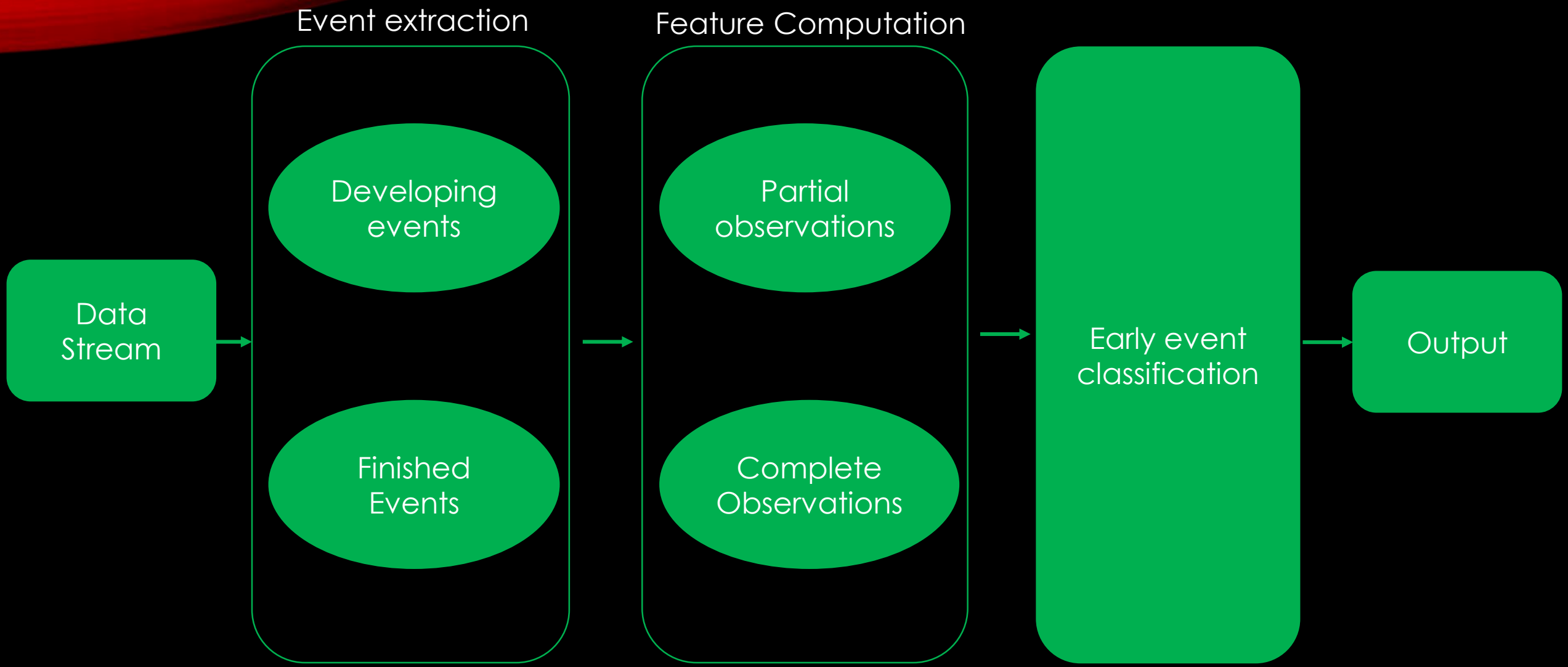


- Tulip bud vs daffodil bud
- Epidemiology
- Application in data streams
  - Network intrusion attack

