What is MSPRT-TANDEM

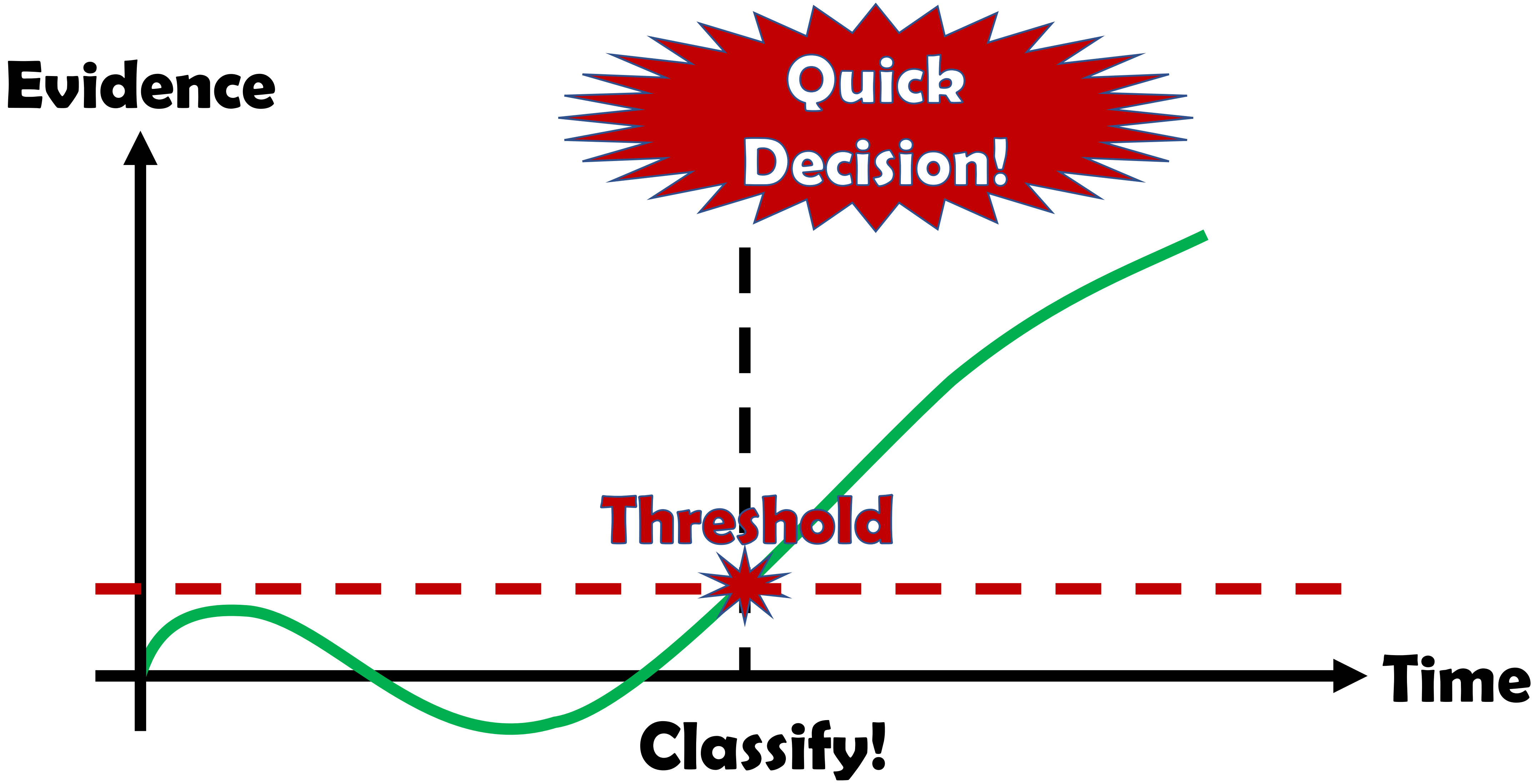
Early classification of time series

Is a task to classify sequential data as **early** and as **accurately** as possible.

