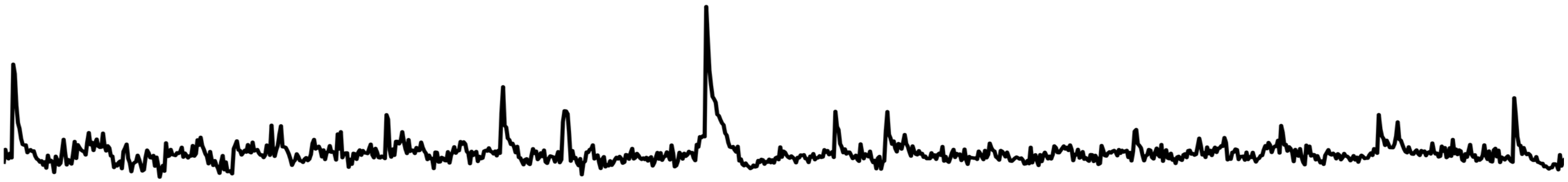
Differentiable Segmentation of Sequences

Erik Scharwächter, Jonathan Lennartz, Emmanuel Müller

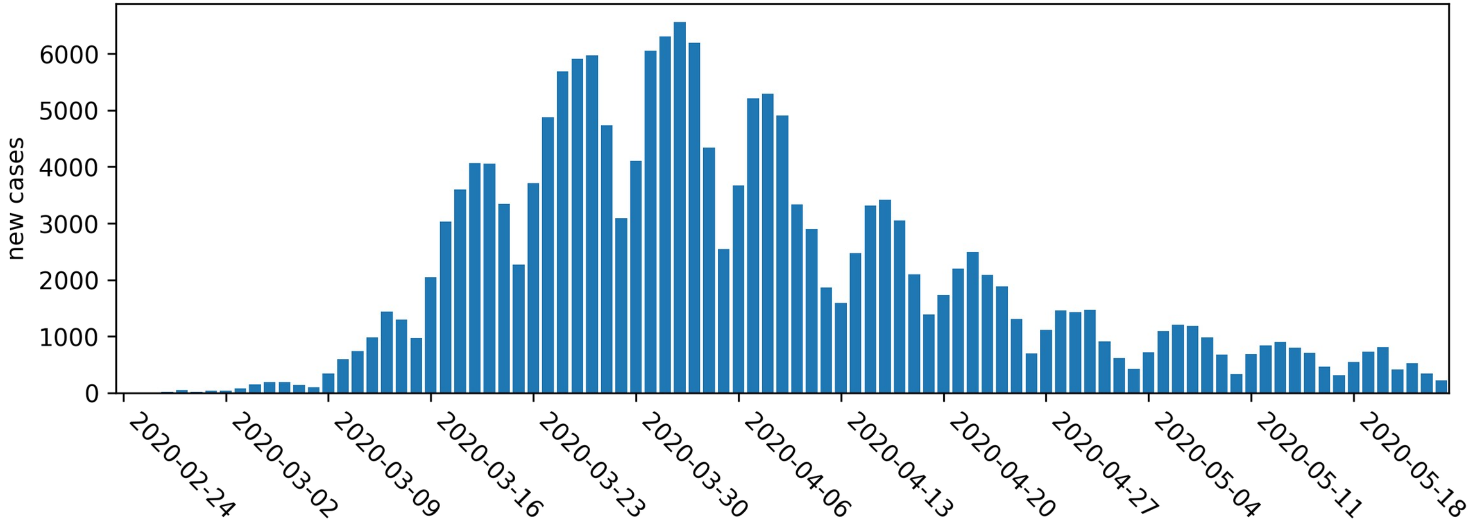
TU Dortmund University, Germany

@ ICLR 2021



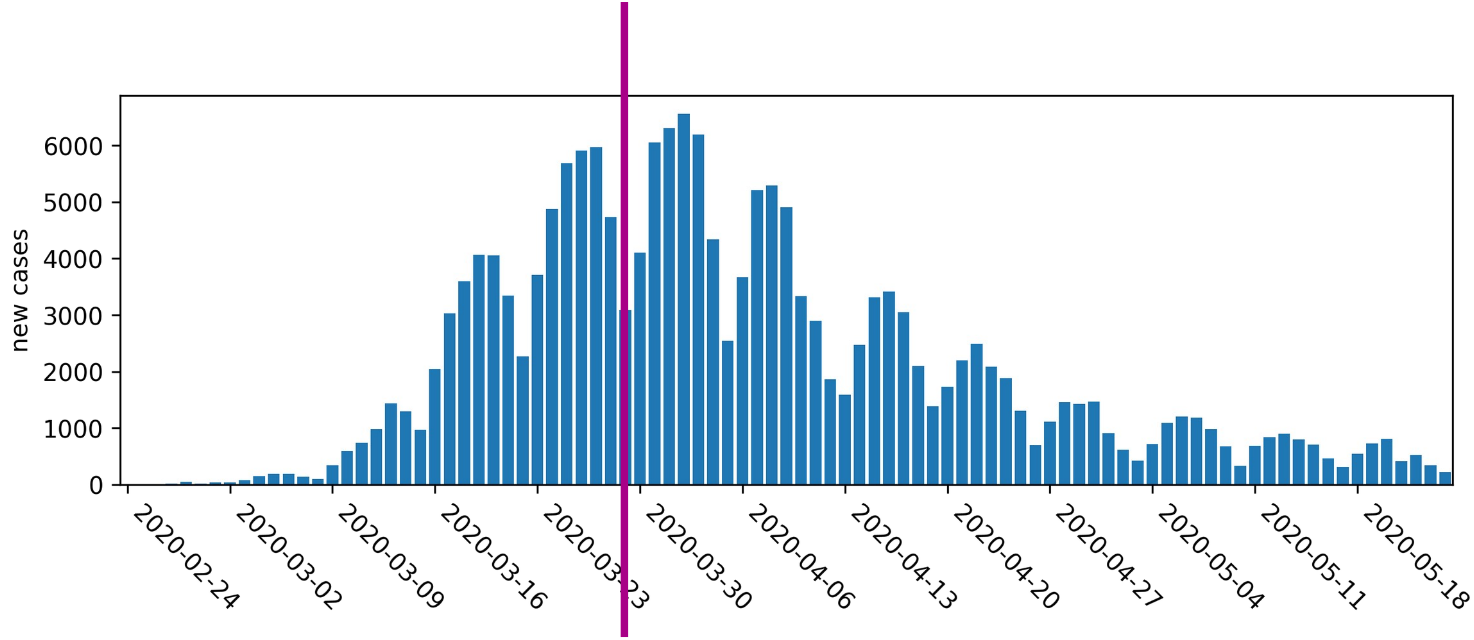
technische universität
dortmund

Example: COVID-19 in Germany (2020)