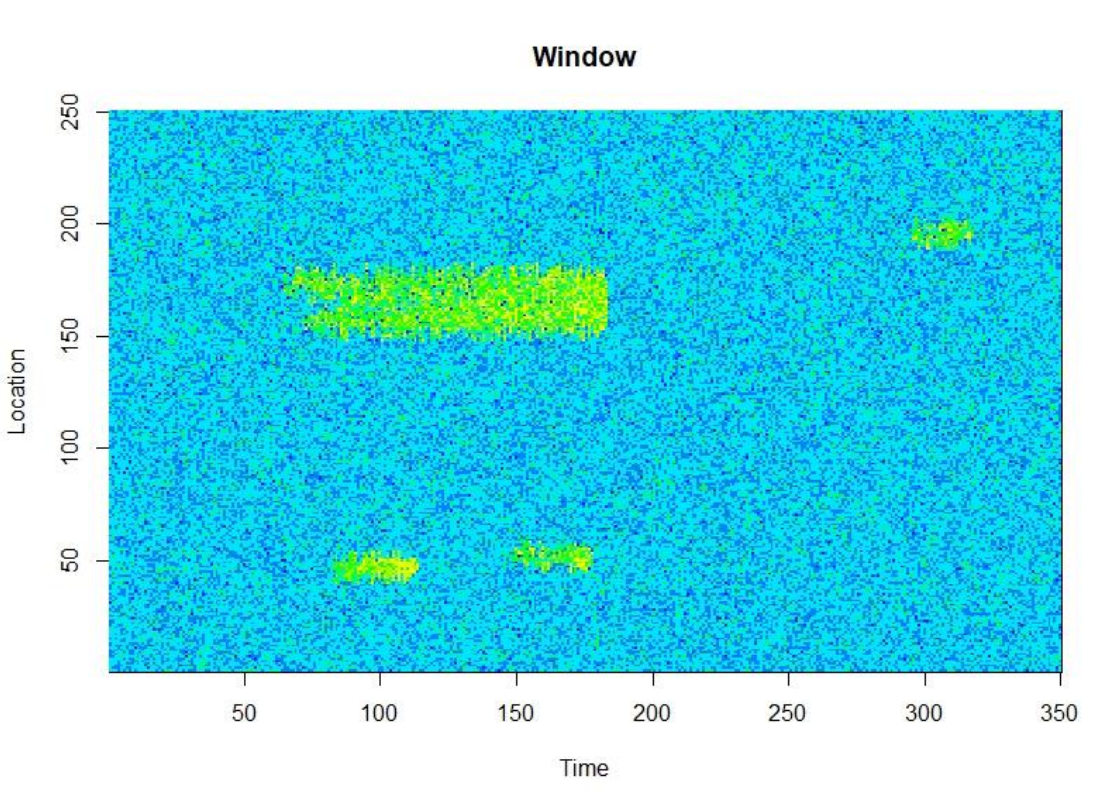
CHARACTERISTICS

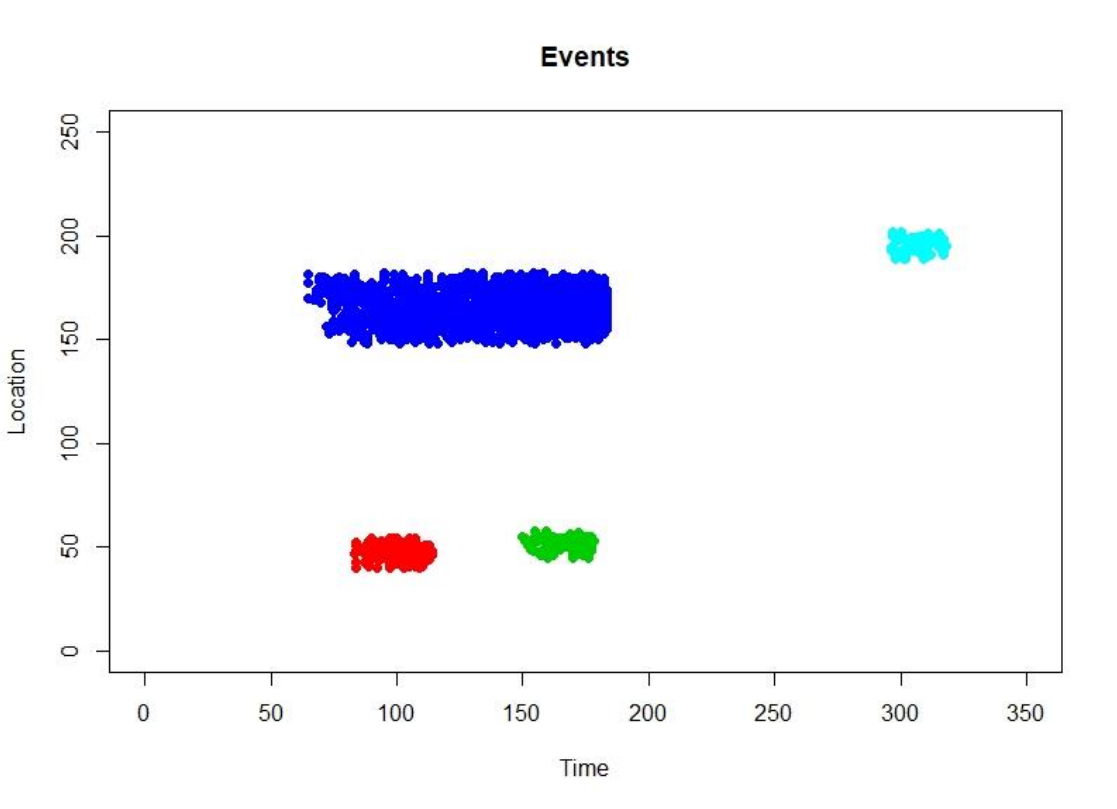


FRAMEWORK


Event extraction
Feature Computation
Data Stream
Developing events
Finished Events
Partial observations
Complete Observations
Early event classification
Output