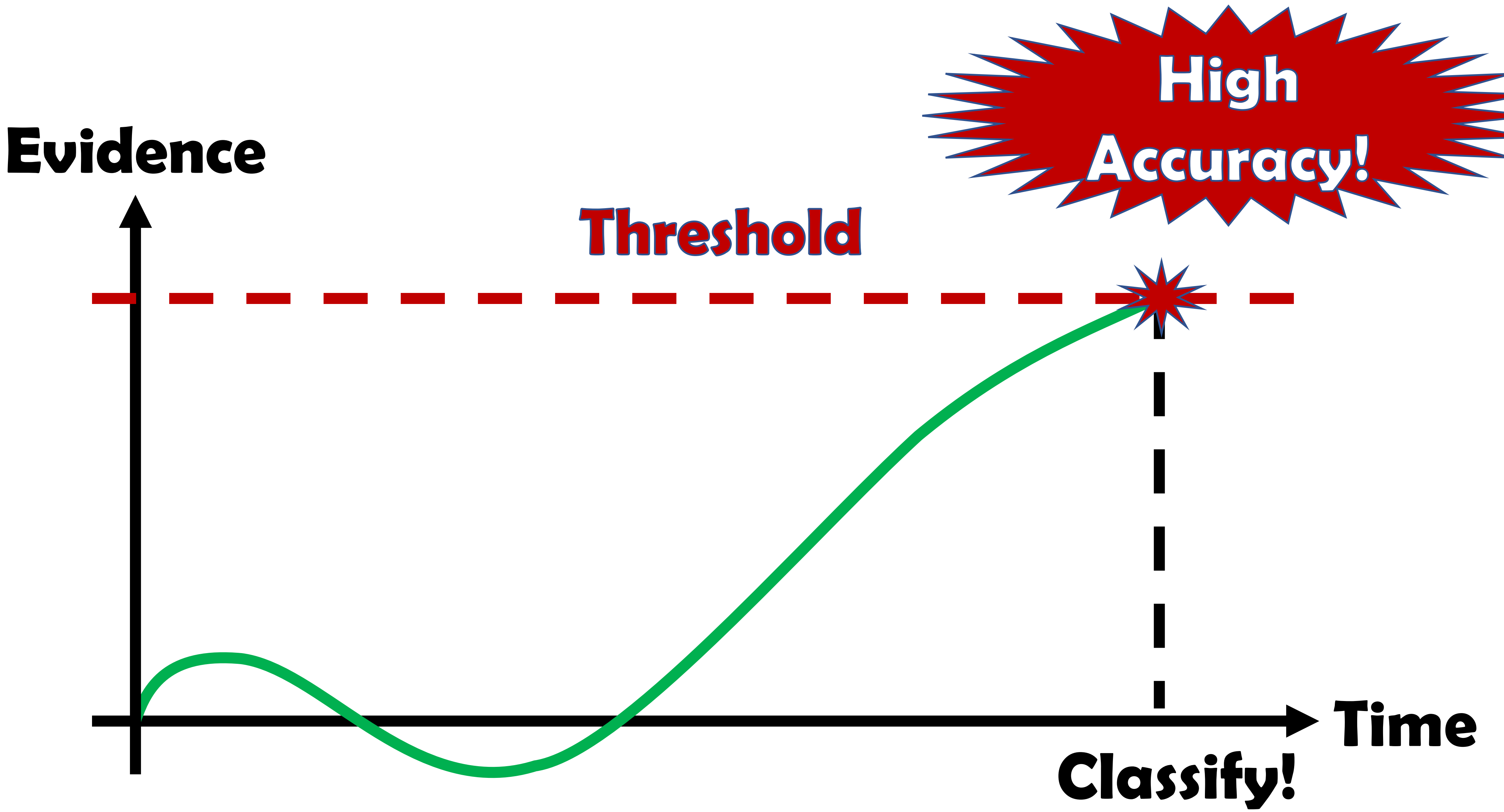
Multi-hypothesis
Sequential
Probability
Ratio
Test

is a nice algorithm for early multiclass classification.

MSPRT has a threshold.