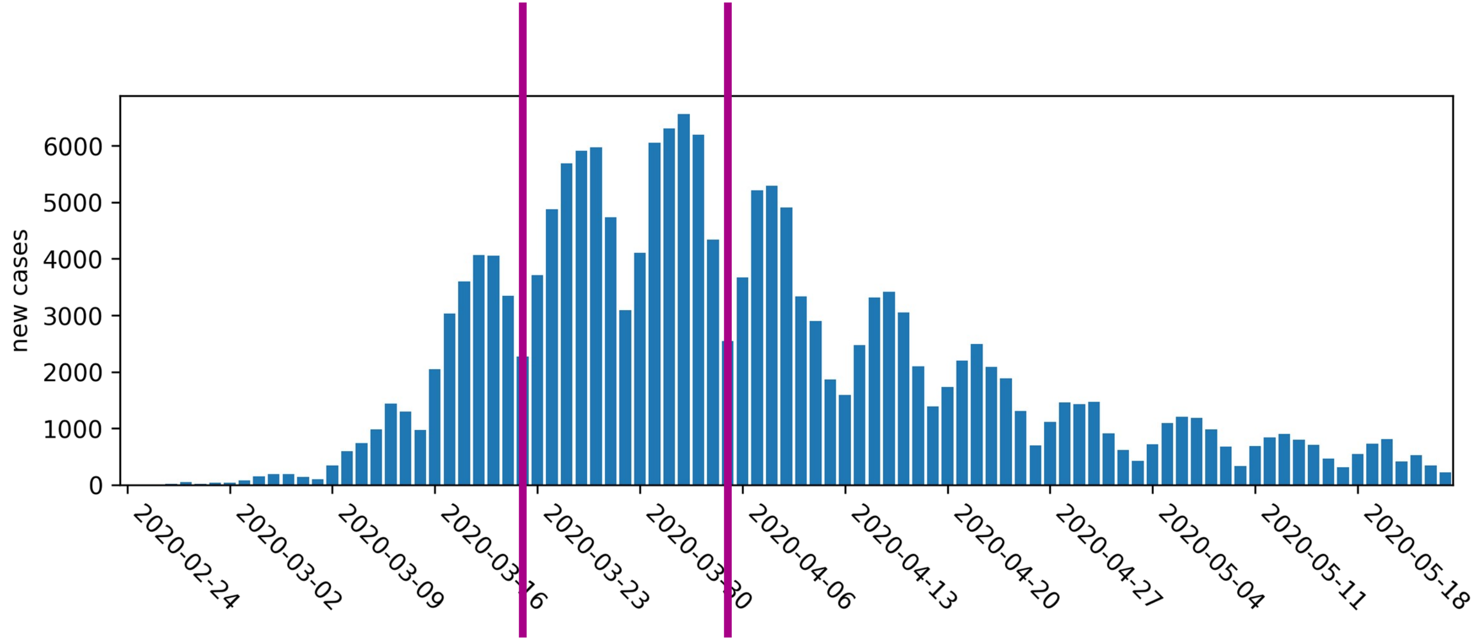
data source: Robert Koch Institute

Example: COVID-19 in Germany (2020)



data source: Robert Koch Institute

Example: COVID-19 in Germany (2020)



