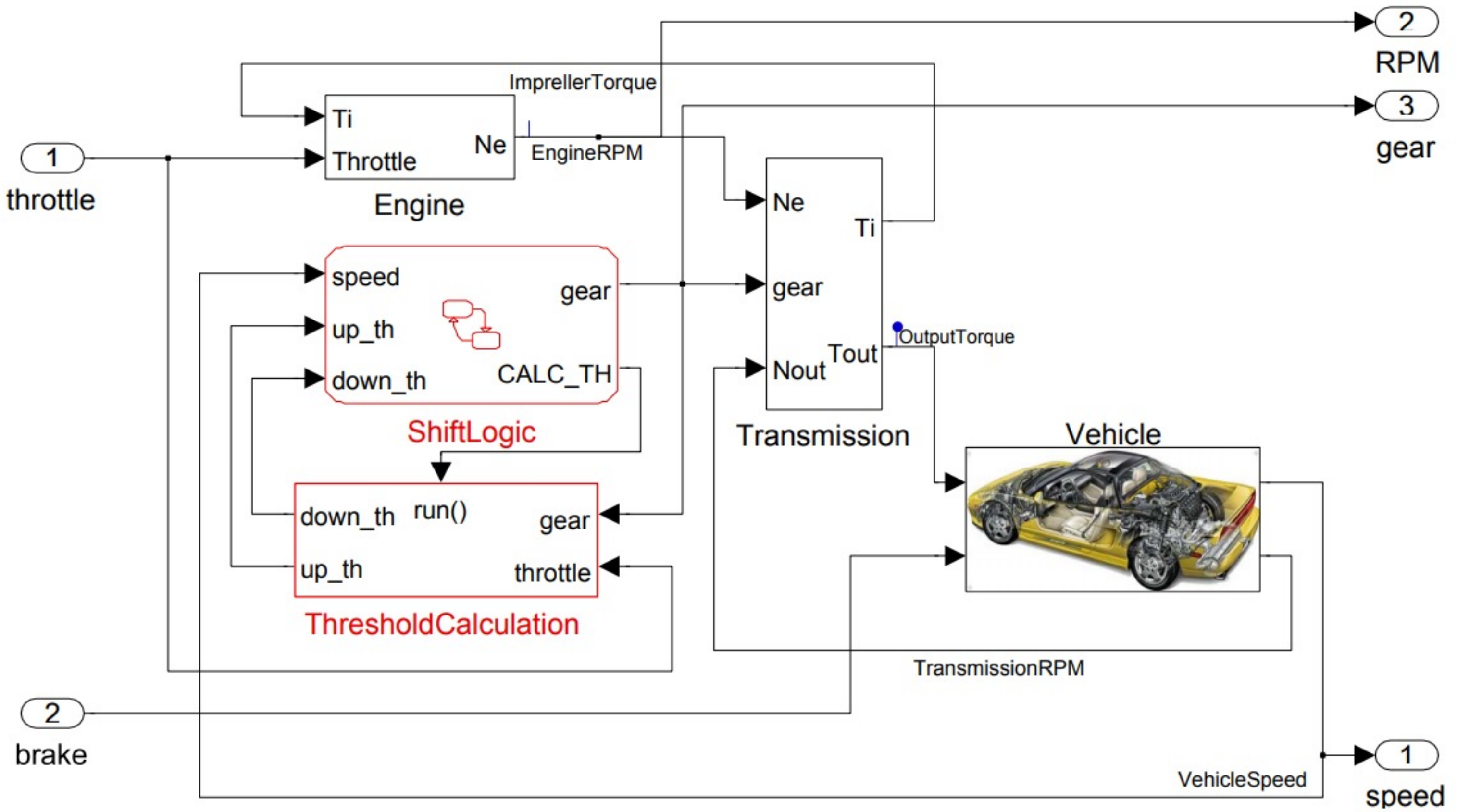
Cyber-Physical Systems

Laura Nenzi

Università degli Studi di Trieste
II Semestre 2021

Lecture: STL learning 2

Specification Mining



- What is the maximum speed that the vehicle can reach ?
- What is the minimum d well time in a given gear ?

Parametric Signal Temporal Logic

Definition (PSTL syntax)

$$\phi := (x_i \bowtie \pi) \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \, \mathcal{U}_{[\tau_1,\tau_2]} \, \varphi_2$$

with $\bowtie \in \{>, \leq\}$

- π is **threshold** parameter
- τ_1, τ_2 are **temporal** parameters

- $\mathbb{K} = (\mathcal{T} \times \mathcal{C})$ be the **parameter space**
- $\boldsymbol{\theta} \in \mathbb{K}$ is a **parameter configuration**

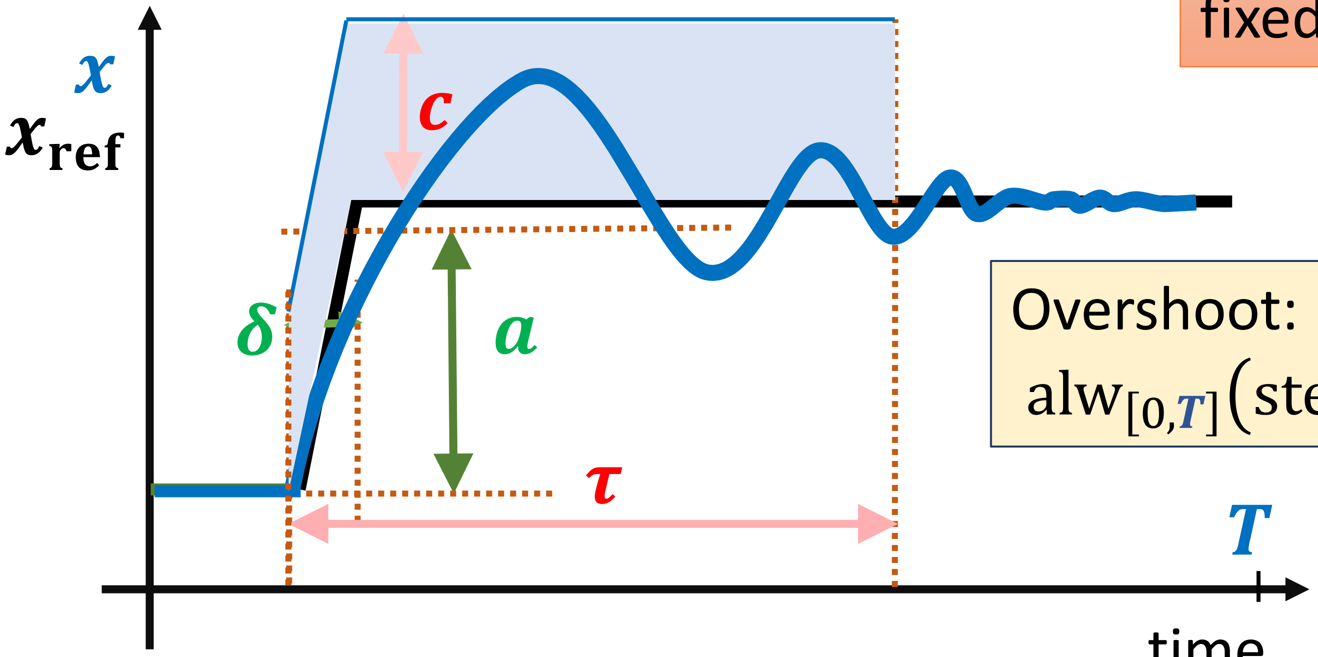
e.g., $\phi = \mathcal{F}_{[a,b]}(x_i > k), \boldsymbol{\theta} = (0, 2, 3.5)$ then $\phi_{\boldsymbol{\theta}} = \mathcal{F}_{[0,2]}(x_i > 3.5)$.

Specification Mining

- Specification Mining: Try to find values of parameters of a PSTL formula from a given model

- Why?
  - Good to know “as-is” properties of the model
  - Finds worst-case behaviors of the model
  - Discriminates between regular and anomalous behaviours

Specification Templates using PSTL



In previous lecture, a, c, T, τ, δ were some fixed values, here they represent parameters

Step:

$$\text{step}(y) \coloneqq y(t+\delta) - y(t) > a$$

Overshoot:

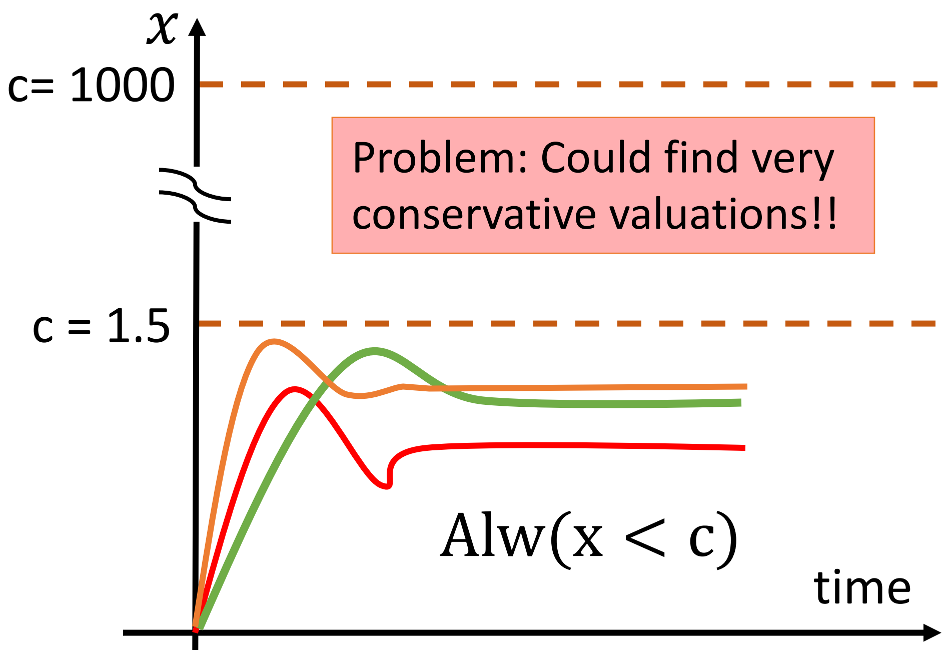
$$\text{alw}_{[0,T]}\big(\text{step}(x_{\text{ref}}) \Rightarrow \text{alw}_{[0,\tau]}(x(t) - x_{\text{ref}}(t) < c)\big)$$

Parameter inference for PSTL

- Given:
  - PSTL formula $\varphi(\mathbf{p})$, $[\mathbf{p} = (p_1, p_2, \dots, p_m)]$
  - Traces $x_1, \dots, x_n$
- Find:
  - ~~Valuation~~ $\nu(\mathbf{p})$ such that: $\forall i : x_i \vDash \varphi\big(\nu(\mathbf{p})\big)$

    δ-tight valuation

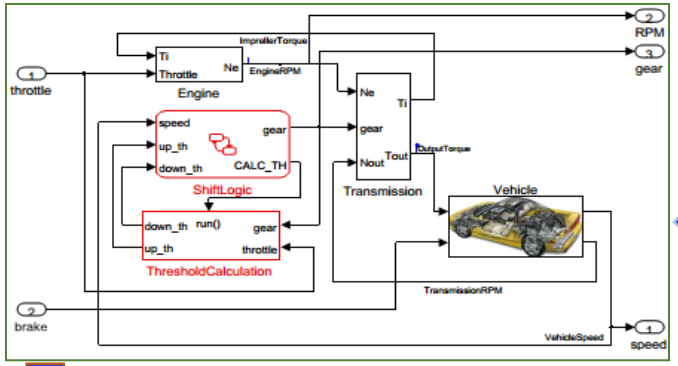
  - and $\exists i: x_i \nvDash \varphi(\nu(\mathbf{p}) \pm \delta)$:
    i.e. small perturbation in $\nu(\mathbf{p})$ makes some trace not satisfy formula



formula sat for given valuation $\Rightarrow$ $\forall$ greater (or lesser) valuations sat

Finding δ-tight valuations hard in general, but **efficient** for **Monotonic PSTL**

Binary search on parameter space



Falsification:
∃ **trace** ⊭ Property?

YES

NO

Specification Mining

Secret Sauce:

- Infer parameters for a given PSTL formula from traces
- Falsify given STL formula

Parar

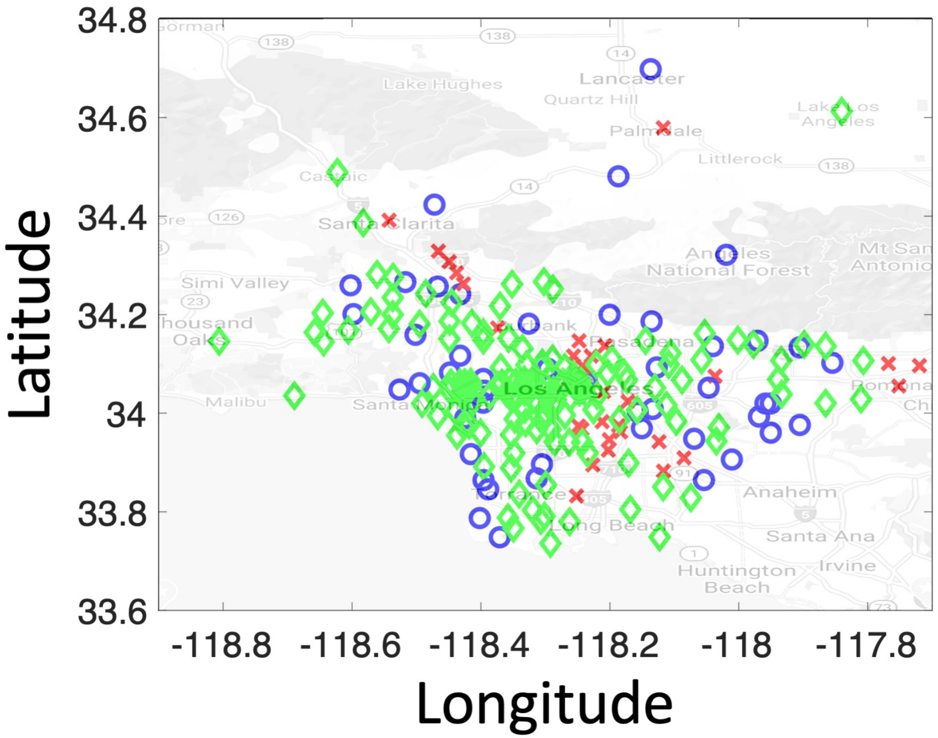
Find “Tightest” Answers

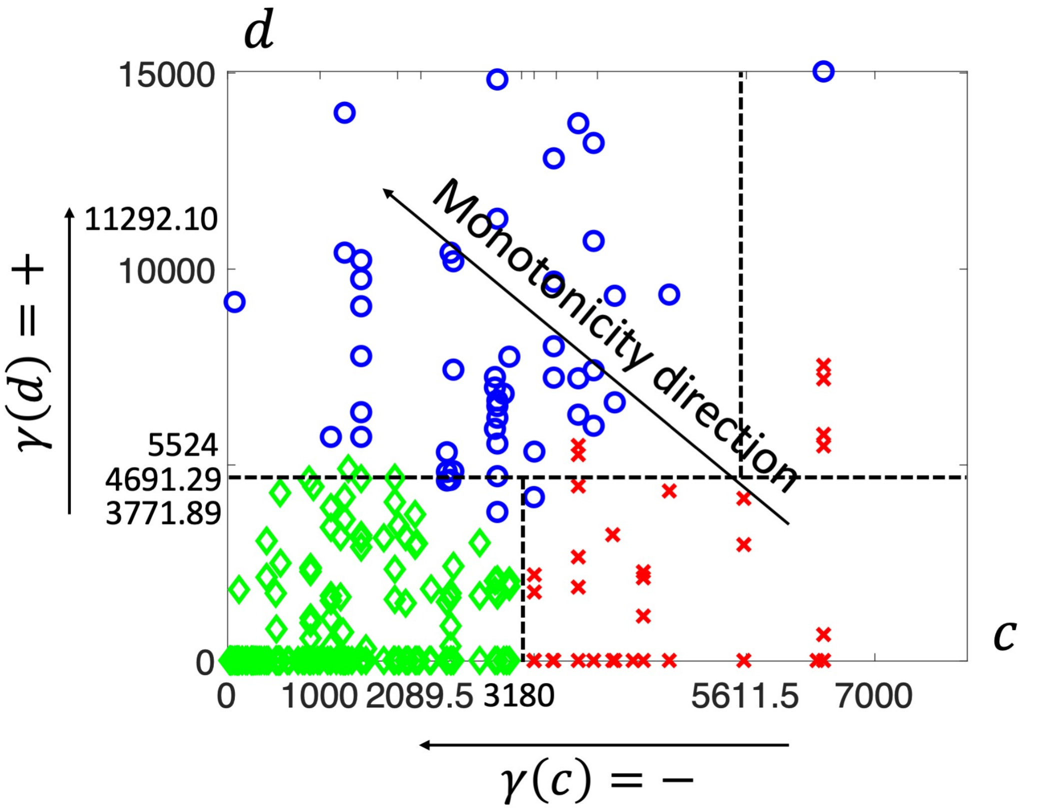
Overshoot is 5.5 KPa
Upper Bound on x is 5