EVENT EXTRACTION



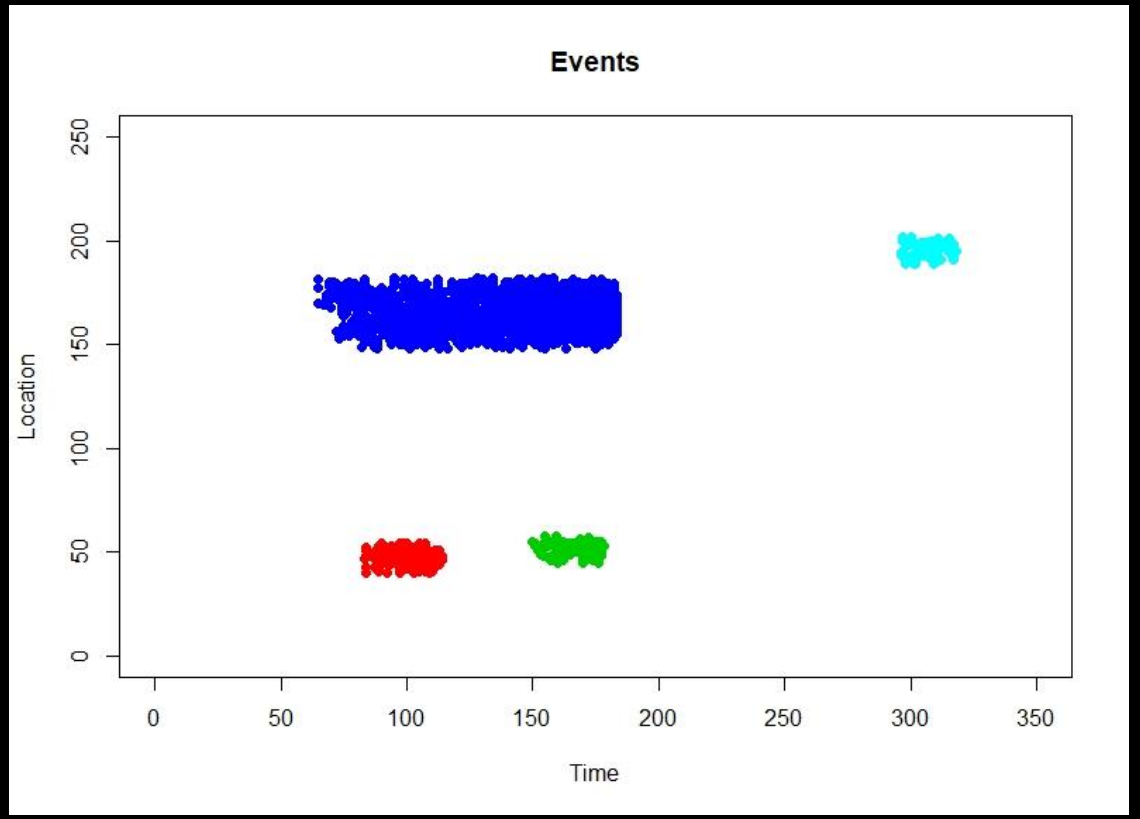


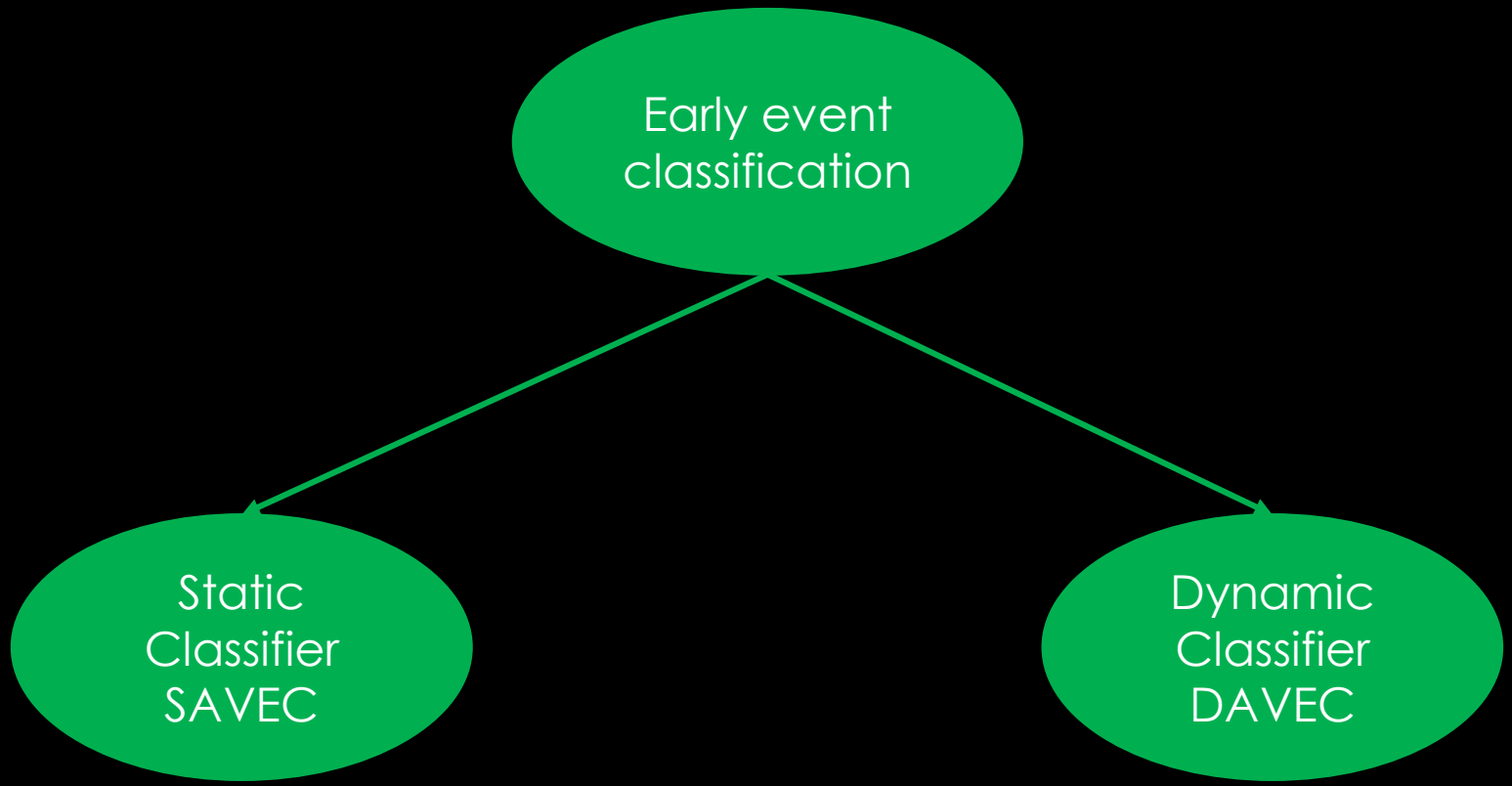
EVENT EXTRACTION ALGORITHM

- Find points greater than 95^{th} percentile
- Cluster these points using DBSCAN algorithm
- Call each cluster an event