data source: Robert Koch Institute

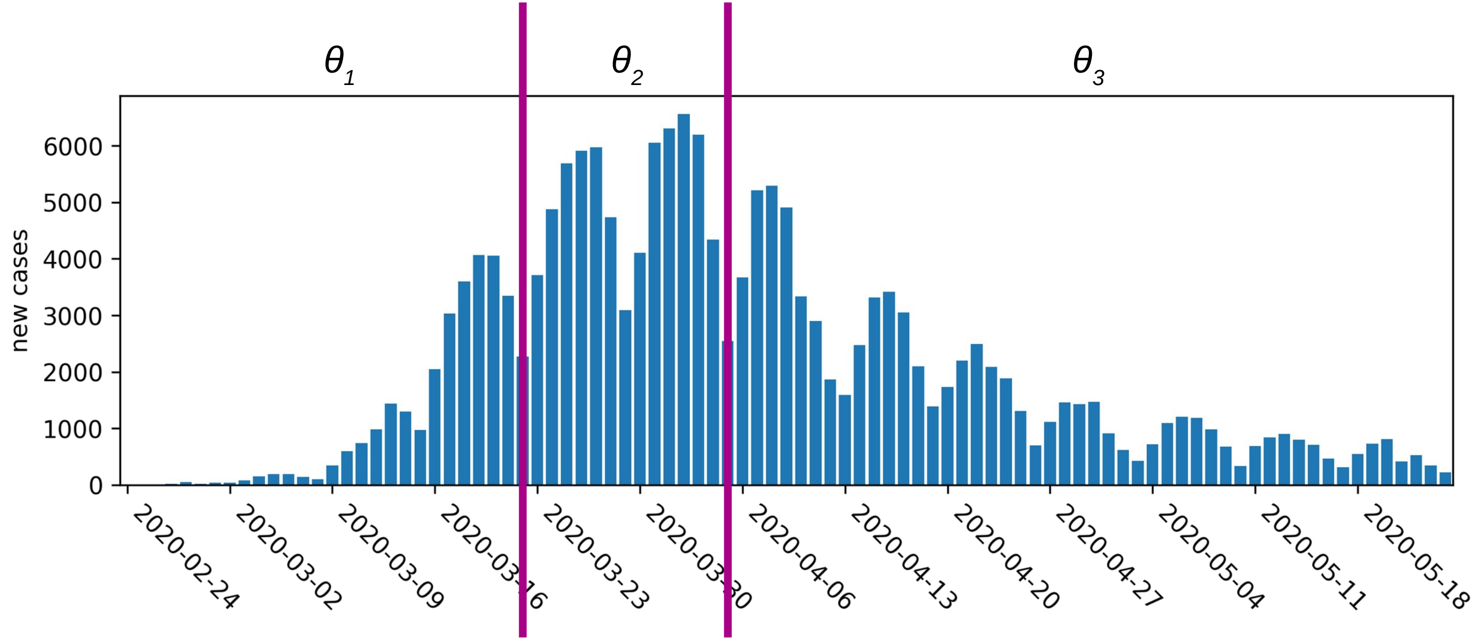
Problem: Fitting a segmented model

$$x_t \mid z_t \sim f_{\text{DGP}}\left(z_t, \theta_k\right)$$

observation — covariates — data generating process — parameters

if *t* belongs to segment *k*

find optimal segmentation and parameters



data source: Robert Koch Institute

Problem: Fitting a segmented model

$$x_t \mid z_t \sim f_{\text{DGP}}\left(z_t, \theta_k\right)$$

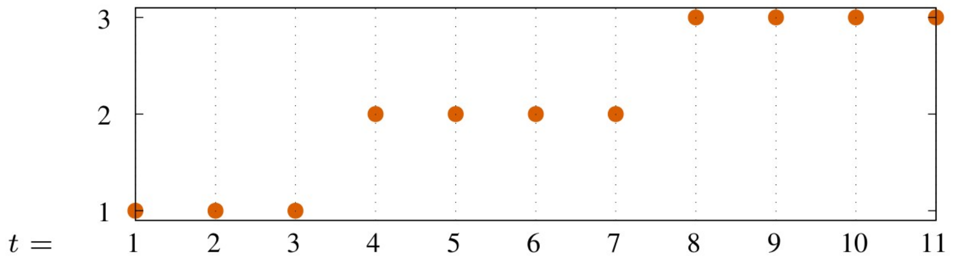
if t belongs to segment k

- x_t: observation
- z_t: covariates
- f_{DGP}: data generating process
- θ_k: parameters

find optimal segmentation and parameters

segmentation function

$$\zeta : \{1, \ldots, T\} \longrightarrow \{1, \ldots, K\}$$



order-preserving
boundary constraints

Problem: Fitting a segmented model

$x_t \mid z_t \sim f_{\text{DGP}}\left(z_t, \theta_k\right)$ if $\zeta(t) = k$

$\zeta : \{1, ..., T\} \longrightarrow \{1, ..., K\}$

find optimal segmentation and parameters

loss function

parameters

$$\min_{\zeta,\theta} \mathcal{L}(\zeta, \theta)$$

segmentation function

Existing approaches: Highly specialized algorithms

$$x_t \mid z_t \sim f_{\mathrm{DGP}}\left(z_t, \theta_k\right)$$ if $\zeta(t) = k$

$$\zeta : \{1, ..., T\} \longrightarrow \{1, ..., K\}$$

find optimal segmentation and parameters

discrete — continuous

$$\min_{\zeta,\theta} \mathcal{L}(\zeta,\theta) = \min_{\zeta} \min_{\theta} \mathcal{L}(\zeta,\theta)$$

grid search
dynamic programming
greedy approaches

Lerman, P. M. (1980). **Fitting Segmented Regression Models by Grid Search.** Journal of the Royal Statistical Society. Series C (Applied Statistics), 29(1), 77–84.

Hawkins, D. M. (1976). **Point Estimation of the Parameters of Piecewise Regression Models.** Journal of the Royal Statistical Society C, 25(1).