Settling Time is **??**
Overshoot is **??**
Bounds on *x* are **??**

Settling Time is 6 ms
Overshoot is 5.5 KPa
Upper Bound on x is 5

Learning STL-based clustering (Unsupervised Learning)





Goal: clusterizing spatio-temporal data using formal logic

Motivation

- Spatially distributed systems generate a large volume of spatio-temporal data
- Designers are interested in analyzing and extracting high-level structure from such data

Traditional ML for time-series clustering:

- Popular techniques:
  - Kmeans
  - Hierarchical Clustering
  - Agglomerative clustering
  - Shapelets
- Pros:
  - Fast
- Cons:
  - Based on shape-similarity
  - May lack interpretability

STL-based clustering of time-series data:

- Considerable interest in learning logical properties of temporal data using logics such as Signal Temporal Logic (STL)

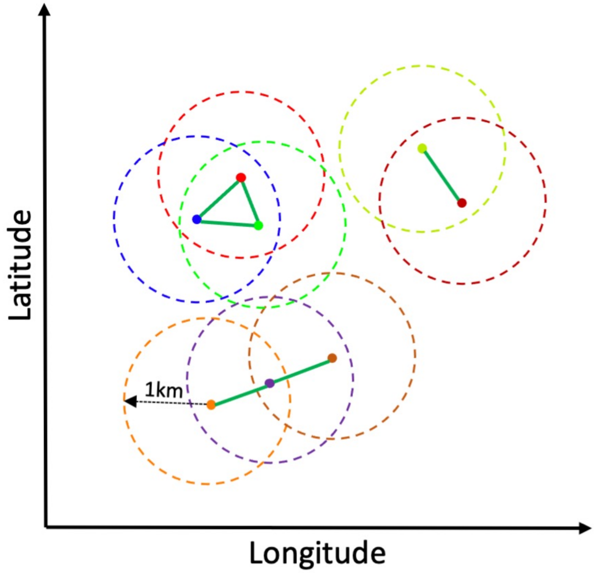
- Algorithms for unsupervised learning of spatio-temporal data using formal logics

Spatial Model:

We model the spatial configuration as a weighted graph $S = \langle L, W \rangle$

L: set of locations

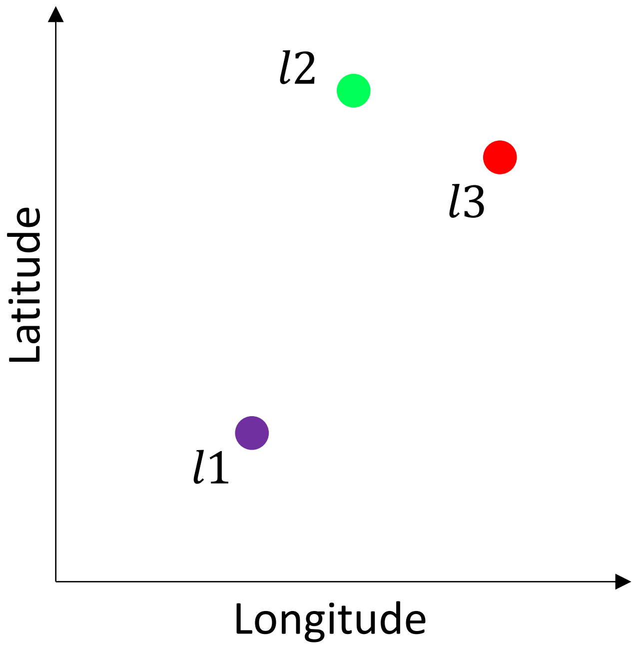
W: proximity relation between locations

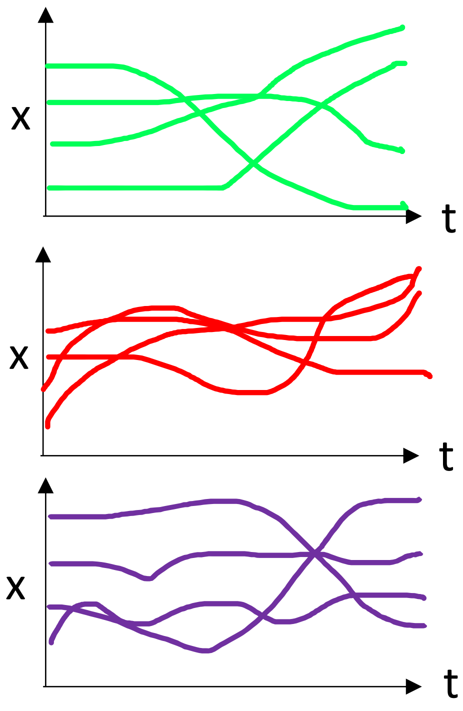


Connectivity graph
$W: spatial\ proximity$

Spatio-temporal trace:

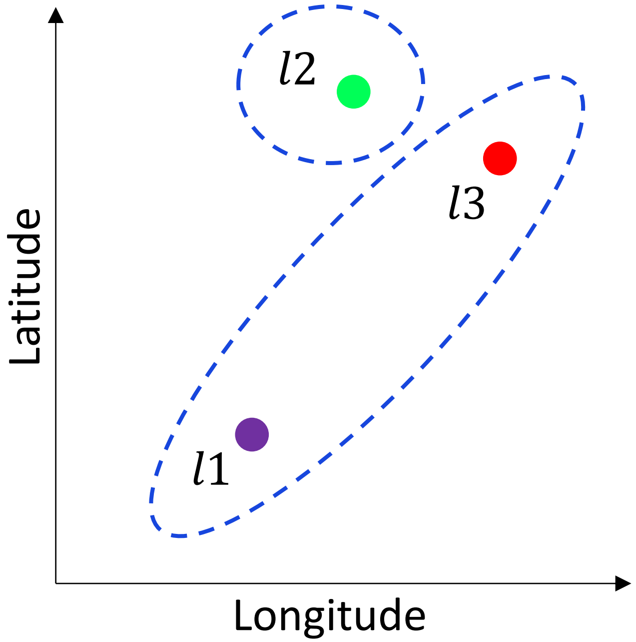
- Time-series data (trace/signal): a sequence of data values indexed by time stamps
- A spatio-temporal trace associates each location in a spatial model with a time-series trace

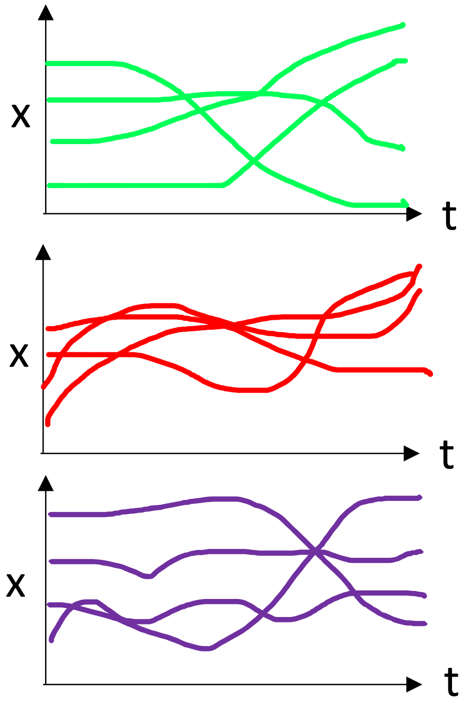




Spatio-temporal data clustering:

- It is a process of grouping data with similar spatial attributes, temporal attributes, or both





Parametric STREL (PSTREL):

- Replacing values in STREL by parameters

$$\boxed{\varphi_1 R_{[0,1000]} \varphi_2}$$

$$\boxed{\varphi_1 R_{[d_1,d2]} \varphi_2}$$

Monotonic PSTREL $\varphi(p)$:

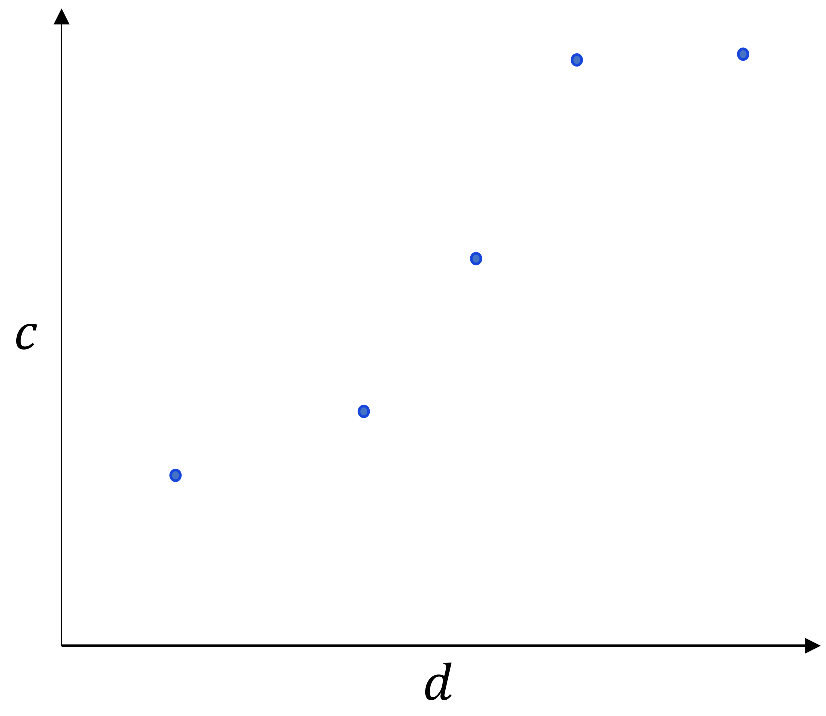
- The polarity of a parameter p is:
  - $+$ if it is easier to satisfy φ as we increase the value of p
  - $-$ if it is easier to satisfy φ as we decrease the value of p
- Monotonic PSTREL:
  - All parameters have either $+$ or $-$ polarity
- Example: $\square_{[0,d]}\varphi$
  - Polarity of d is $-$

Validity Domain of PSTREL $\varphi(p)$

- Given a location l
- A set of spatio-temporal traces X associated with l
- The set of all valuations to p such that each trace in X satisfies the STREL formula
- Boundary of the validity domain:

  The robustness value with respect to at least one trace in X is ≈ 0
- Robustness means distance to satisfaction or violation

$\square_{[0,d]} y < c$



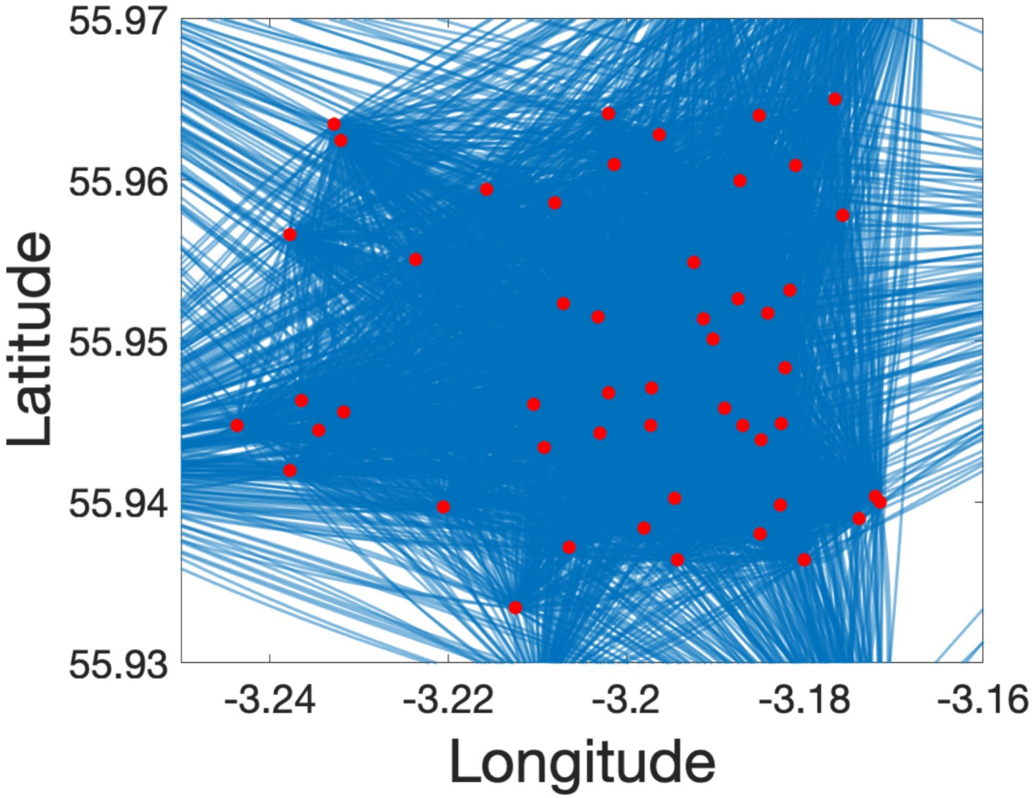
High-level steps:

- Constructing the spatial model
- Projecting each spatio-temporal trace to a tight valuation in the parameter space of a given PSTREL formula
- Clustering the trace projections
- Learning bounding boxes for each cluster using a Decision Tree based approach
- Learning a STREL formula for each cluster
- Improving the interpretability of the learned STREL formulas

Constructing Spatial Model:

Approach 1: fully connected graph

- Pros: gives the most accurate result
- Cons: computationally expensive

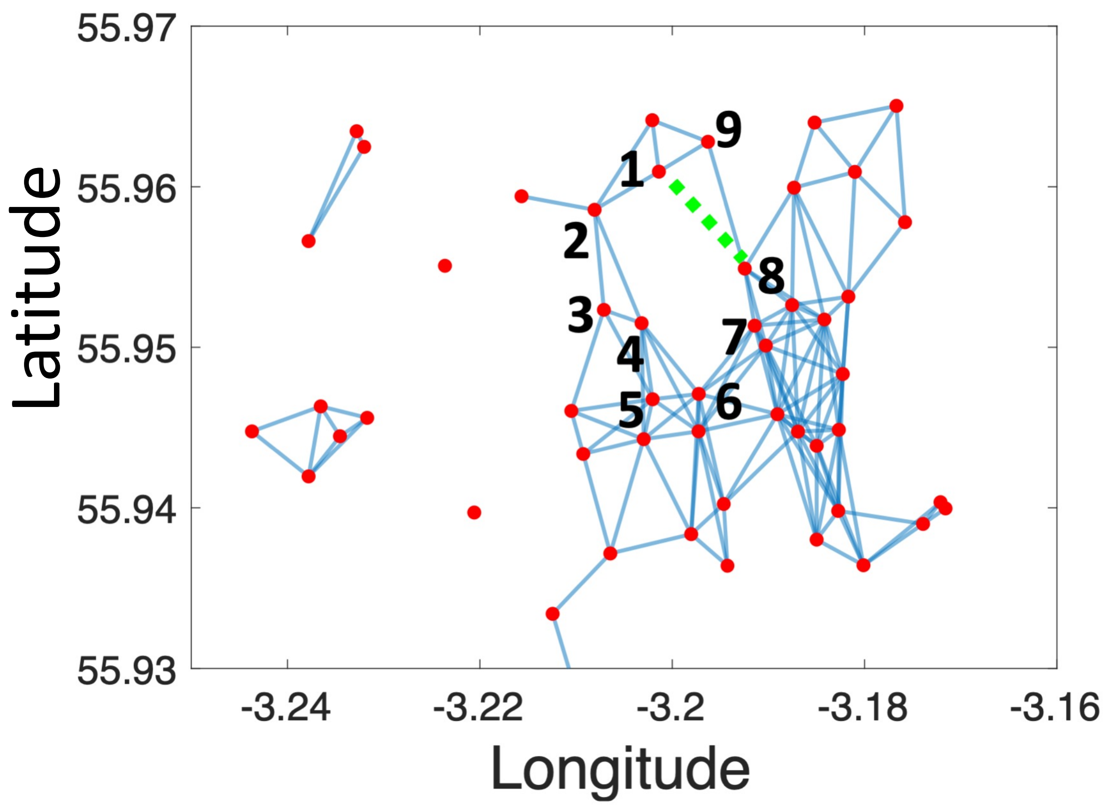


Constructing Spatial Model:

Approach 2: Connectivity graph that connects locations with distance less than a threshold

- Pros: lower cost
- Cons: disconnected spatial model

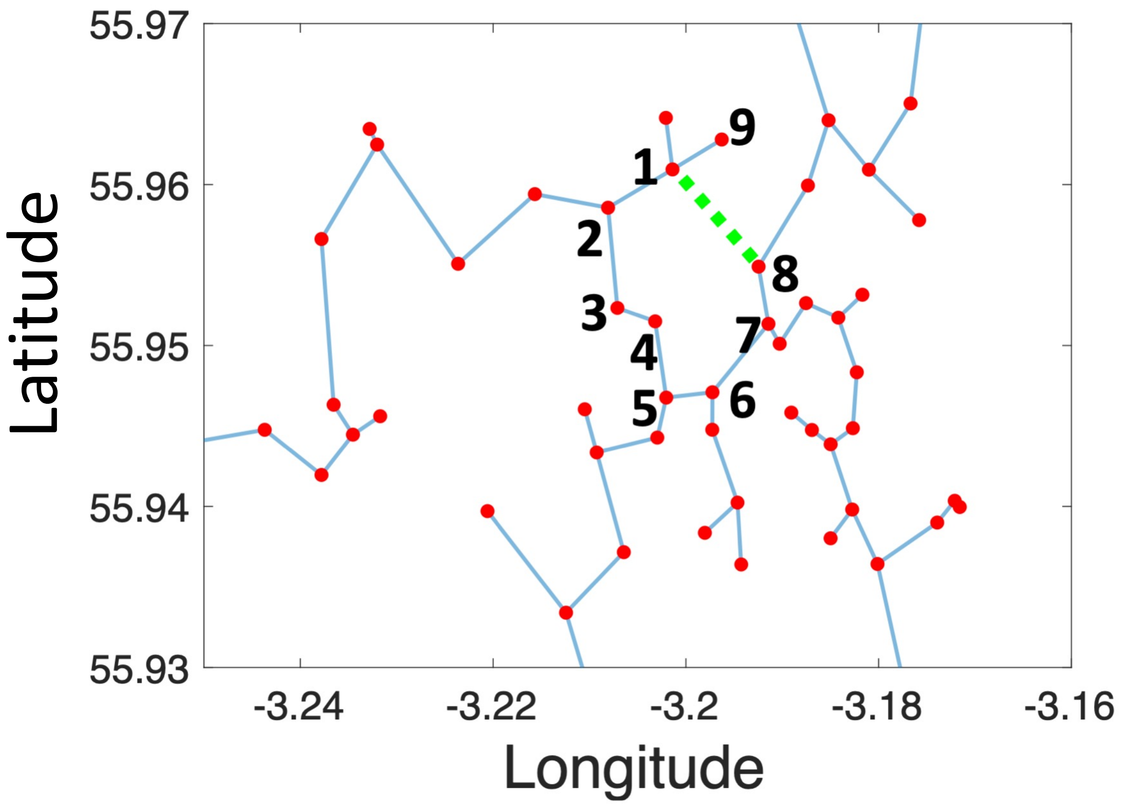
which affects the accuracy



Constructing Spatial Model:

Approach 3: Minimum Spanning Tree (MST)

- Pros: low cost and connected graph
- Cons: overestimation of distance between some nodes



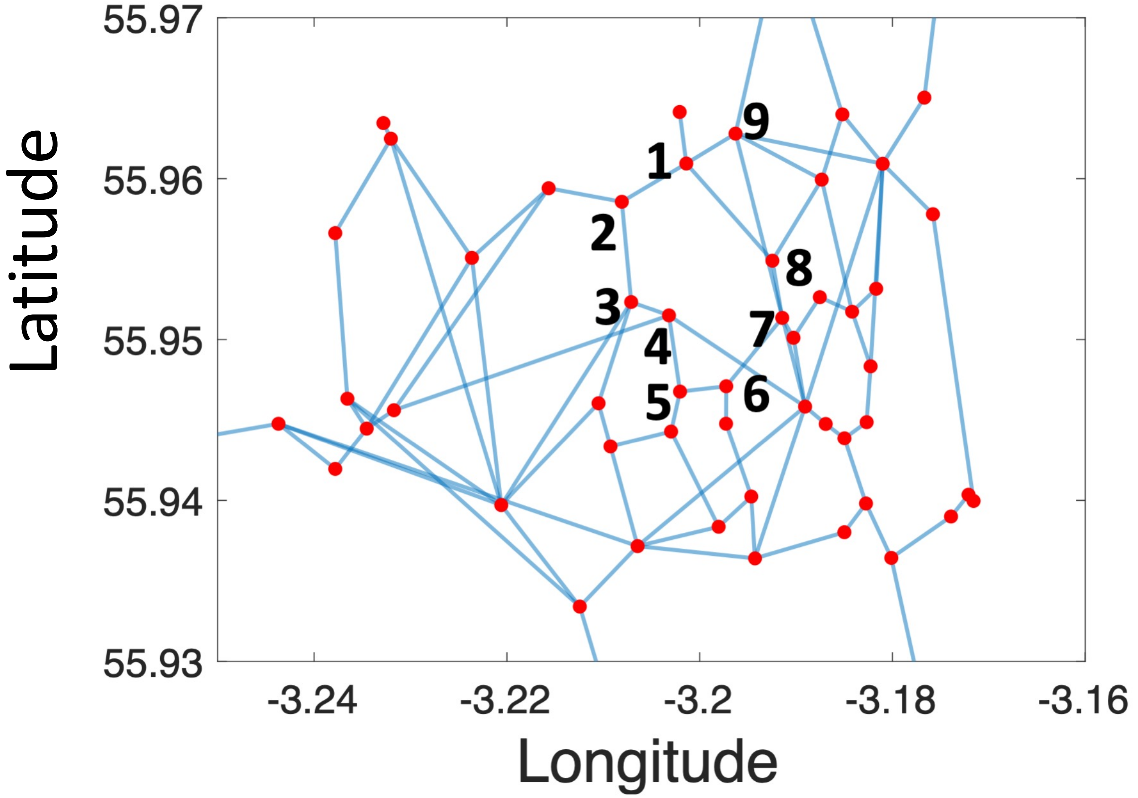
Constructing Spatial Model:

Approach 4: Enhanced Minimum Spanning Graph

Step1: create an MST

Step2: connect nodes that their shortest distance through MST is more than α times their actual distance (default $\alpha = 2$)

- Pros: low cost, connected graph and more accurate distance between nodes
- Cons: not as accurate as fully connected graph

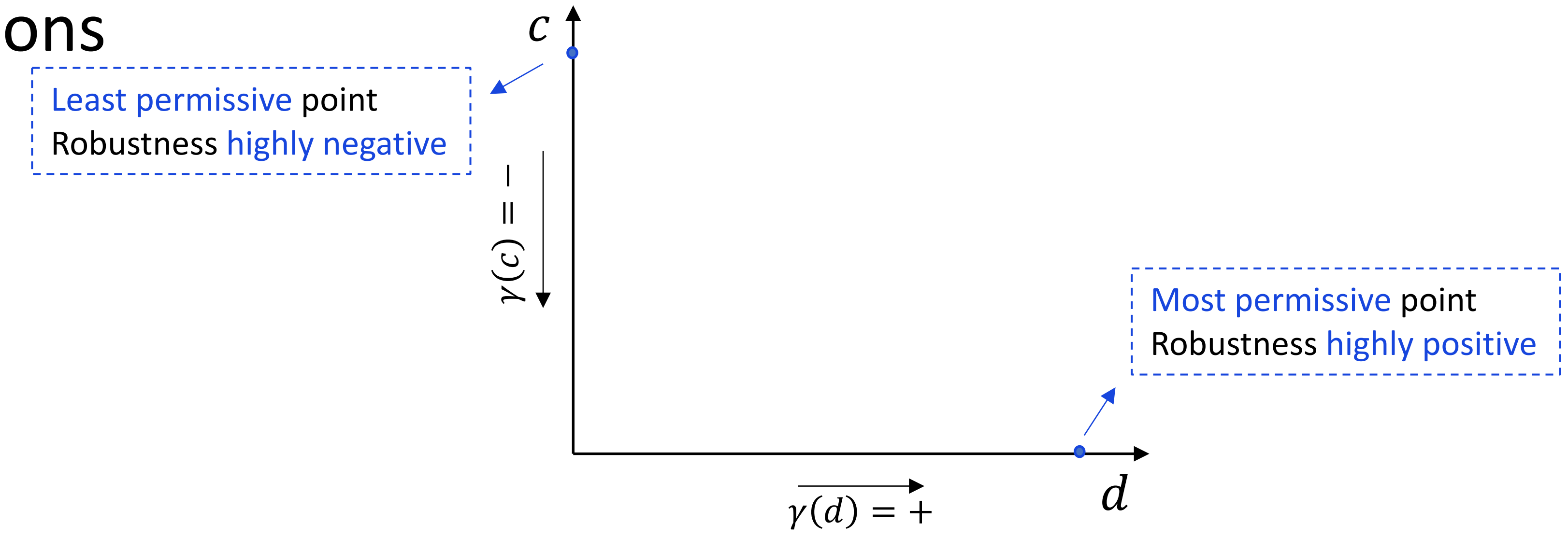


Spatio-temporal trace projection :

- The user provides a PSTREL formula

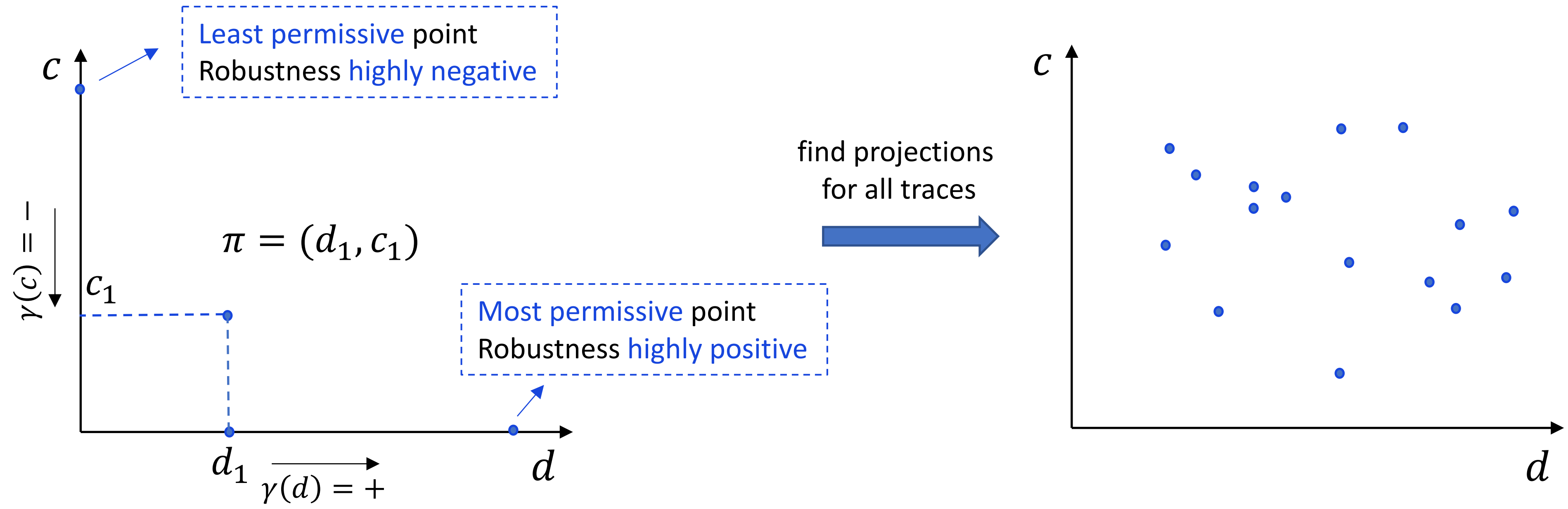
$$G_{[0,3hours]} \diamond_{[0,d]} (Bikes > c)$$

- The goal is to learn the tight parameter valuations for each spatio-temporal trace
- Tight parameter valuation is not unique, and each point on the boundary of validity domain corresponds to a tight parameter valuations



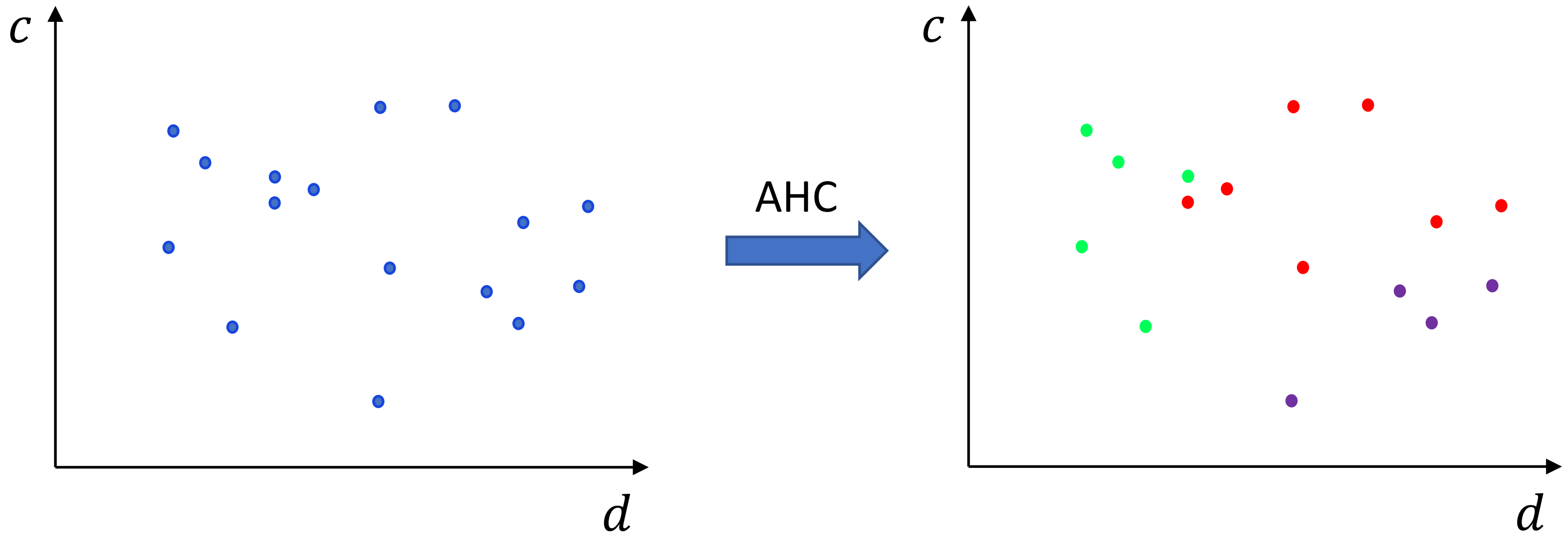
Spatio-temporal trace projection :

- We assume some ordering or priority on parameter space, e.g., $d >_p c$, provided by user
  1. Bisection search on d
  2. Bisection search on c



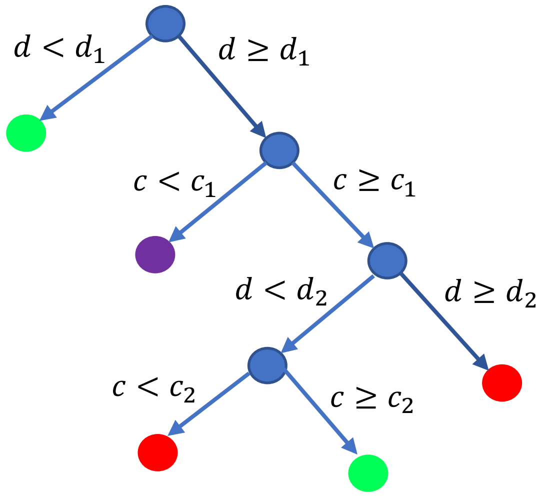
Clustering:

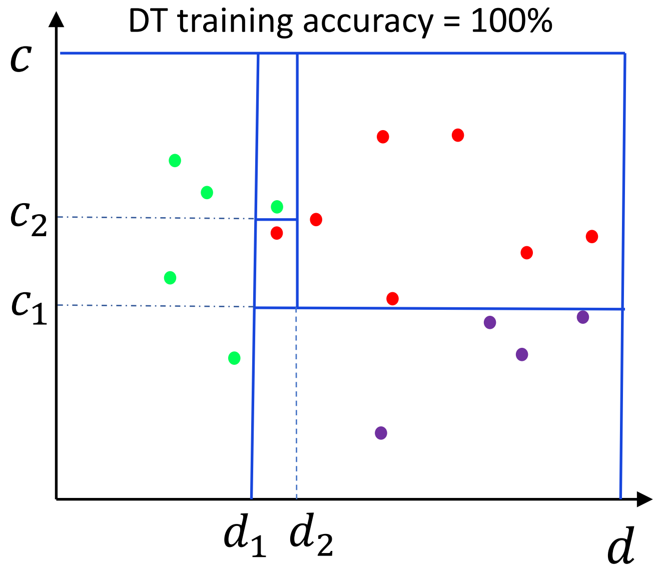
- The parameter valuation points serve as features for off-the-shelf clustering algorithms
- We use the Agglomerative Hierarchical Clustering (AHC) technique
- Number of clusters to choose:
  - Domain knowledge/Silhouette metric



Learning bounding boxes for each cluster:

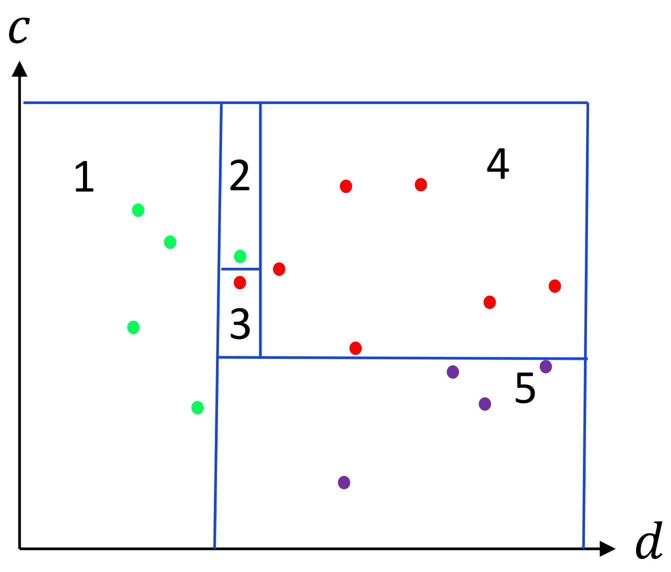
- We label each parameter valuation with its cluster
  - Labels = (green, red, purple)
- We use off-the-shelf Decision Tree (DT) algorithms to learn bounding boxes



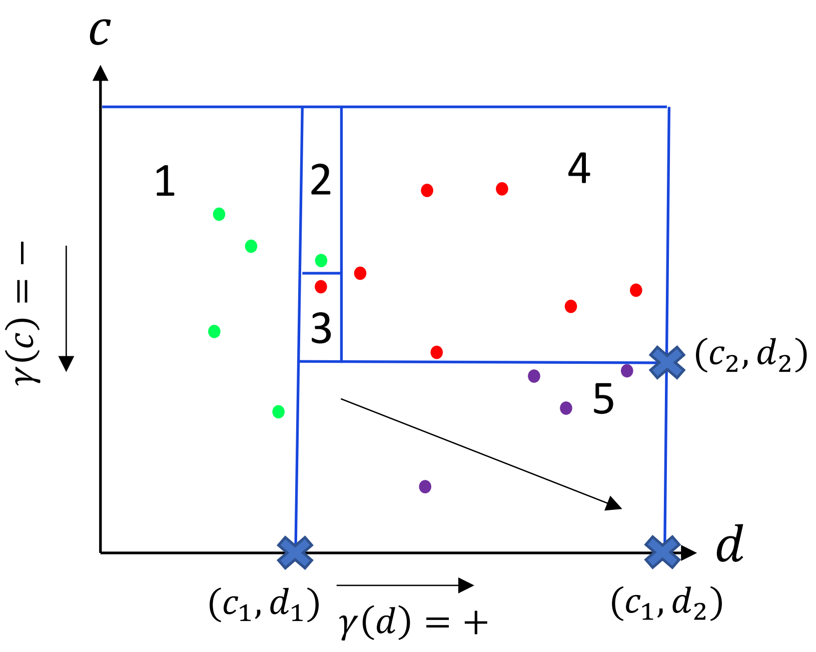


Learning a STREL Formula for each Cluster:

- $\varphi_{green} = \varphi_1 \vee \varphi_2$
- $\varphi_{red} = \varphi_3 \vee \varphi_4$
- $\varphi_{purple} = \varphi_5$



Learning a STREL Formula for each Cluster:



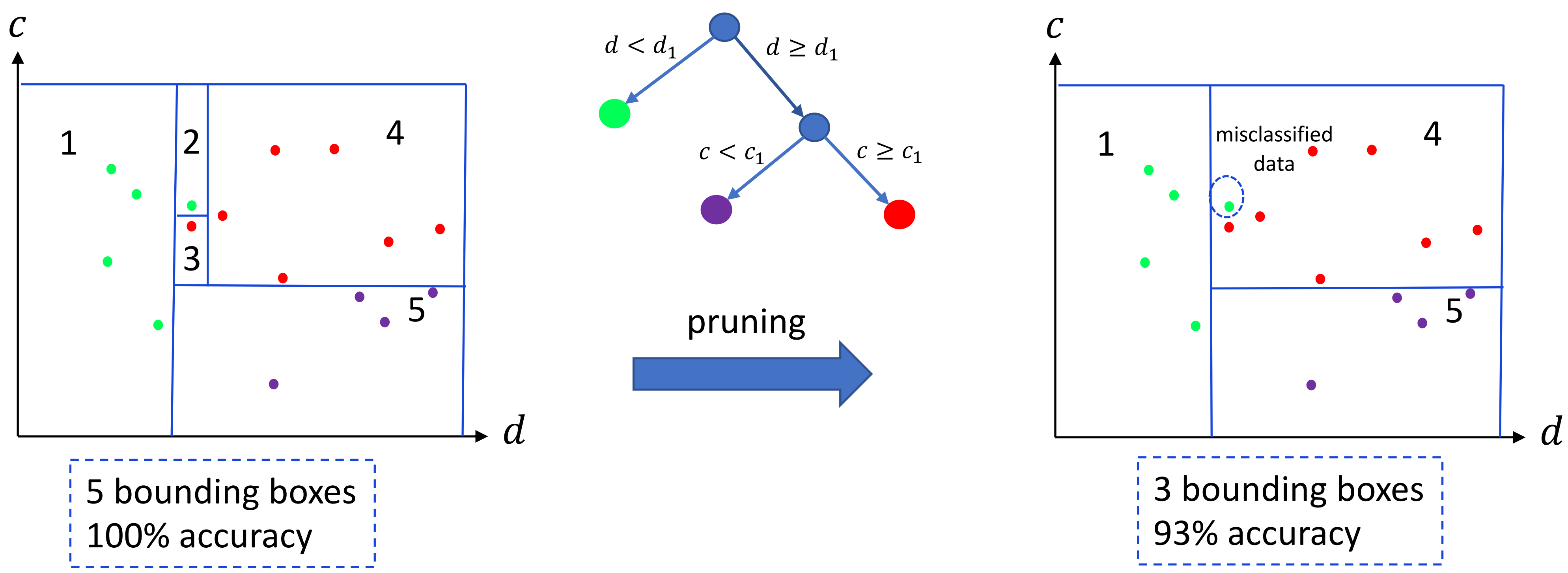
$$\varphi_5 = \varphi(\mathrm{c}_1, \mathrm{d}_2) \wedge \neg\varphi(\mathrm{c}_1, \mathrm{d}_1) \wedge \neg\varphi(c_2, d_2)$$

$$\varphi = G_{[0,3hours]} \diamond_{[0,d]} (Bikes > c)$$

$$\varphi_5 = G_{[0,3hours]} \diamond_{[0,d_2]} (Bikes > c_1)$$
$$\wedge \neg G_{[0,3hours]} \diamond_{[0,d_1]} (Bikes > c_1)$$
$$\wedge \neg G_{[0,3hours]} \diamond_{[0,d_2]} (Bikes > c_2)$$

Pruning the Decision Tree:

- In some cases, achieving 100% accuracy can result in long and hence less interpretable formulas
- We prune the DT using a K-fold cross validation approach



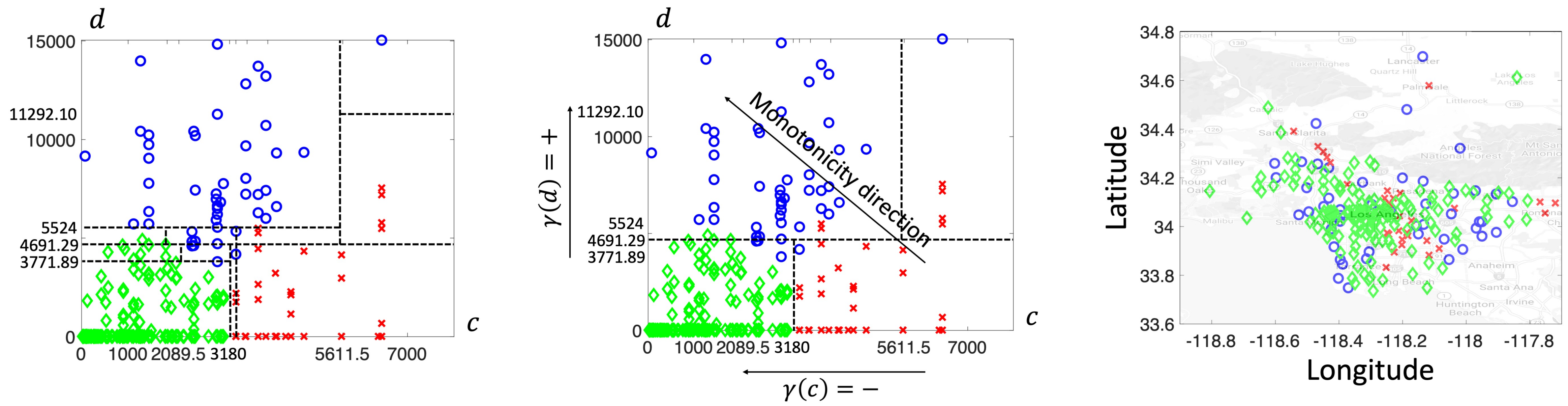
Benchmarks:

- COVID-19 data from LA County
  - COVID-19 pandemic has extremely affected our lives
  - Understanding the spread pattern of COVID-19 in different areas is vital to stop the spread of the disease.
  - We focus on number of new positive cases in each region of the LA county
- BSS data from the city of Edinburgh
  - The BSS consists of a number of bike stations, distributed over a geographic area
  - We focus on the number of bikes (B) and empty slots (S) in each bike station
  - We are interested in analyzing the behavior of each station
- Outdoor Air Quality data from California
- Synthetic data for a food court building

COVID-19 data from LA County

PSTREL formula: $\Diamond_{[0,d]}\{F_{[0,\tau]}(x > c)\}$

- We fix τ to 10 days
- Small d and large c are **hot spots**



$$\varphi_{red} = \Diamond_{[0,4691.29]}\{F_{[0,10]}(x \geq 3181)\} \vee \Diamond_{[0,15000]}\{F_{[0,10]}(x \geq 5612)\}$$

BSS data from the city of Edinburgh

PSTREL formula:

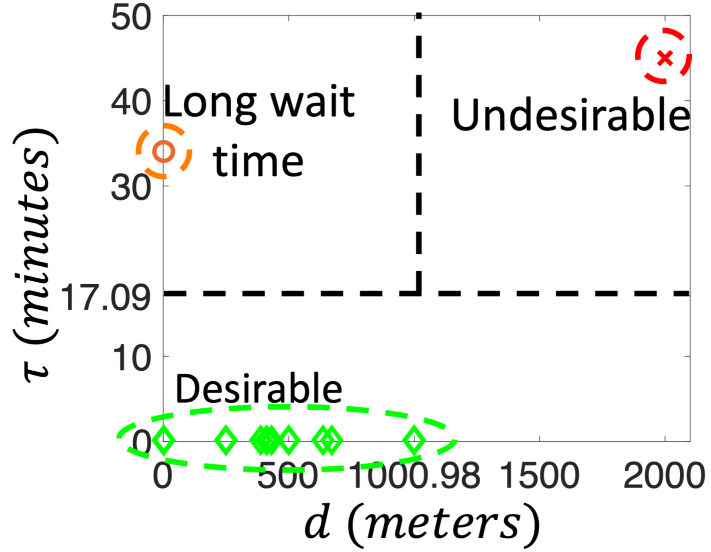
$$\varphi(\tau, d) = G_{[0,\tau]}\big(\varphi_{wait}(\tau) \vee \varphi_{walk}(d)\big)$$

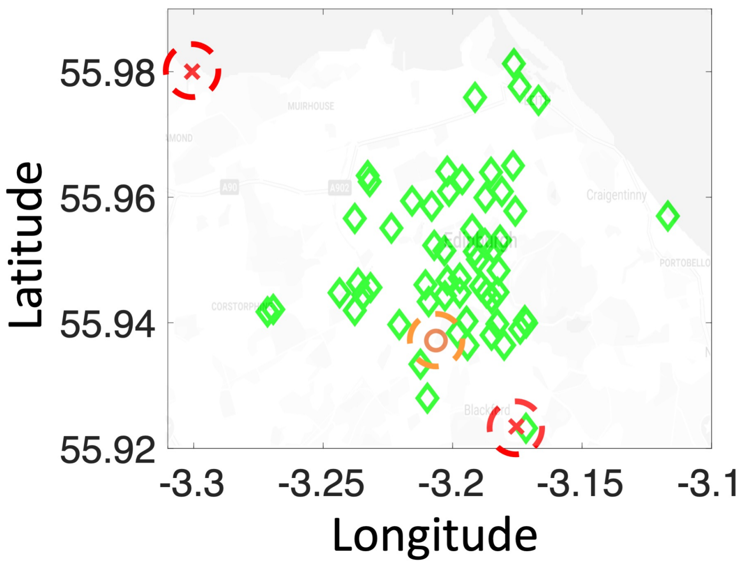
- Within the next 3 hours either $\varphi_{wait}(\tau)$ or $\varphi_{walk}(d)$ is True

$$\varphi_{wait}(\tau) = F_{[0,\tau]}(B \geq 1) \wedge F_{[0,\tau]}(S \geq 1),$$
$$\varphi_{walk}(d) = \diamond_{[0,d]}\, (B \geq 1) \wedge \diamond_{[0,d]}\, (S \geq 1)$$

- Locations with large τ: long wait times
- Locations with large d: far from stations with Bikes/Slots availability

BSS data from the city of Edinburgh





$$\varphi_{red} = \neg G_{[0,3]}\big(\varphi_{wait}(17.09) \vee \varphi_{walk}(2100)\big) \wedge \neg G_{[0,3]}(\varphi_{wait}(50) \vee \varphi_{walk}(1000.98))$$

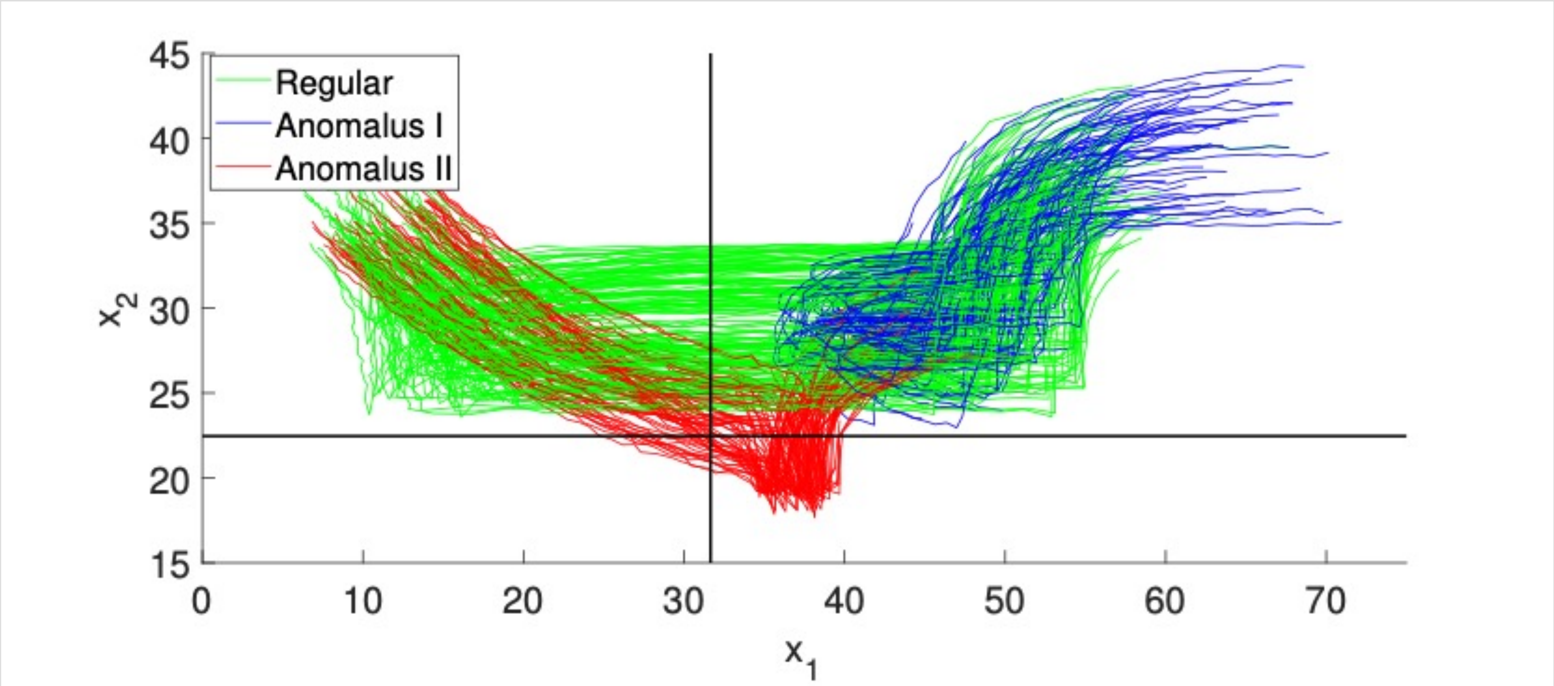
Results summary:

Case	$\|L\|$	$\|W\|$	$runtime(secs)$	numC	$\|\varphi_{cluster}\|$
COVID-19	235	427	813.65	3	$3.\|\varphi\| + 4$
BSS	61	91	681.78	3	$2.\|\varphi\| + 4$
Air Quality	107	60	136.02	8	$5.\|\varphi\| + 7$
Food Court	20	35	78.24	8	$3.\|\varphi\| + 4$

In a nutshell:

- We proposed a technique to learn interpretable STREL formulas from spatio-temporal data
- We proposed a new method for creating a spatial model with a restrict number of edges that preserves connectivity of the spatial model.
- We leveraged robustness of STREL combined with bisection search to extract features for spatiotemporal time-series clustering.
- We applied AHC on the extracted features followed by a DT based approach to learn an interpretable STREL formula for each cluster
- The results show that our method performs slower than ML approaches, but it is more interpretable

Learning STL classifiers



Goal: learning a specification/ classifier as a temporal logic formula to discriminate as much as possible between bad and good behaviours

Advantages: explicability, easy to build monitors

Application: anomaly detection, specification synthesis

Methodology

- *Single-level* variant: learning formula structure and parameter using Context Free Grammar Genetic Programming (CFGGP)

- *Bi-level* variant:
  - learning formula structure CFGGP
  - learn parameters of the formula using by **Bayesian Optimisation**

A fitness function f measures the quality of candidate solutions and depends on the kind of problem at hand (two-classes, one-class)

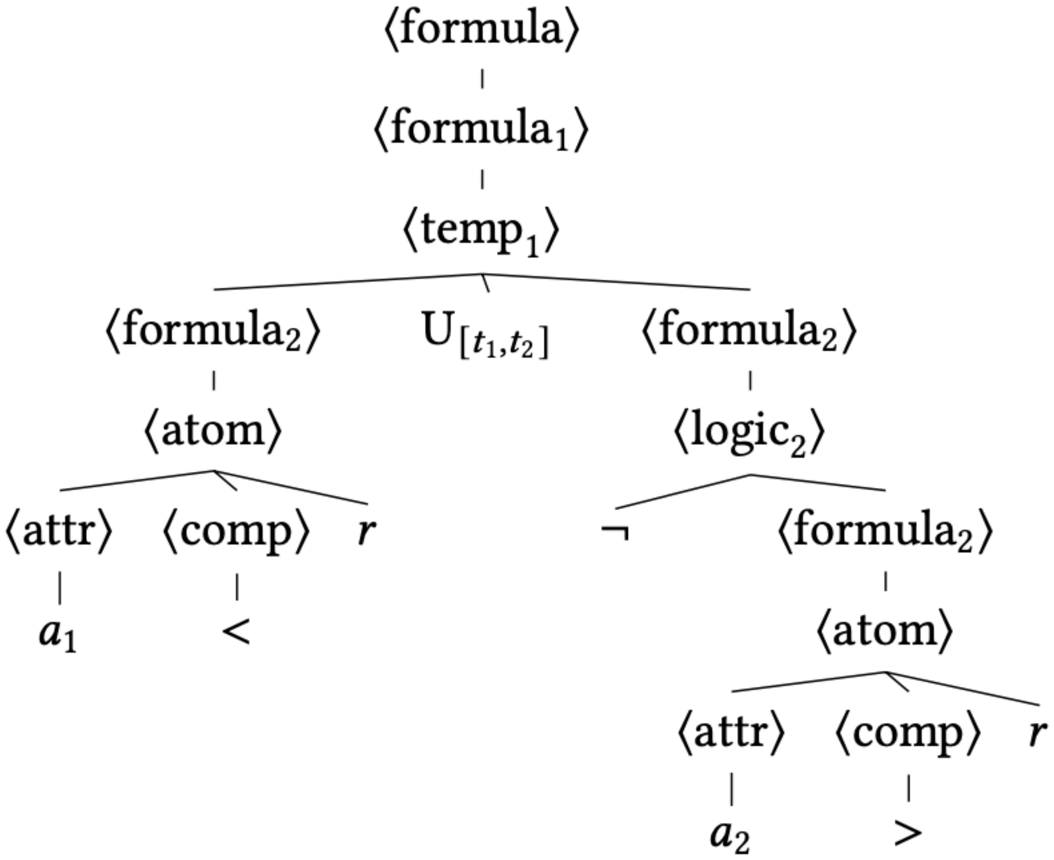
Evolutionary algorithm

- It builds the offspring population P'
- It merges the parent and offspring populations
- It shrinks the resulting new population P

```
function evolve():
    P ← initialize(𝒢, n_pop)
    foreach i ∈ {1, ..., n_gen} do
        P' ← ∅
        while |P'| ≤ n_pop do
            i ← 0
            repeat
                if ~ U(0, 1) ≤ p_xover then
                    (φ_{p,1}, f_{p,1}) ← select(P)
                    (φ_{p,2}, f_{p,2}) ← select(P)
                    φ_c ← crossover(φ_{p,1}, φ_{p,2}; 𝒢)
                else
                    (φ_p, f_p) ← select(P)
                    φ_c ← mutate(φ_p; 𝒢)
                end
                i ← i + 1
            until (φ_c ∉ P ∪ P') ∧ (i ≤ n_atts)
            P' ← P' ∪ {(φ_c, f_opt(φ_c; ℒ))}
        end
        P ← P ∪ P'
        while |P| ≥ n_pop do
            P ← P \ {worst(P)}
        end
    end
    return best(P)
end
```

Building the populations

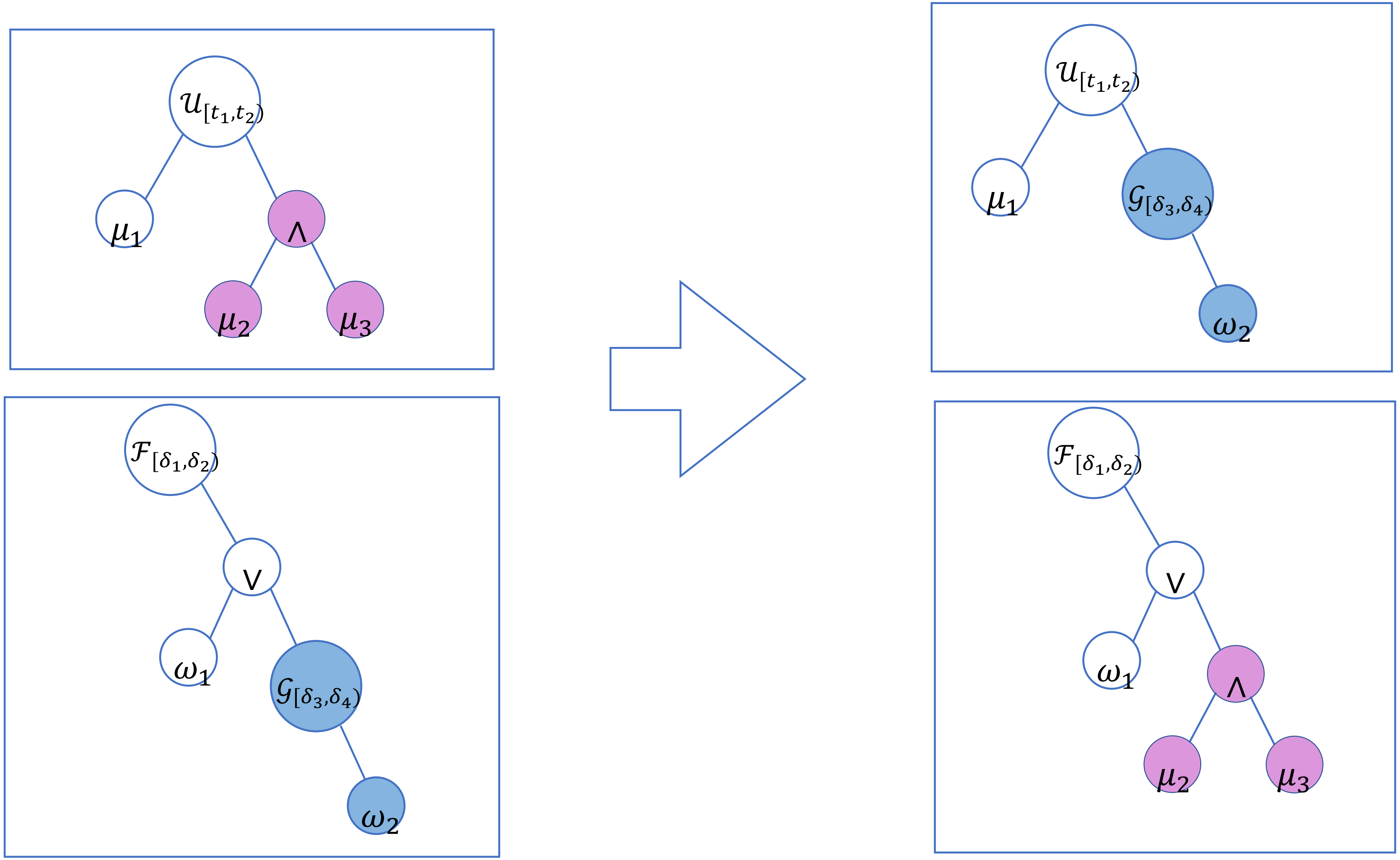
- Candidate formulas are represented as derivation trees of a grammar



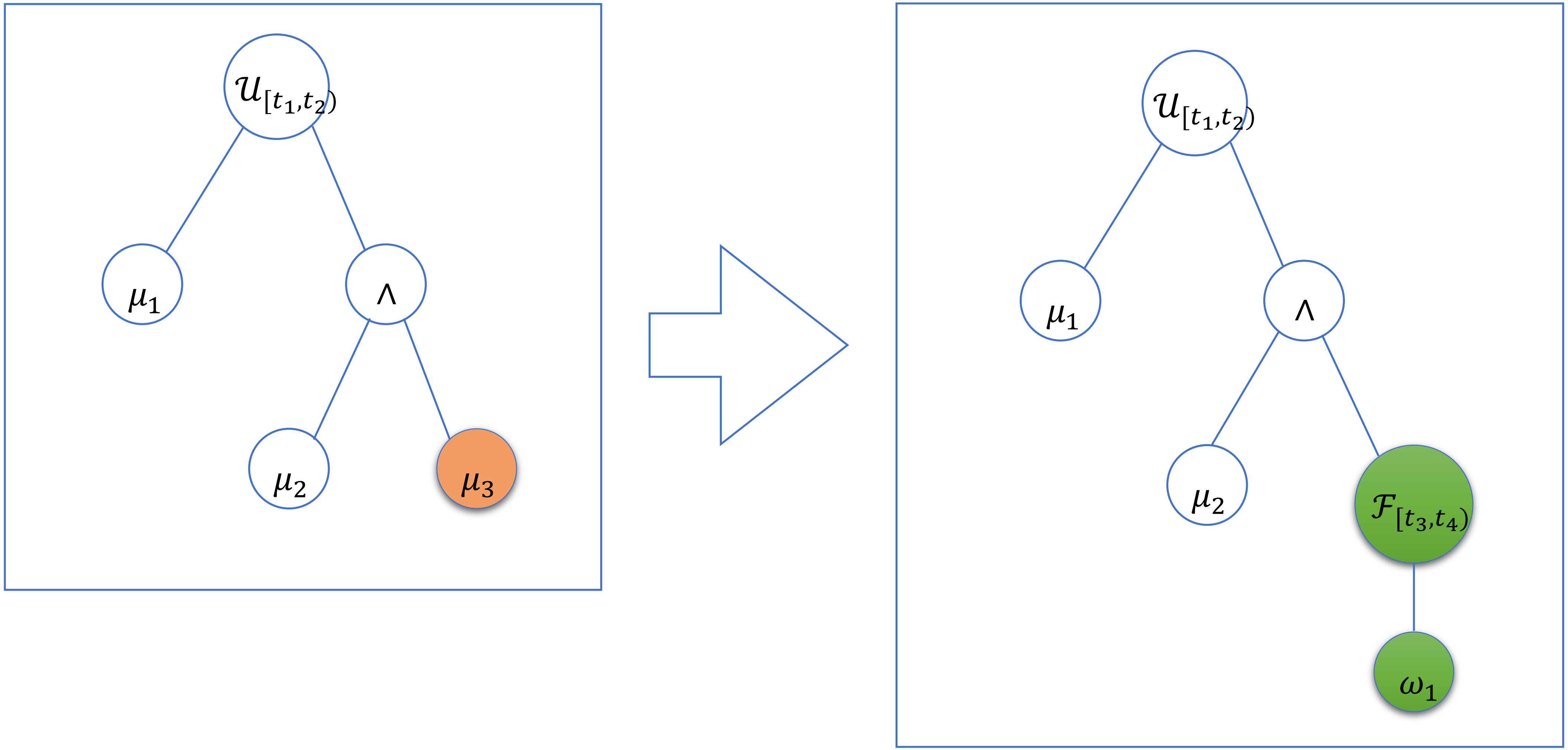
Context Free Grammar

$$
\begin{aligned}
\langle\text{formula}\rangle &::= \langle\text{formula}_1\rangle \\
\langle\text{formula}_i\rangle &::= \begin{cases} \langle\text{atom}\rangle \mid \langle\text{logic}_i\rangle \mid \langle\text{temp}_1\rangle & \text{if } i < i_{\text{max}} \\ \langle\text{atom}\rangle \mid \langle\text{logic}_i\rangle & \text{otherwise} \end{cases} \\
\langle\text{logic}_i\rangle &::= \neg\langle\text{formula}_i\rangle \mid \langle\text{formula}_i\rangle \wedge \langle\text{formula}_i\rangle \\
\langle\text{temp}_i\rangle &::= \langle\text{formula}_{i+1}\rangle \mathrm{U}_{\langle\text{interval}\rangle} \langle\text{formula}_{i+1}\rangle \mid \\
&\quad\ \ \mathrm{G}_{\langle\text{interval}\rangle}\langle\text{formula}_{i+1}\rangle \mid \mathrm{F}_{\langle\text{interval}\rangle}\langle\text{formula}_{i+1}\rangle \\
\langle\text{interval}\rangle &::= [\langle\text{num}\rangle, \langle\text{num}\rangle] \\
\langle\text{atom}\rangle &::= \langle\text{attr}\rangle\langle\text{comp}\rangle\langle\text{num}\rangle \\
\langle\text{attr}\rangle &::= a_1 \mid a_2 \mid \ldots \mid a_{|A|} \\
\langle\text{comp}\rangle &::= \ < \ \mid \ > \\
\langle\text{num}\rangle &::= \langle\text{digit}\rangle\langle\text{digit}\rangle \\
\langle\text{digit}\rangle &::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{aligned}
$$