FEATURE COMPUTATION

- Length, width of event
- Total pixel value
- Mean
- Standard deviation
- Fit a line to average pixel value for each time
- Slope of fitted line




Early event classification
Static Classifier SAVEC
Dynamic Classifier DAVEC

WHY TWO PARTIAL OBSERVATION CLASSIFIERS

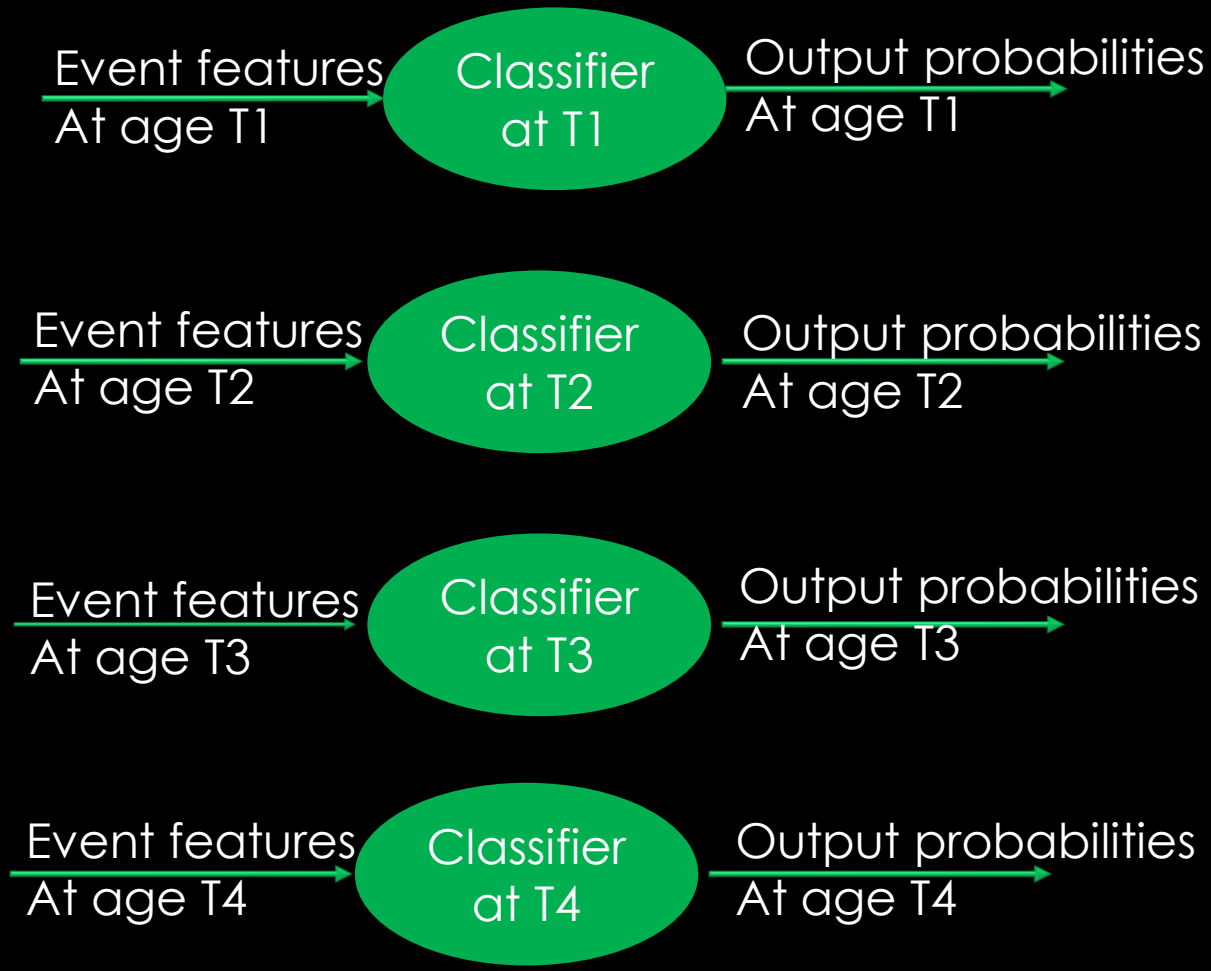
Static classifier

- Suitable for stable conditions

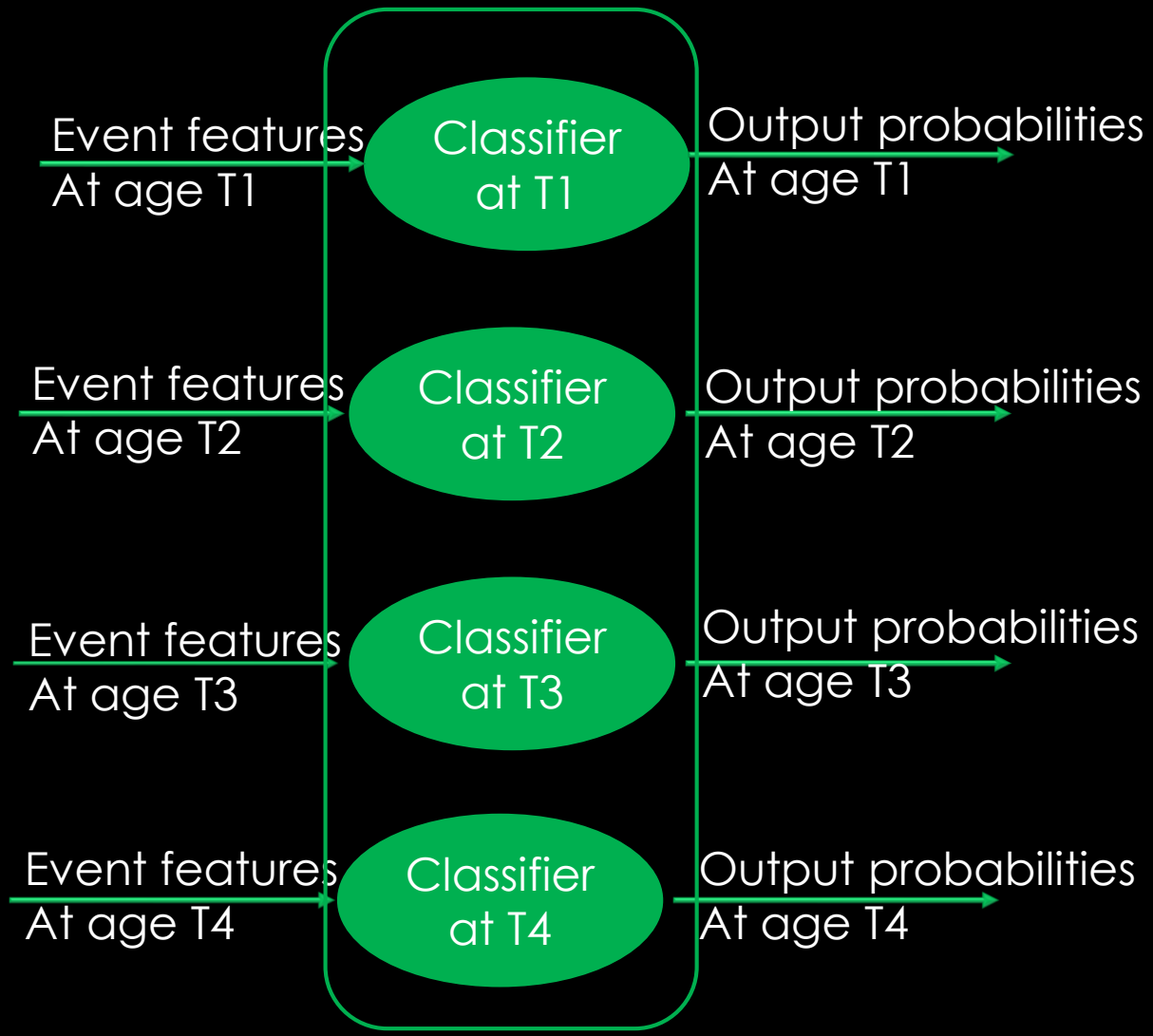