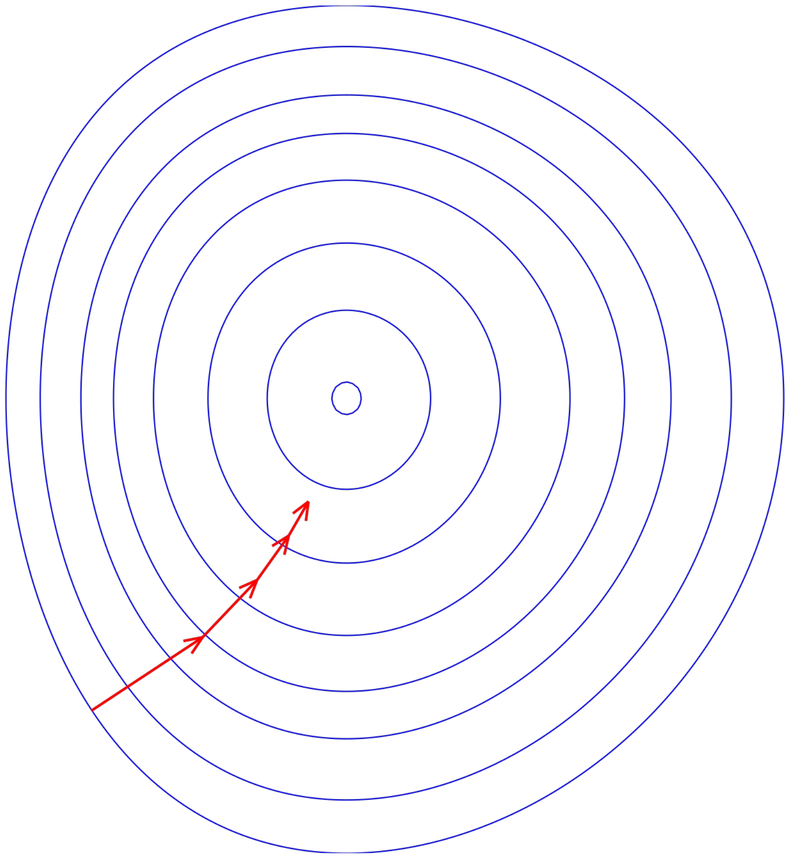
Acharya, J., Diakonikolas, I., Li, J., & Schmidt, L. (2016). **Fast algorithms for segmented regression.** In: ICML.