Crossover operator



Mutation operator



Learning the Parameters

Problem

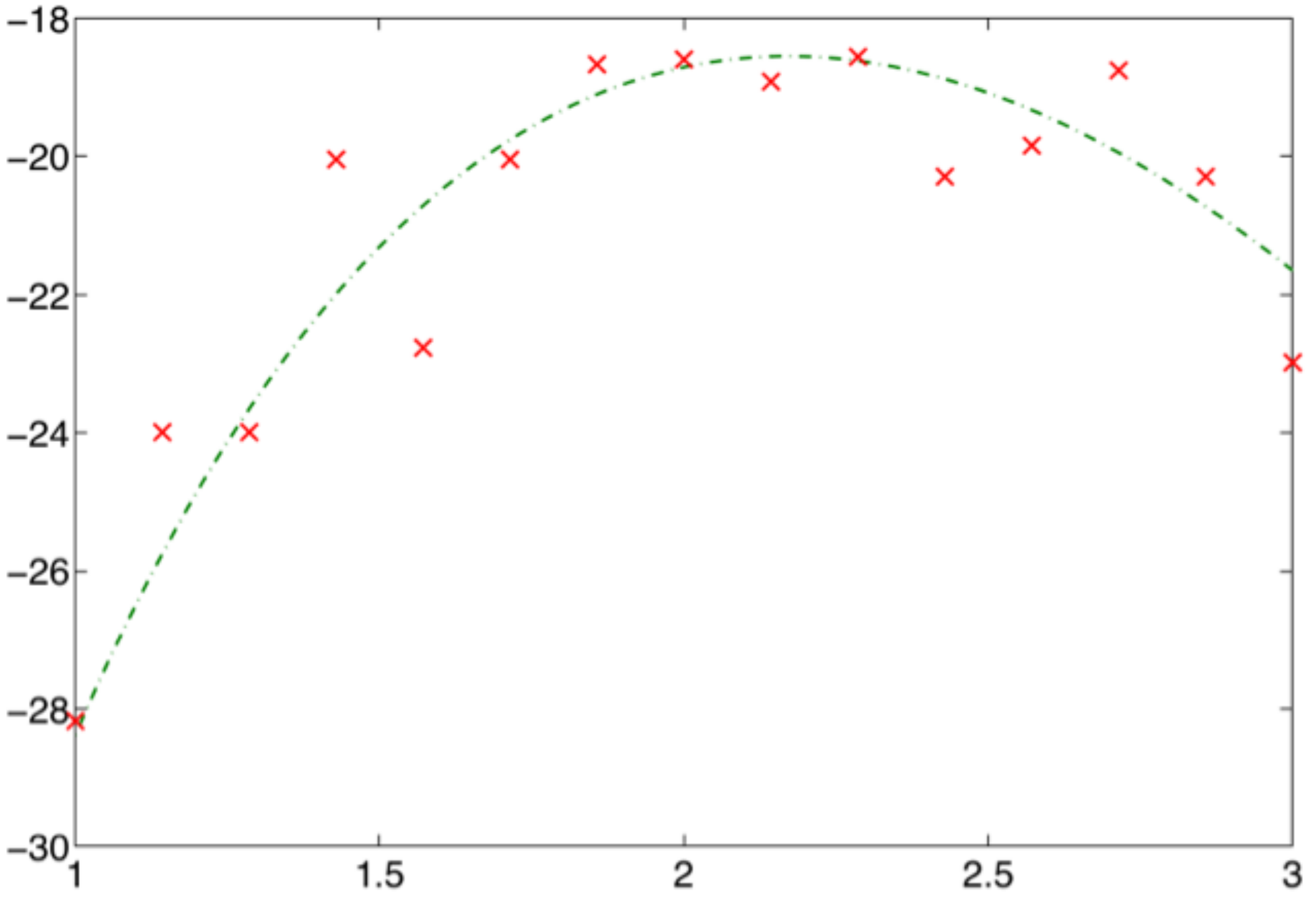
Given a PSTL formula ϕ, a parameter space K, find Θ^* that maximises the discrimination function $f_{opt}(\varphi_\Theta)$

Methodology

1. Sample $\{(\theta_{(i)}, y_{(i)}),\ i = 1,\ldots,n\}$
2. Emulate (**GP Regression**): $G[R_\phi] \sim GP(\mu,k)$
3. Optimize the emulation via **GP-UCB algorithm**, new $\theta_{(n+1)}$

(1) The G($\phi_\circ$) Computation

Collection of the training set $\{(\theta^{(i)}, y^{(i)}),\ i = 1,\dots,m\}$ for parameters values θ.

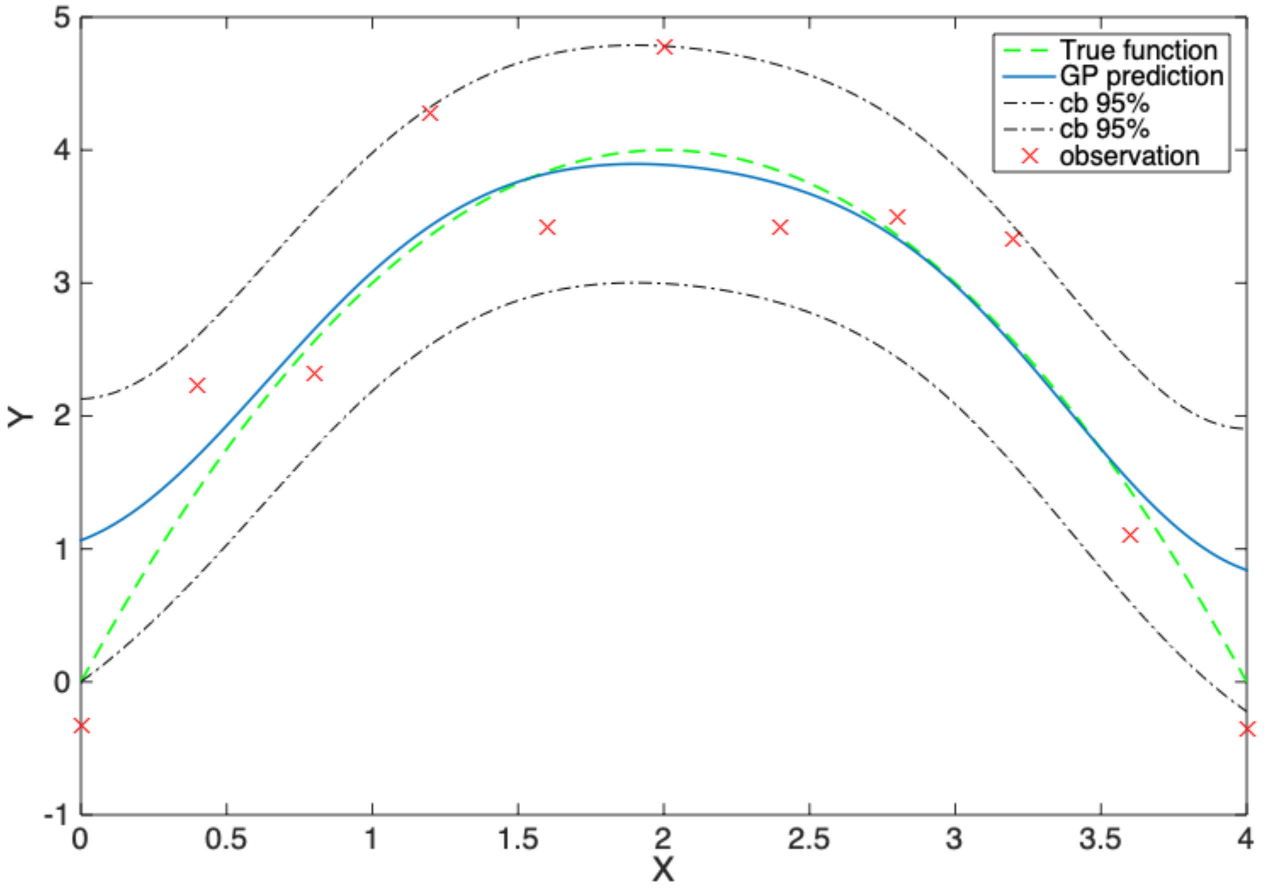


(2) The GP Regression

We have noisy observations y of the function value distributed around an unknown true value $f(\theta)$ with spherical Gaussian noise

(2) The GP Regression

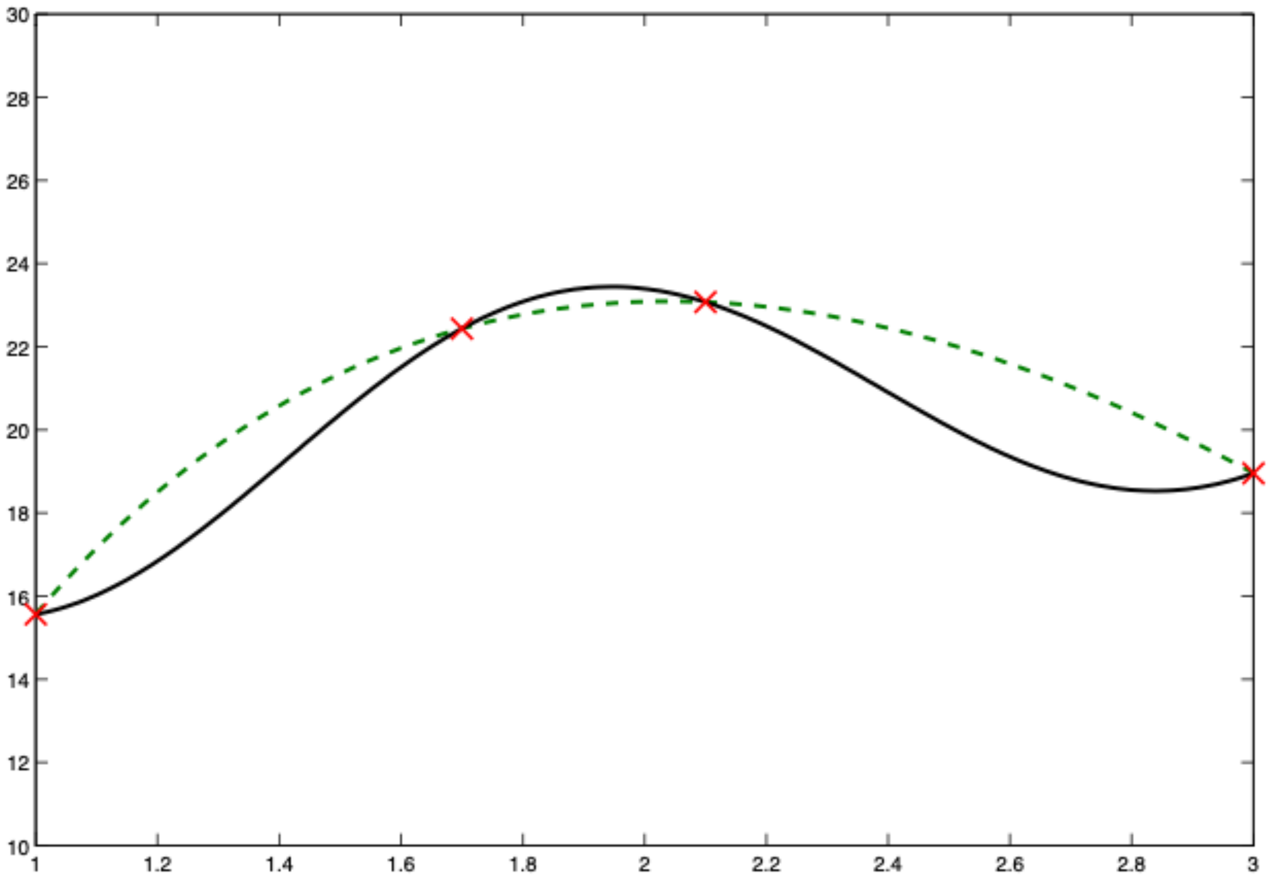
We have noisy observations y of the function value distributed around an unknown true value f (θ) with spherical Gaussian noise



(3) The GP-UCB Algorithm

Balance Exploration and Exploitation: we maximise the

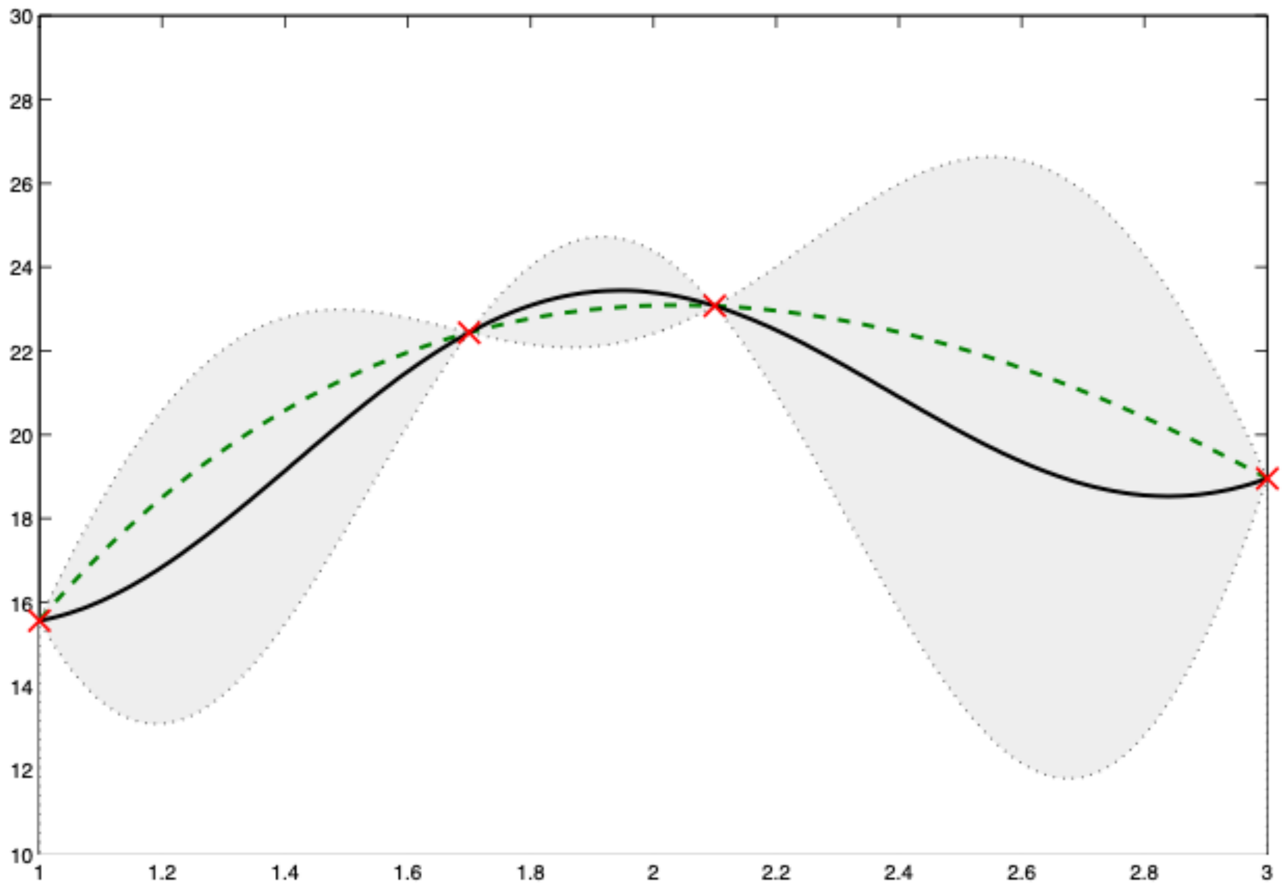
95% upper quantile of the distribution: $\theta_{t+1} = argmax_\theta[\mu^*(\theta) + \beta_t\sqrt{k^*(\theta,\theta)}]$