Dynamic Classifier

- When situations change fast

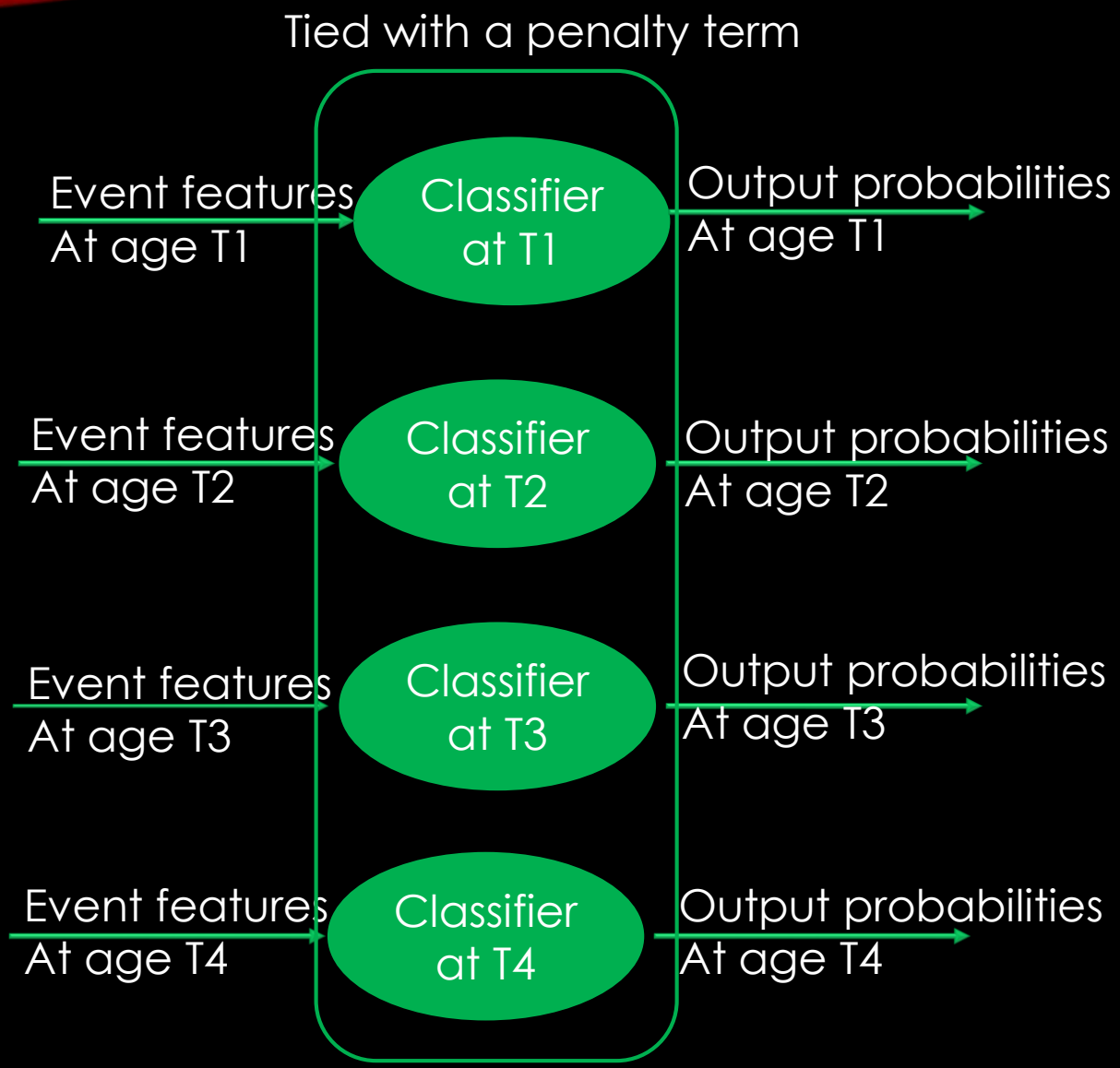
SAVEC


Event features
At age T1
Classifier
at T1
Output probabilities
At age T1
Event features
At age T2
Classifier
at T2
Output probabilities
At age T2
Event features
At age T3
Classifier
at T3
Output probabilities
At age T3
Event features
At age T4
Classifier
at T4
Output probabilities
At age T4