Our goal: Optimization with gradient descent

$x_t \mid z_t \sim f_{\mathrm{DGP}}\left(z_t, \theta_k\right)$ if $\zeta(t) = k$

$\zeta : \{1, ..., T\} \longrightarrow \{1, ..., K\}$

$$\min_{\zeta,\theta} \mathcal{L}(\zeta, \theta) =$$

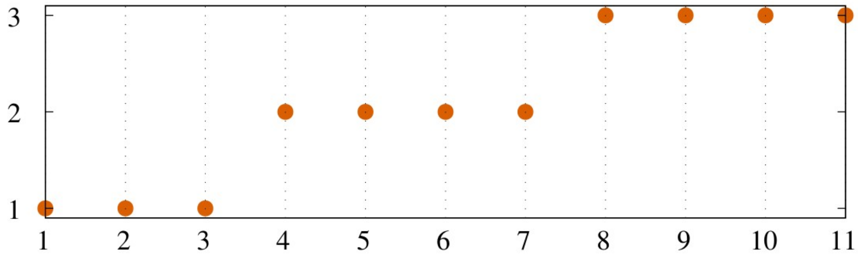


modelling flexibility
deep learning
algorithmic advances

Continuous relaxation

segmentation function

$$\zeta : \{1, ..., T\} \longrightarrow \{1, ..., K\}$$

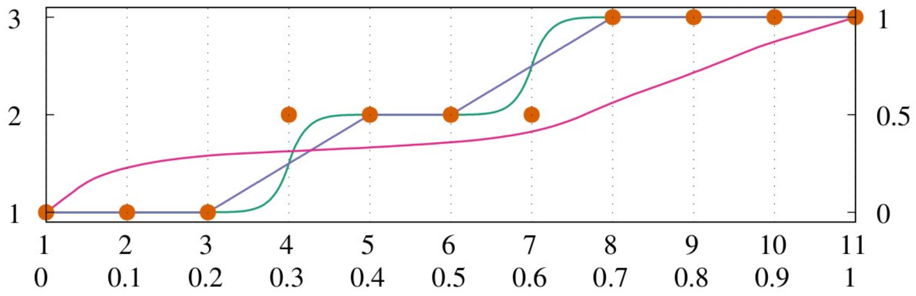


order-preserving
boundary constraints

Continuous relaxation

replace segmentation function
with warping function

$$\gamma : [0,1] \longrightarrow [0,1]$$

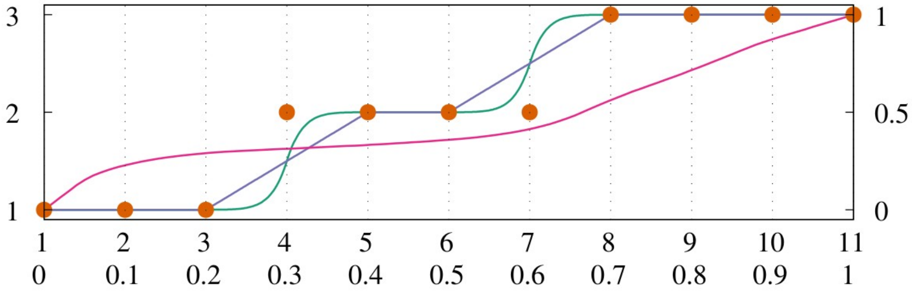


order-preserving
boundary constraints

Continuous relaxation