(3) The GP-UCB Algorithm

Balance Exploration and Exploitation: we maximise the

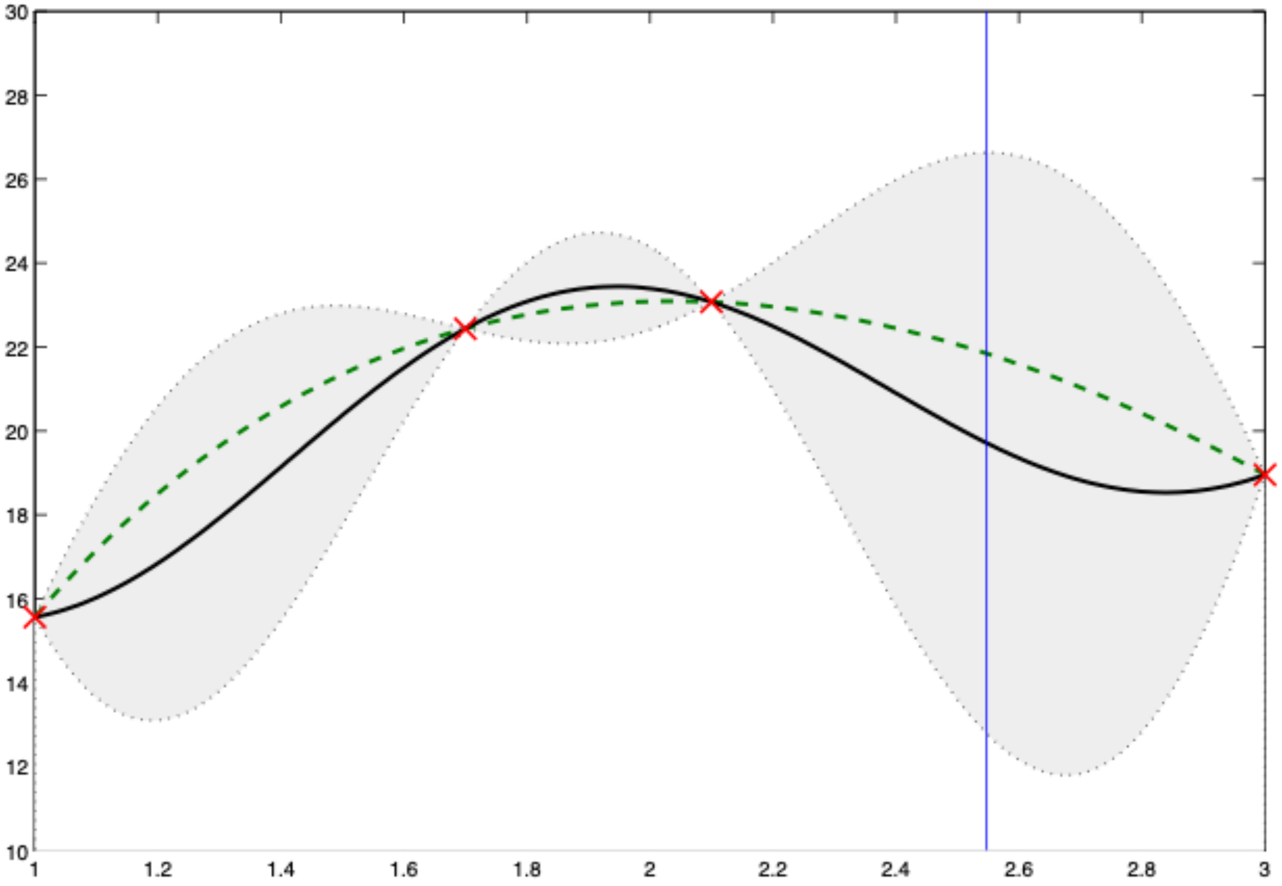
95% upper quantile of the distribution: $\theta_{t+1} = argmax_{\theta}[\mu^*(\theta) + \beta_t\sqrt{k^*(\theta,\theta)}]$



(3) The GP-UCB Algorithm

Balance Exploration and Exploitation: we maximise the

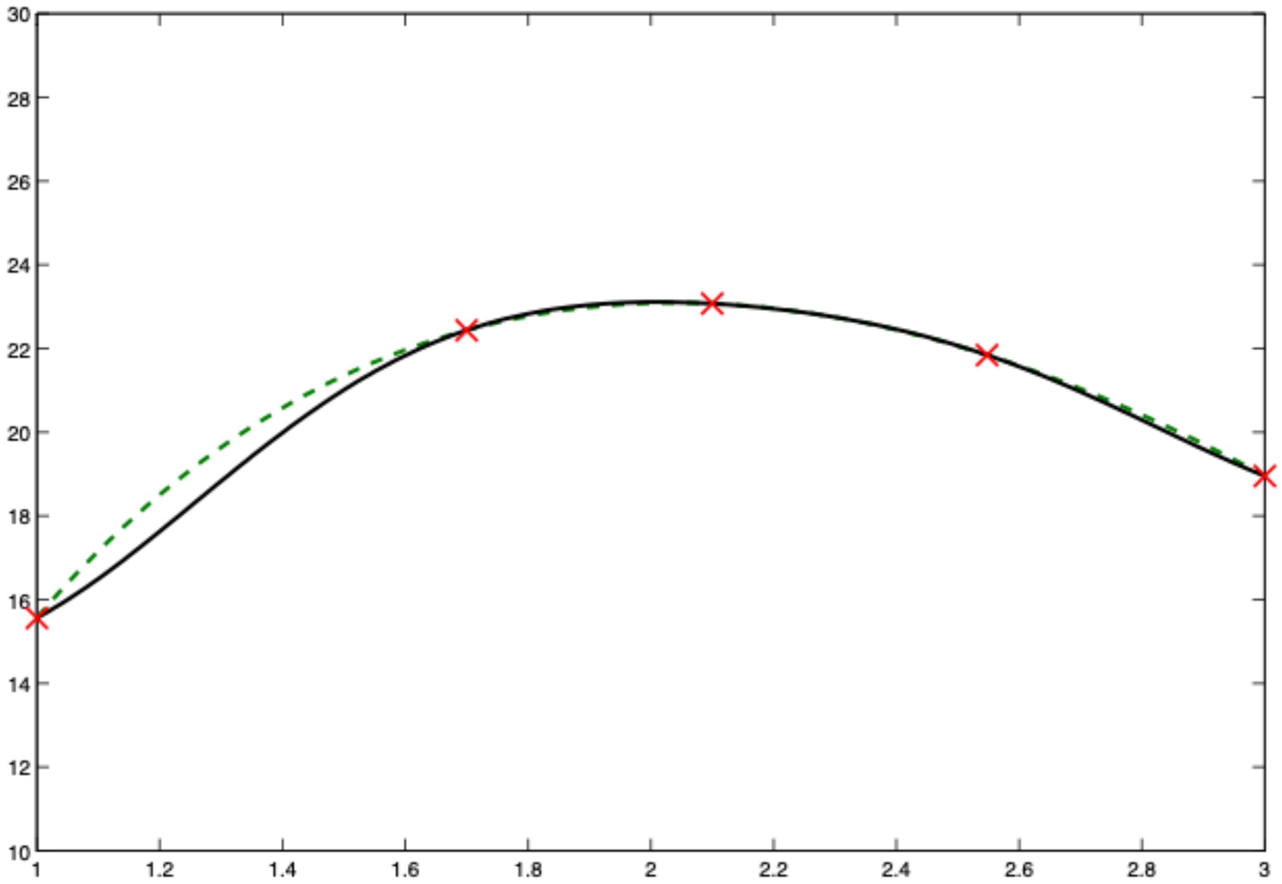
95% upper quantile of the distribution: $\theta_{t+1} = argmax_{\theta}[\mu^*(\theta) + \beta_t \sqrt{k^*(\theta,\theta)}]$



(3) The GP-UCB Algorithm

Balance Exploration and Exploitation: we maximise the

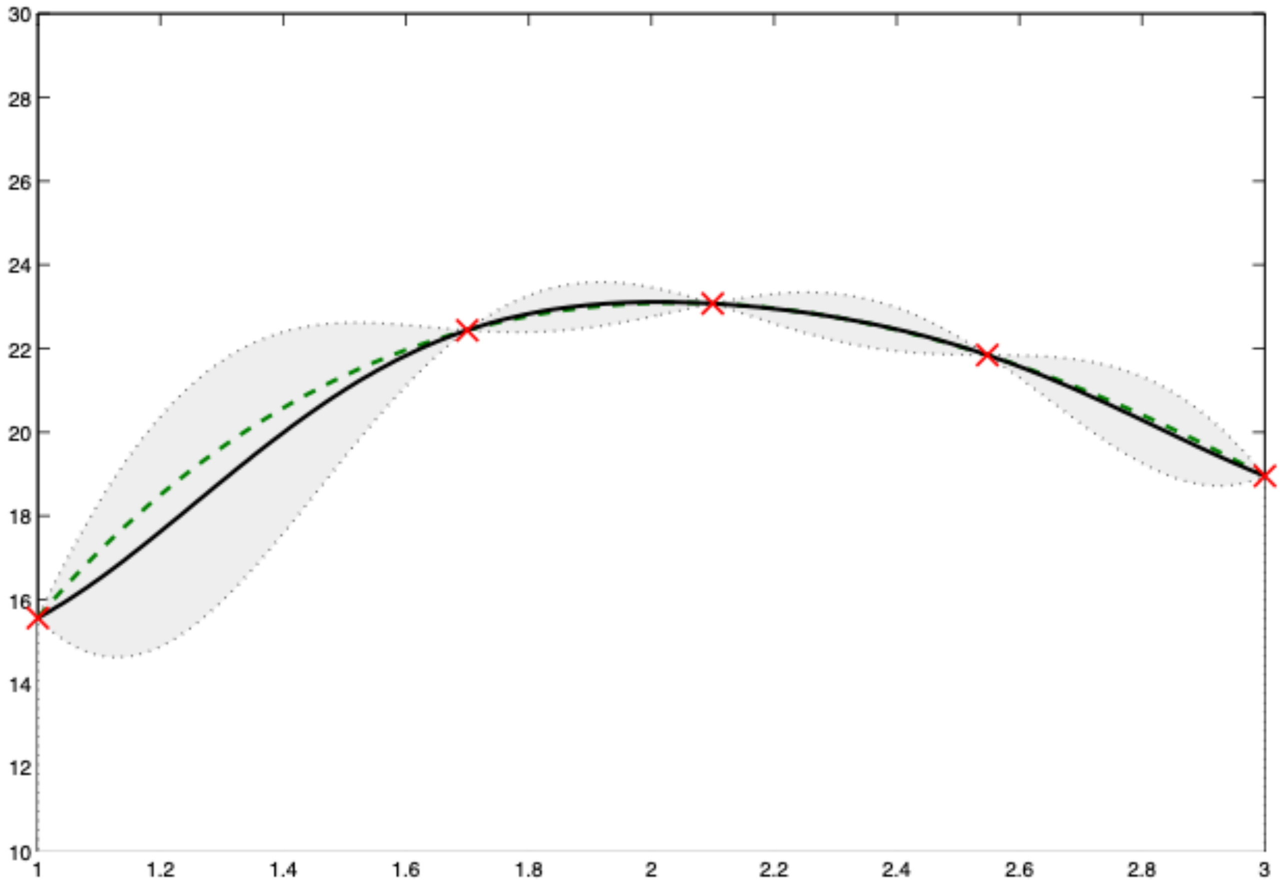
95% upper quantile of the distribution: $\theta_{t+1} = argmax_{\theta}[\mu^*(\theta) + \beta_t \sqrt{k^*(\theta,\theta)}]$



(3) The GP-UCB Algorithm

Balance Exploration and Exploitation: we maximise the

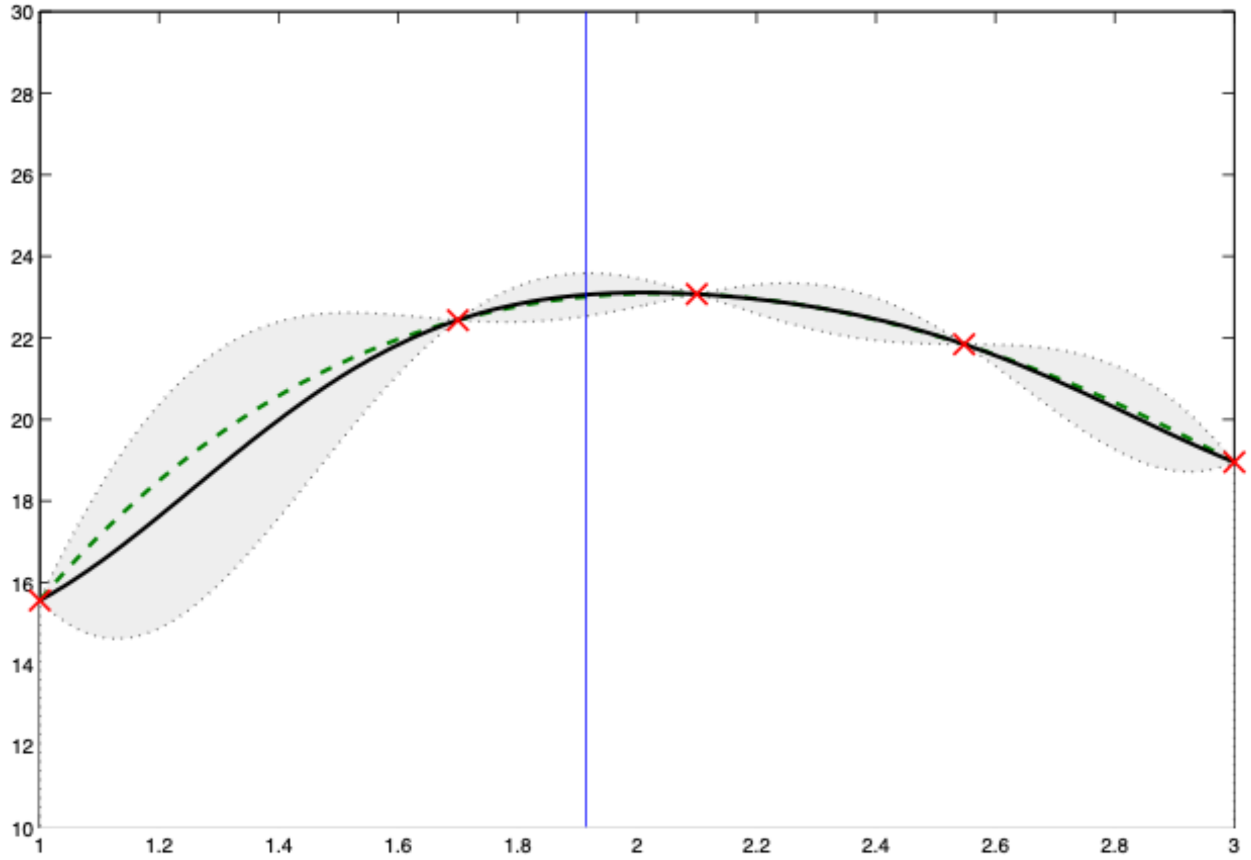
95% upper quantile of the distribution: $\theta_{t+1} = argmax_{\theta}[\mu^*(\theta) + \beta_t \sqrt{k^*(\theta,\theta)}]$



(3) The GP-UCB Algorithm

Balance Exploration and Exploitation: we maximise the

95% upper quantile of the distribution: $\theta_{t+1} = argmax_\theta[\mu^*(\theta) + \beta_t\sqrt{k^*(\theta,\theta)}]$



Fitness Function for the two-classes problem

$$f(\varphi; X_{\mathcal{L}}^{+}, X_{\mathcal{L}}^{-}) = -\frac{\mu_{\varphi,X_{\mathcal{L}}^{+}} - \mu_{\varphi,X_{\mathcal{L}}^{-}}}{\sigma_{\varphi,X_{\mathcal{L}}^{+}} + \sigma_{\varphi,X_{\mathcal{L}}^{-}}}$$

$$\mu_{\varphi,X} = \frac{1}{|X|} \sum_{\boldsymbol{x} \in X} \rho(\varphi, \boldsymbol{x})$$

$$\sigma_{\varphi,X} = \sqrt{\frac{1}{|X|} \sum_{\boldsymbol{x} \in X} \left(\rho(\varphi, \boldsymbol{x}) - \mu_{\varphi,X}\right)^2}$$