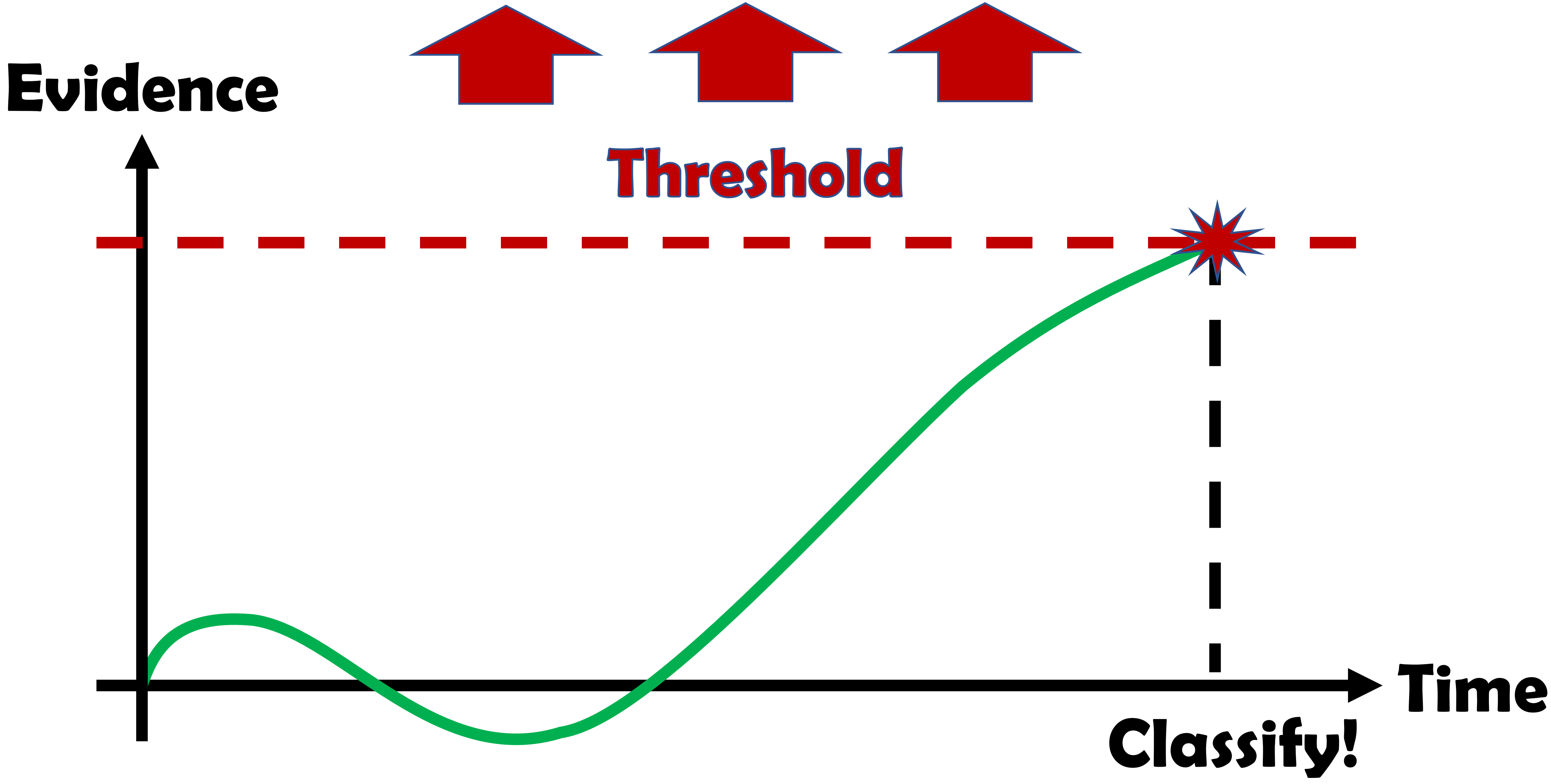


M
S
P
R
T has a threshold.



M S P R T

Asymptotic optimality



is known to be asymptotically optimal.

M

S Evidence := Log-Likelihood Ratio Matrix

$$\lambda(t) = \left(\log \frac{p(\text{input} = X(t) \mid \text{class} = k)}{p(\text{input} = X(t) \mid \text{class} = l)} \right)_{k,l}$$

P Prediction = $\underset{\text{rows}}{\operatorname{argmax}} \; \underset{\text{columns}}{\min} \; \lambda(\text{Hitting time})$

R

T observes LLR matrix.

M

S

P

R

T has a problem.

Evidence := Log-Likelihood Ratio Matrix

$$\lambda(t) = \left(\log \frac{p(\,\mathbf{input} = X(t) \mid \mathbf{class} = k)}{p(\,\mathbf{input} = X(t) \mid \mathbf{class} = l)} \right)_{k,l}$$

Prediction = $\underset{\text{rows}}{\operatorname{argmax}} \; \underset{\text{columns}}{\min} \; \lambda(\text{Hitting time})$

Who knows!?

Our approach

is

to estimate $\lambda(t)$

from a dataset...

...but it is difficult...