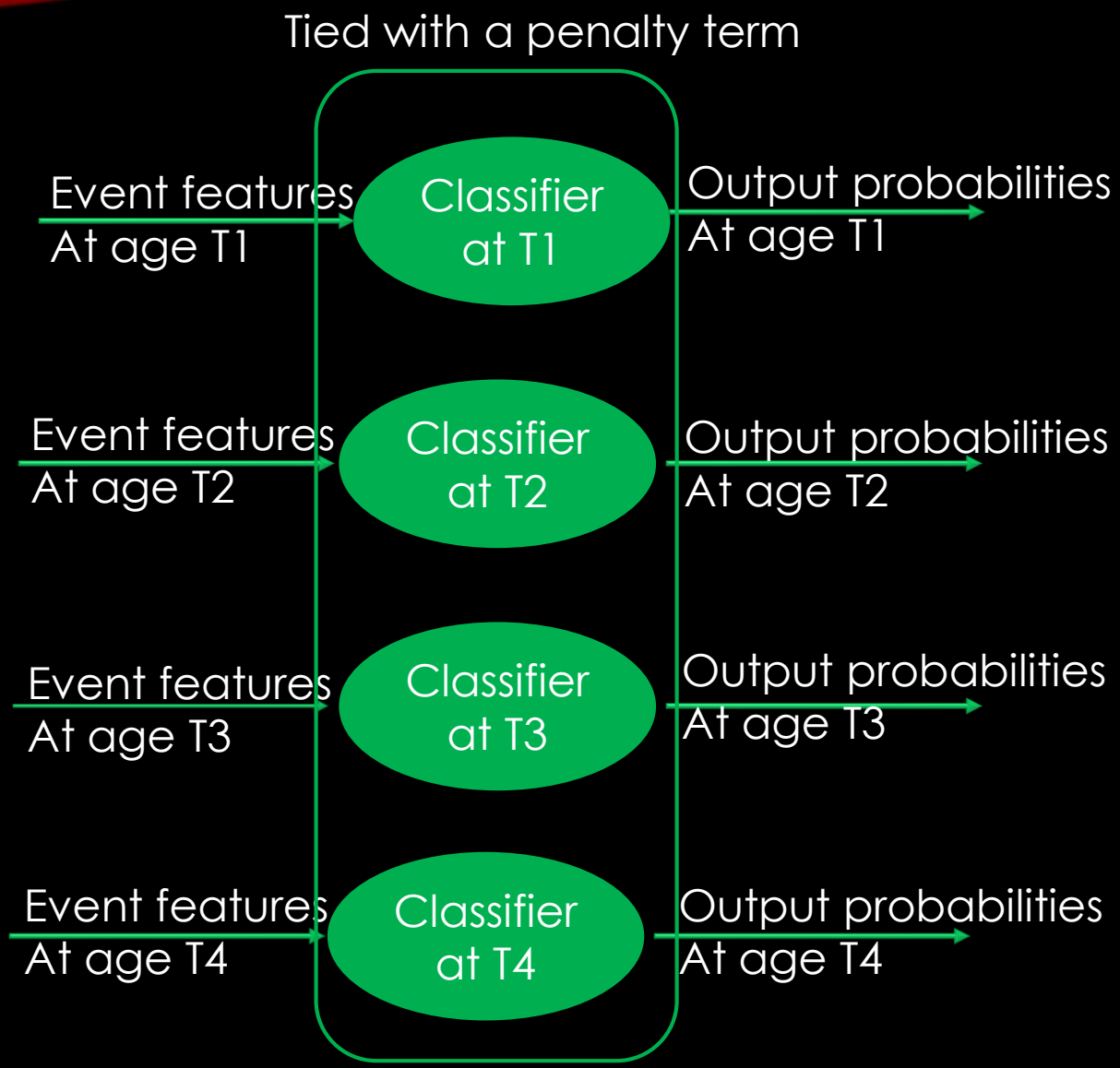
SAVEC


Event features
At age T1
Classifier
at T1
Output probabilities
At age T1
Event features
At age T2
Classifier
at T2
Output probabilities
At age T2
Event features
At age T3
Classifier
at T3
Output probabilities
At age T3
Event features
At age T4
Classifier
at T4
Output probabilities
At age T4

SAVEC