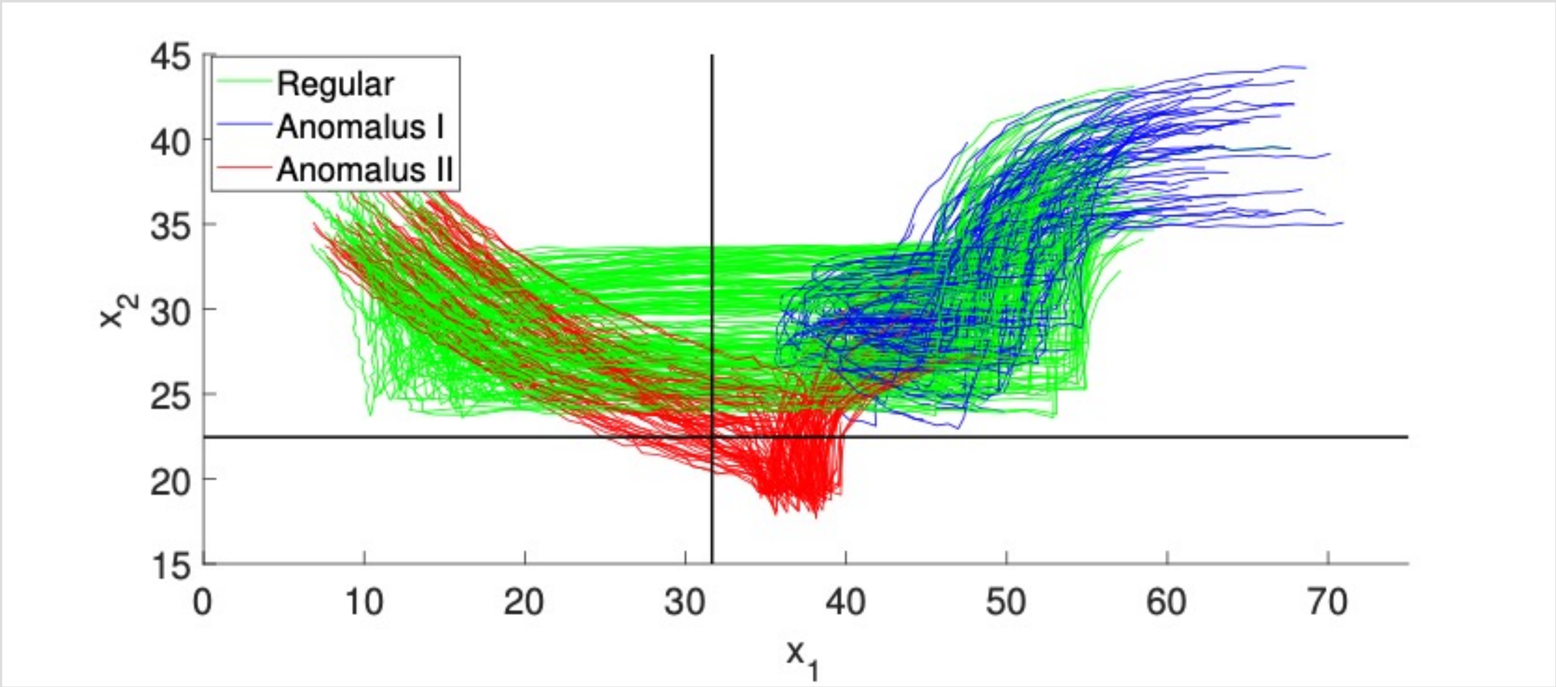
Fitness Function for the one-class problem

$$f(\varphi; X^+_{\mathcal{L}}) = \alpha \frac{1}{\left|X^+_{\mathcal{L}}\right|} \left|\{\boldsymbol{x} \in X^+_{\mathcal{L}} : \boldsymbol{x} \not\models \varphi\}\right| + \frac{1}{\sigma'_{\varphi, X^+_{\mathcal{L}}} \left|X^+_{\mathcal{L}}\right|} \sum_{\boldsymbol{x} \in X^+_{\mathcal{L}}} |\rho(\varphi, \boldsymbol{x})|$$

$$\sigma'_{\varphi, X} = \sqrt{\frac{1}{|X|} \sum_{\boldsymbol{x} \in X} \left(|\rho(\varphi, \boldsymbol{x})| - \frac{1}{|X|} \sum_{\boldsymbol{x} \in X} |\rho(\varphi, \boldsymbol{x})| \right)^2}$$

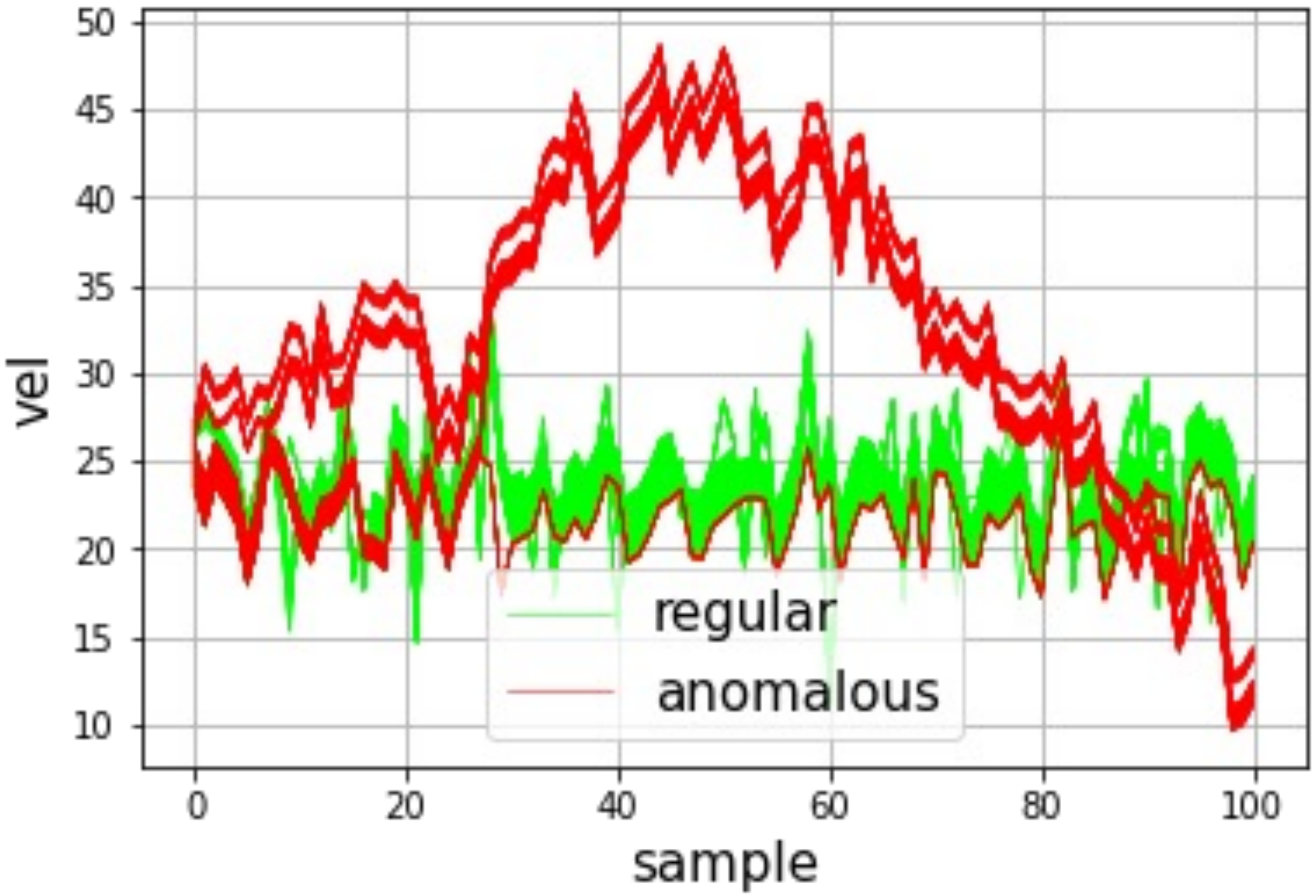
Maritime Surveillance

Synthetic dataset of naval surveillance of 2-dimensional coordinates traces of vessels behaviours.



$$\phi = ((x_2 > 22.46)\,\mathcal{U}_{[49,287]}\,(x_1 \leq 31.65))$$

Train Cruise

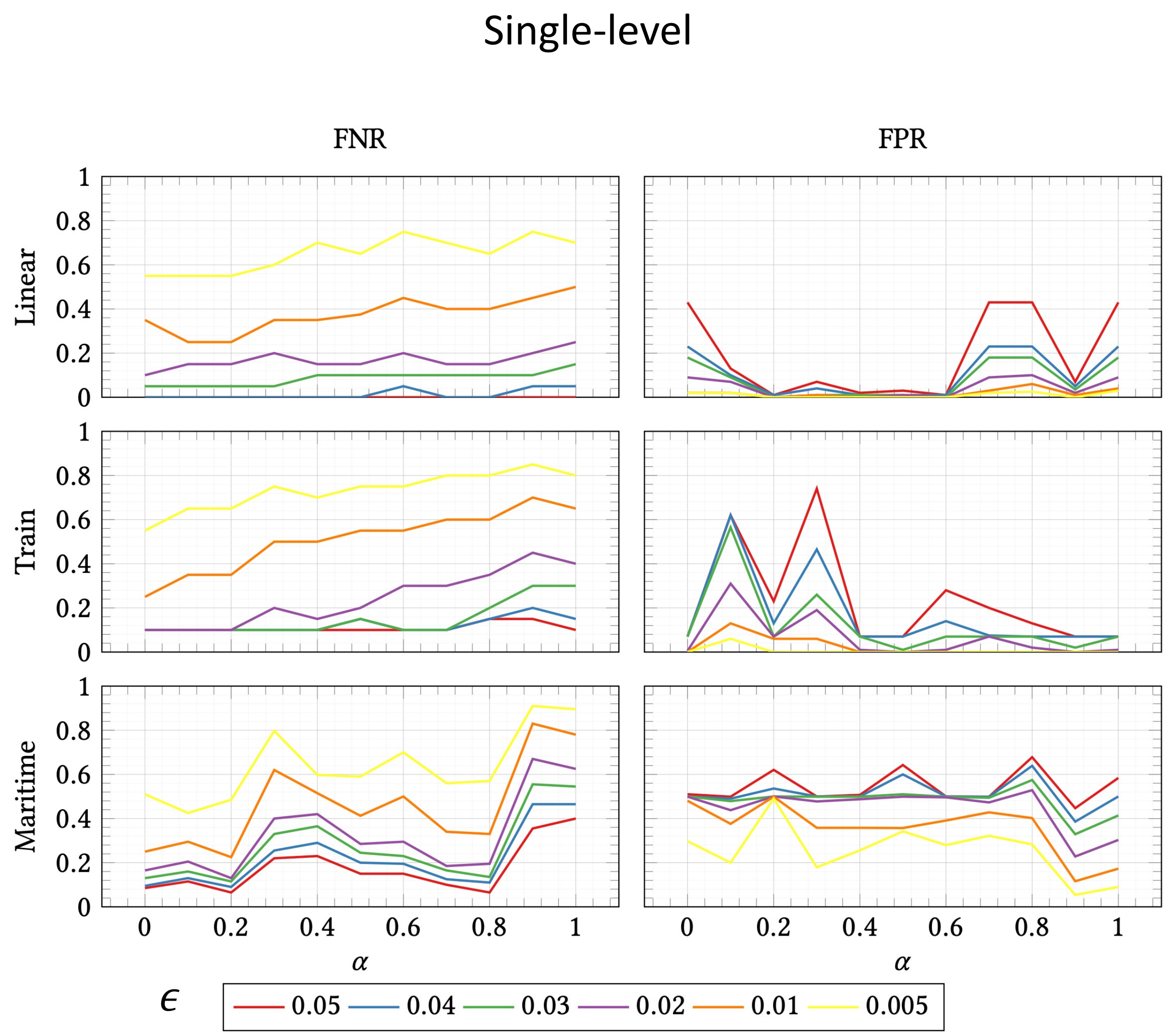


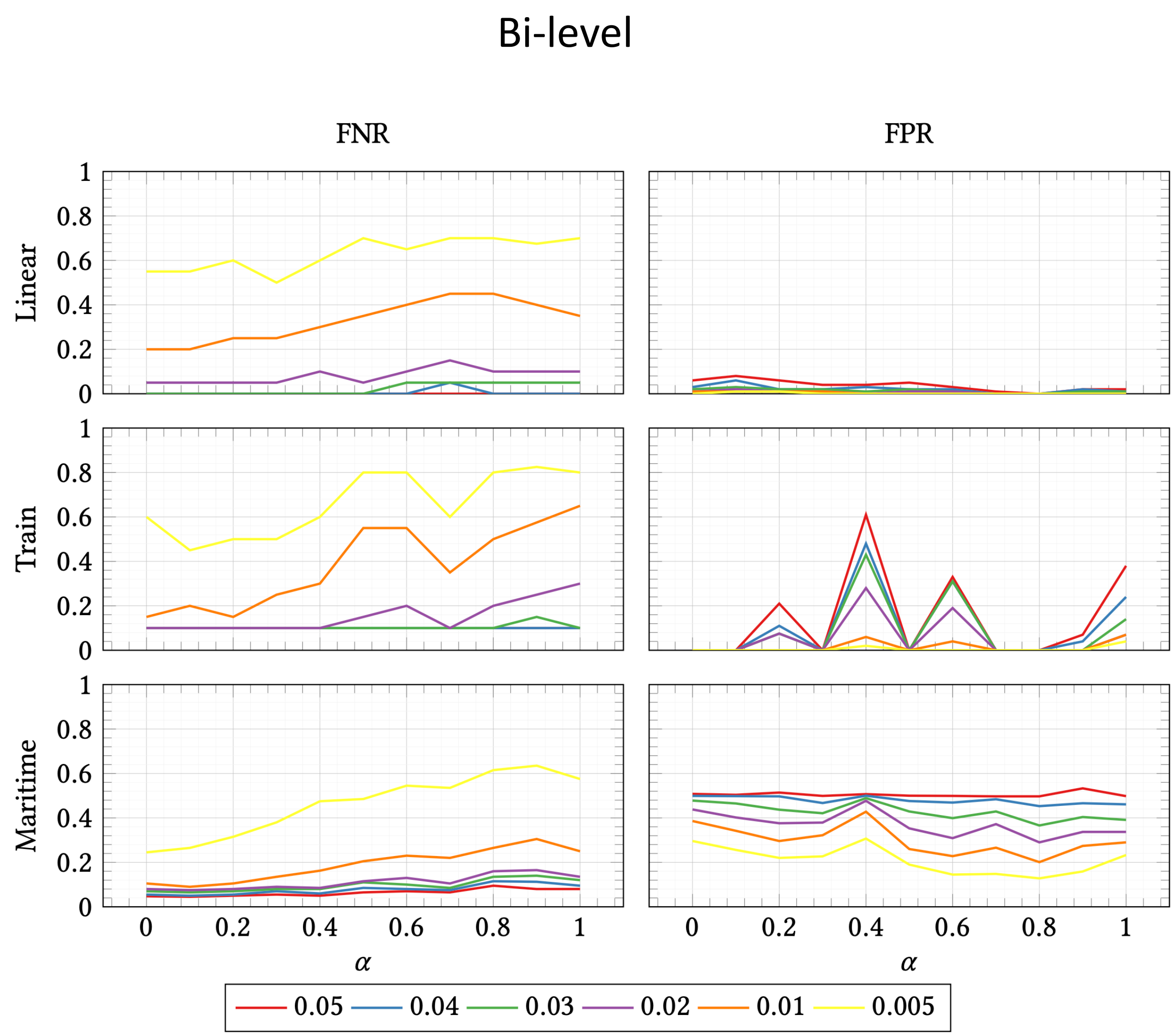
$$(\mathrm{F}_{[22,40]}(\mathtt{vel} > 24.48)) \wedge (\mathrm{F}_{[46,49]}(19.00 < \mathtt{vel} < 26.44))$$

Results (supervised learning)

Dataset	Algorithm	FNR	FPR	MCR	Time
Maritime	BUSTLE (single-level)	0.00	0.00	0.00	109
	BUSTLE (bi-level)	0.00	0.00	0.00	1477
	[23]	0.00	0.00	0.00	N/A
	[22]	0.05	0.02	0.04	73
	[6]	N/A	N/A	0.02	140
Linear	BUSTLE (single-level)	0.00	0.00	0.00	15
	BUSTLE (bi-level)	0.00	0.00	0.00	112
	[23]	0.01	0.01	0.01	N/A
	[22]	N/A	N/A	0.02	39
Train	BUSTLE (single-level)	0.03	0.05	0.04	26
	BUSTLE (bi-level)	0.00	0.03	0.02	523
	[23]	0.07	0.32	0.19	N/A
	[22]	N/A	N/A	0.02	32

Results (semi-supervised learning)





Bibliography

Mining Requirements:

- Jyo Deshmukh et al. Mining Requirements from Closed-loop Control Models (HSCC '13, IEEE Trans. On Computer Aided Design '15)

- Mohammadinejad et al., Mining Interpretable Spatio-temporal Logic Properties for Spatially Distributed Systems, ATVA, 2021.
- V-Chanlatte et al., Logical clustering and learning for time-series data, 2017.

- Bartocci, E., Bortolussi, L., Sanguinetti, G.: Data-driven statistical learning of temporal logic properties, FORMATS, 2014
- Bufo, S., Bartocci, E., Sanguinetti, G., Borelli, M., Lucangelo, U., Bortolussi, L.i,Temporal logic based monitoring of assisted ventilation in intensive care patients, ISoLA, 2014.
- Nenzi L., Silvetti S., Bartocci E., Bortolussi L. (2018) *A Robust Genetic Algorithm for Learning Temporal Specifications from Data.* QEST 2018. LNCS, vol 11024. Springer, Cham.
- Bombara, G et all, A Decision Tree Approach to Data Classification Using Signal Temporal Logic. In: Proc. of HSCC, 2016.