Multiple density ratios

$$\lambda(t) = \left(\log \frac{p(X(t) \mid k)}{p(X(t) \mid l)} \right)_{k,l}$$

Simultaneous optimization

$$= \begin{bmatrix} 0 & \log \frac{p(X(t) \mid 1)}{p(X(t) \mid 2)} & \cdots & \log \frac{p(X(t) \mid 1)}{p(X(t) \mid K-1)} & \log \frac{p(X(t) \mid 1)}{p(X(t) \mid K)} \\ \log \frac{p(X(t) \mid 2)}{p(X(t) \mid 1)} & \ddots & & \log \frac{p(X(t) \mid 1)}{p(X(t) \mid K-1)} & \log \frac{p(X(t) \mid 2)}{p(X(t) \mid K)} \\ \vdots & & 0 & & \vdots \\ \log \frac{p(X(t) \mid K-1)}{p(X(t) \mid 1)} & \log \frac{p(X(t) \mid K-1)}{p(X(t) \mid 2)} & & \ddots & \log \frac{p(X(t) \mid K-1)}{p(X(t) \mid K)} \\ \log \frac{p(X(t) \mid K)}{p(X(t) \mid 1)} & \log \frac{p(X(t) \mid K)}{p(X(t) \mid 2)} & \cdots & \log \frac{p(X(t) \mid K)}{p(X(t) \mid K-1)} & 0 \end{bmatrix}$$

Our LSEL is a solution!!

Minimizing LSEL

gives a precise estimate of

$$\lambda(t)$$

LSEL

is a log-sum-exp-type loss.

$$L \sim \sum_{\substack{k \\ \text{(classes)}}} \sum_{\substack{t \\ \text{(time)}}} \sum_{\substack{i \\ \text{(sequences} \\ \text{in class } k)}} \log \sum_{\substack{l \\ \text{(columns)}}} \exp\left(-\lambda_{kl}(X_i(t))\right)$$

We exploit this structure and prove three theoretical properties,
All of which contribute to the performance of the LLR matrix estimation...

LSEL

has three theoretical properties.

Consistency

By minimizing LSEL, we can get the true λ in the large training set limit.

Hard class weighting

LSEL gives large gradients for hard class examples.

Guess-aversion

LSEL returns discriminative "scores" λ even on class-imbalanced datasets.

The overall DNN-based model for early multiclass classificaiton

MSPRT + LSEL

= MSPRT-TANDEM

outperforms baseline models statistically significantly.