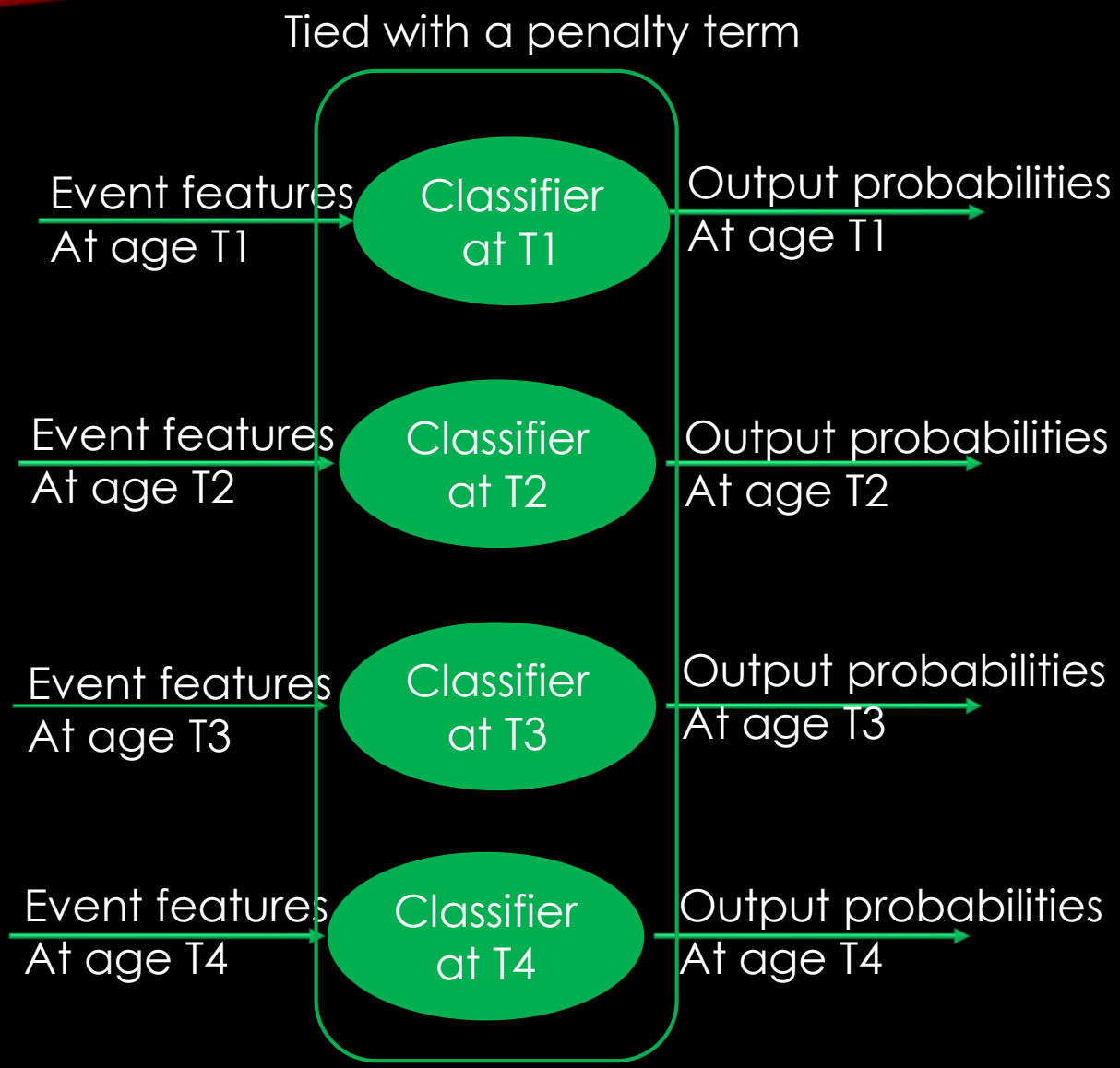

Tied with a penalty term
Event features
At age T1
Classifier
at T1
Output probabilities
At age T1
Event features
At age T2
Classifier
at T2
Output probabilities
At age T2
Event features
At age T3
Classifier
at T3
Output probabilities
At age T3
Event features
At age T4
Classifier
at T4
Output probabilities
At age T4