replace segmentation function with warping function

$$\gamma : [0,1] \longrightarrow [0,1]$$



order-preserving boundary constraints

novel family of warping functions

step functions based on the two-sided power distribution

$$\gamma_{\mathrm{TSP}}(u; \mu_1, \ldots, \mu_K)$$

unconstrained real parameters

$$\mu^{(i+1)} = \mu^{(i)} - \eta \frac{\partial \mathcal{L}}{\partial \mu}$$

Continuous relaxation

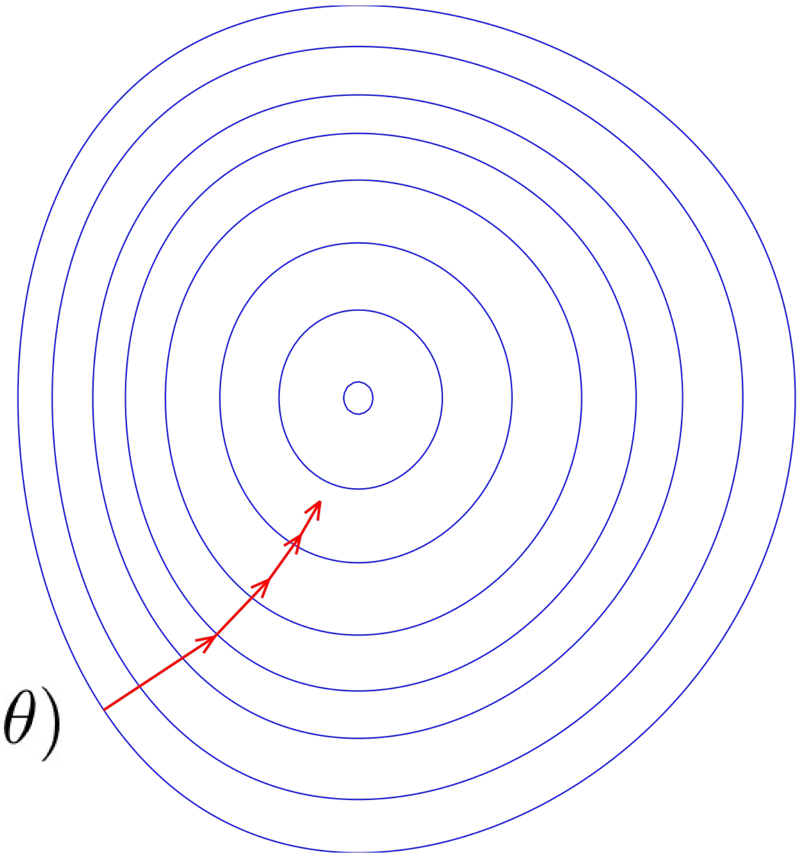
$$x_t \mid z_t \sim f_{\text{DGP}}\left(z_t, \hat{\theta}_t\right)$$

$$\hat{\theta}_t := \sum_k \theta_k \max\left(0, 1 - \left|\hat{\zeta}_t - k\right|\right)$$

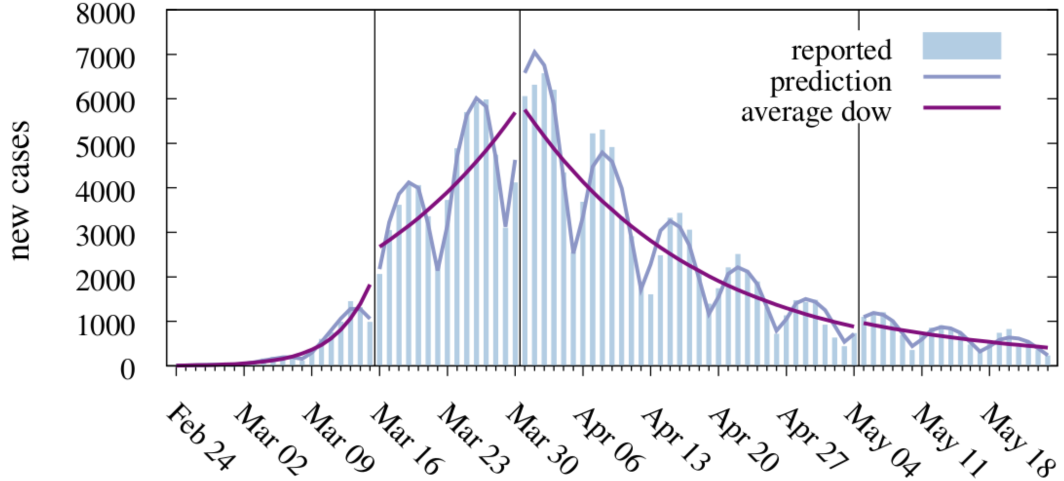
$$\hat{\zeta}_t := 1 + \gamma_{\text{TSP}}\left(\frac{t-1}{T-1}; m\right) \cdot (K-1)$$