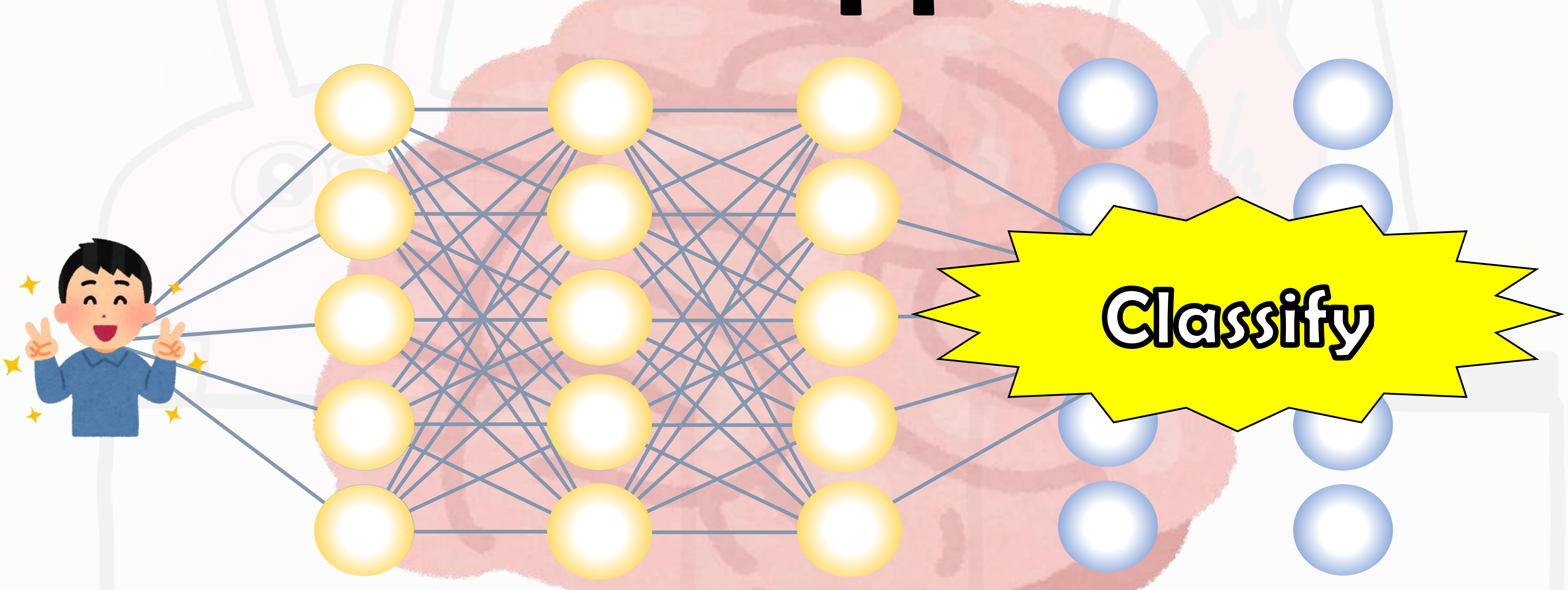
Potential Applications

Early detection of patient's deterioration

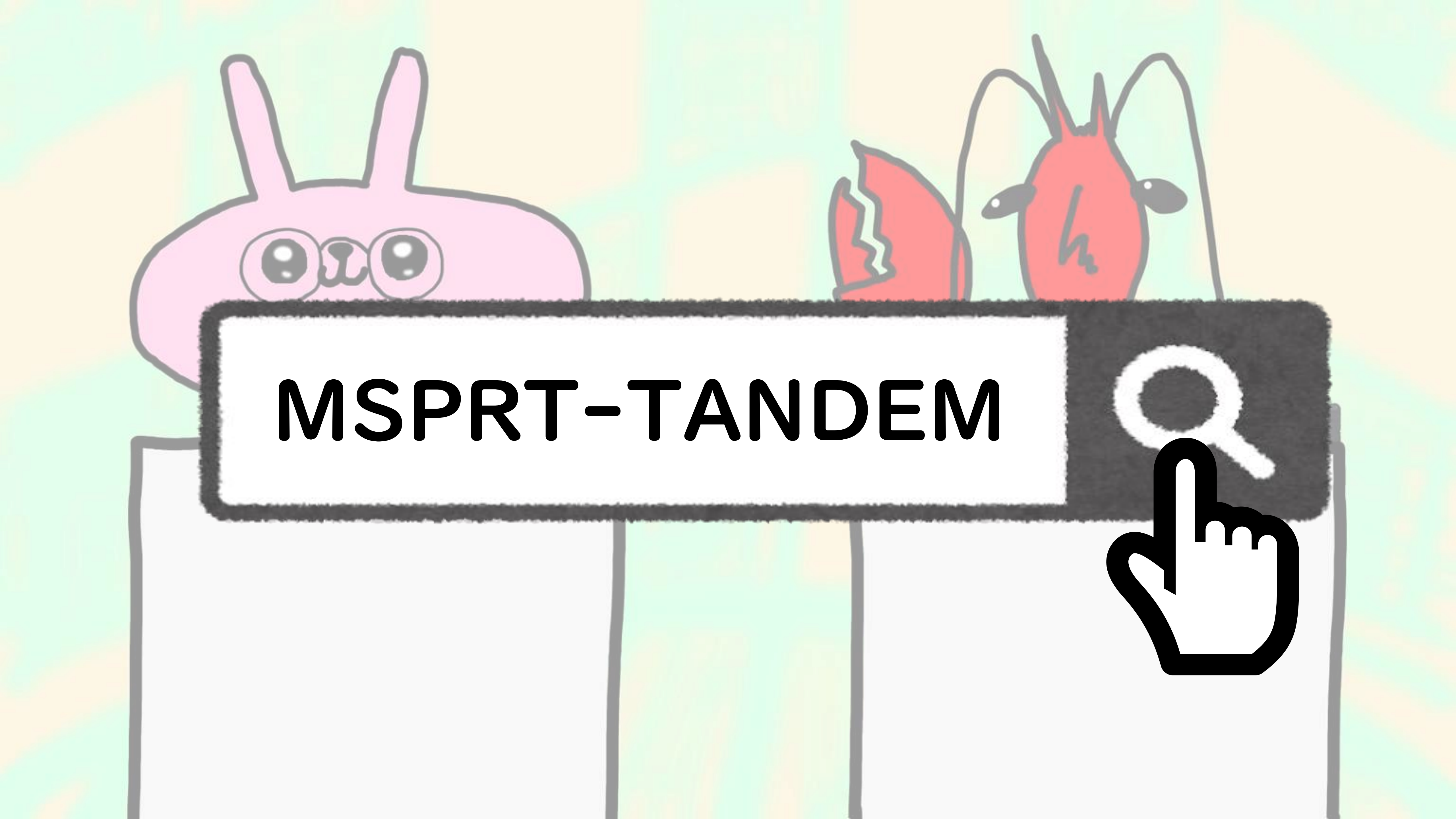
Potential Applications

Early action selection for autonomous driving

Potential Applications



Early exit to save computational costs and avoid overfitting in deep learning


MSPRT-TANDEM

Title

The Power of Log-Sum-Exp: Sequential Density Ratio Matrix Estimation for Speed-Accuracy Optimization

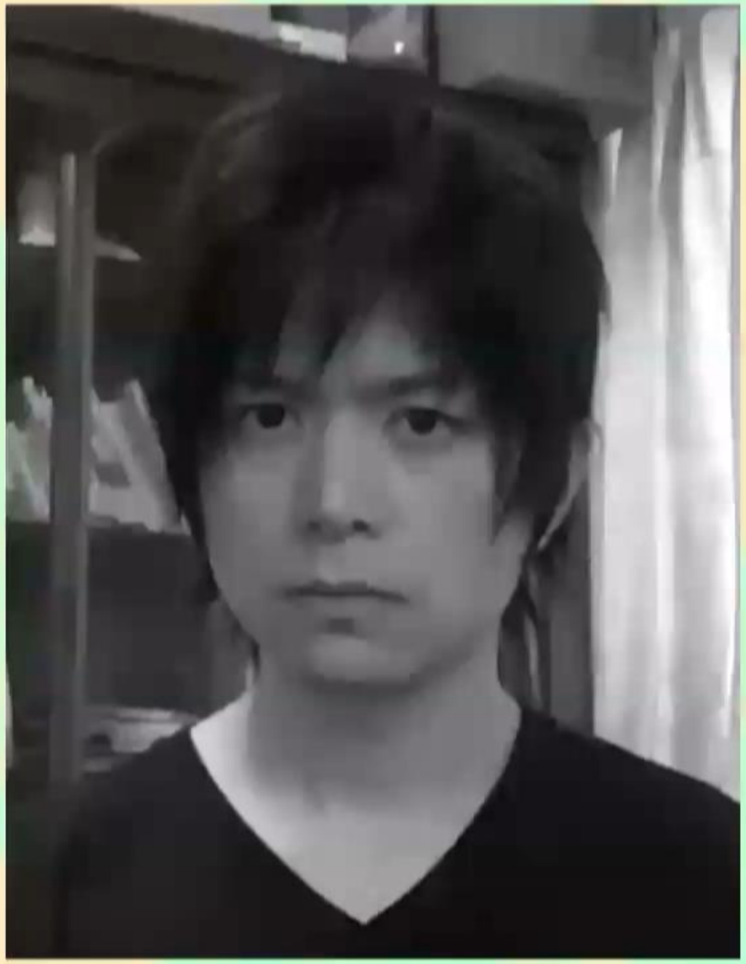
Authors

Taiki Miyagawa (NEC)
Akinori F. Ebihara (NEC)

Taiki Miyagawa

An ordinary rabbit.

Twitter @kanaheinousagi



Akinori F. Ebihara

A neuroscience Ph.D. working in the field of machine learning/computer vision.
An orchid enthusiast. A father of a son.

Twitter @non_iid

Summary

- The **MSPRT** is known to be **asymptotically optimal**, but it requires the true LLR of the input, which limits its real-world applications.
- We relax this critical requirement via **Density Ratio Matrix Estimation (DRME)**.
- We propose a loss function, the **LSEL**, for DRME.
- We prove that the LSEL is **consistent**, **weighs hard classes**, and is **guess-averse**.
- Our experiment shows that MSPRT+LSEL = **MSPRT-TANDEM** outperforms other baseline models on a large-scale action recognition datasets (UCF101 and HMDB51).
- <u>**MSPRT-TANDEM opens up possibilities for the MSPRT in wide variety of tasks.**</u>