SAVEC



$$\varphi\left(\tilde{\beta},\lambda\right)=\frac{1}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n}\mathscr{L}\left(\boldsymbol{p}_{t_j}^{i},y_i;\tilde{\beta}\right)+\lambda\sum_{k=1}^{l}\sum_{j=1}^{n-1}\left(\tilde{\beta}_{j+1,k}-\tilde{\beta}_{j,k}\right)^2$$

SAVEC



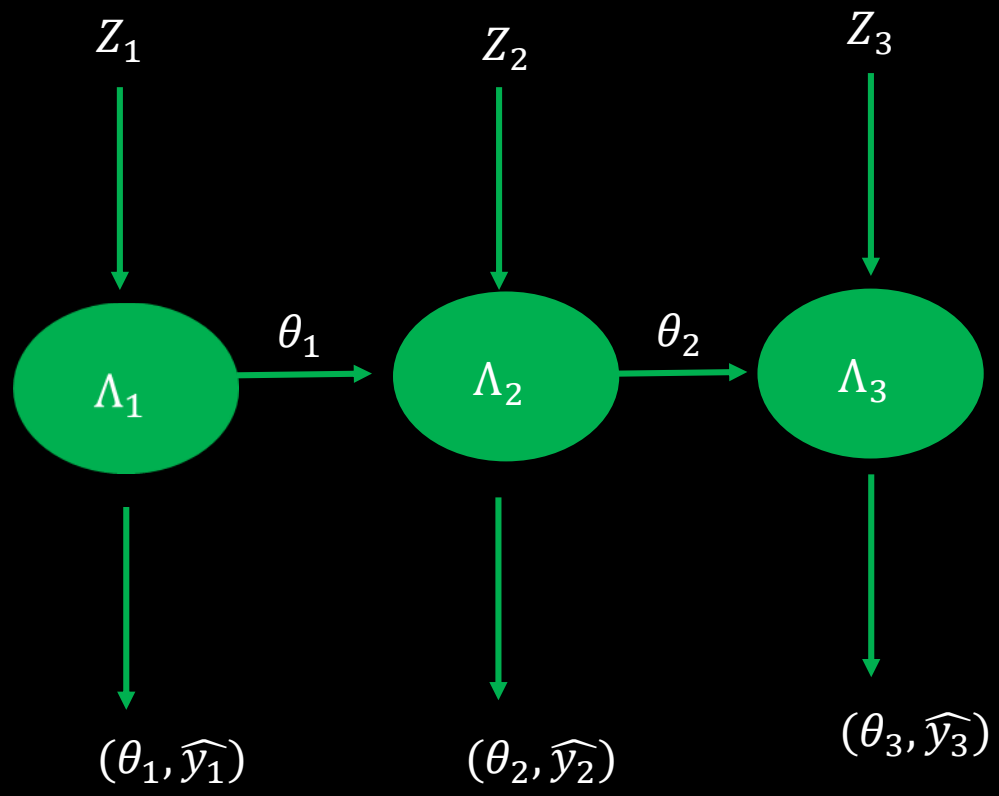
$$\varphi\left(\tilde{\beta},\lambda\right)=\frac{1}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n}\mathscr{L}\left(\boldsymbol{p}_{t_j}^{i},y_i;\tilde{\beta}\right)+\lambda\sum_{k=1}^{l}\sum_{j=1}^{n-1}\left(\tilde{\beta}_{j+1,k}-\tilde{\beta}_{j,k}\right)^2$$

Coefficients from T1 to T2, T2 to T3, T3 to T4 cannot change drastically. It is penalized.

DAVEC

- State space models for Gaussian
- $y_t = Z_t\alpha_t + \epsilon_t$
- $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$
- Z_t contains partial observations

- For a binary response
- $y_t \sim Bernoulli(p)$
- $logit(p_t) = Z_t\alpha_t$
- $\theta_t = Z_t\alpha_t$
- Feed the output to the next model



RESULTS – SYNTHETIC DATA

Classifier	Accuracy				Standard deviation			
	T1	T2	T3	T4	T1	T2	T3	T4
SAVEC	80	89	93	91	9	7	7	6
DAVEC	87	89	89	89	5	4	3	2
Logistic Regression	75	79	79	79	12	10	10	9

RESULTS – REAL DATA

Classifier	Accuracy – Area under ROC				Standard deviation			
	T1	T2	T3	T4	T1	T2	T3	T4
SAVEC	93	94	96	94	5	5	2	5
DAVEC	94	97	94	97	7	0	5	1
Logistic Regression	93	85	85	81	7	15	15	14

- R package eventstream : https://github.com/sevvandi/eventstream
- Preprint on Research gate : https://bit.ly/partialobs
- Monash is now hiring in business analytics – see **bit.ly/monash-ba** for details

- R package eventstream : https://github.com/sevvandi/eventstream
- Preprint on Research gate : https://bit.ly/partialobs
- Monash is now hiring in business analytics – see **bit.ly/monash-ba** for details

THANK YOU!