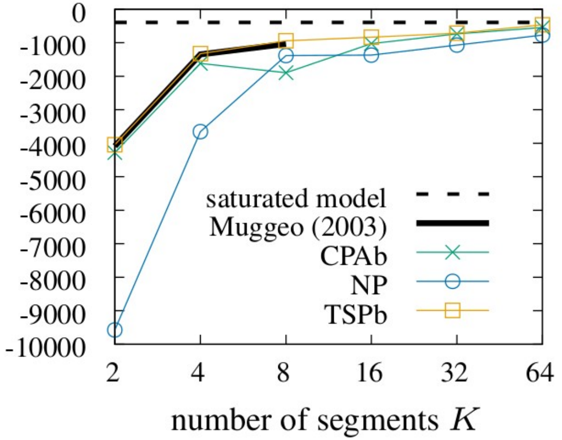
$$m_k := \frac{\sum_{k' \leq k} \exp(\mu_{k'})}{\sum_{k'} \exp(\mu_{k'})}$$

$$\min_{\mu,\theta} \mathcal{L}(\mu, \theta)$$



Example: COVID-19 in Germany

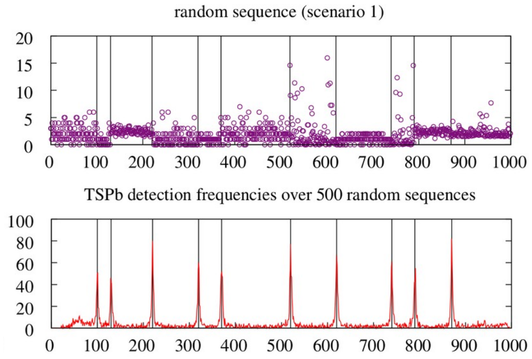




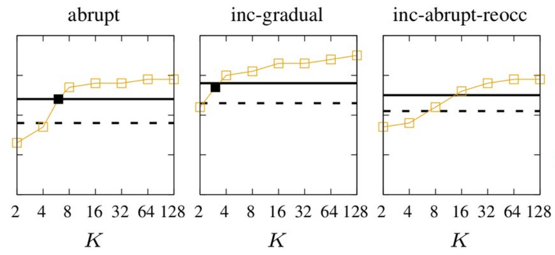
DGP: Poisson regression (GLM)

More experiments in the paper

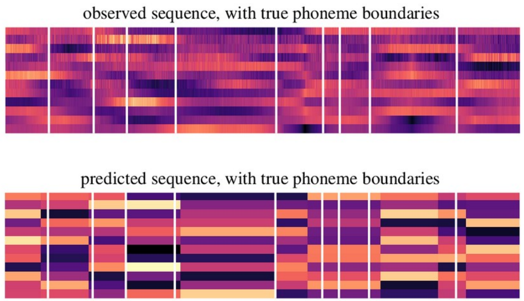
change point detection



classification with concept drift



phoneme segmentation



relaxed segmented model
high modeling capacity for nonstationary sequential data with discrete change points

Arlot, S., Celisse, A., & Harchaoui, Z. (2019). **A Kernel Multiple Change-point Algorithm via Model Selection.** Journal of Machine Learning Research, 20, 1–56.

Souza, V. M. A., dos Reis, D. M., Maletzke, A. G., & Batista, G. E. A. P. A. (2020). **Challenges in benchmarking stream learning algorithms with real-world data.** Data Mining and Knowledge Discovery, 34(6), 1805–1858.