Regionally Additive Models: Explainable-by-design models minimizing feature interactions

Vasilis Gkolemis[1,2] Anargyros Tzerefos[1] Theodore Dalamagas[1]
Eirini Ntoutsi[3] Christos Diou[2]

[1]ATHENA Research and Innovation Center

[2]Harokopio University of Athens

[3]Universitat der Bundeswehr Munchen

September 2023, Turin, Italy

Generalized Additive Models

Wikipedia says:

> *In statistics, a generalized additive model (GAM) is a generalized linear model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions.*

Generalized Additive Models (GAMs)

Wikipedia says:

> *In statistics, a generalized additive model (GAM) is a generalized linear model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions.*

Generalized Additive Models (GAMs)

Wikipedia says:

> *In statistics, a generalized additive model (GAM) is a generalized linear model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions.*

---

y

Generalized Additive Models (GAMs)

Wikipedia says:

> *In statistics, a generalized additive model (GAM) is a generalized linear model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions.*

$$y = \cdot + \ldots + \cdot$$

Generalized Additive Models (GAMs)

Wikipedia says:

> *In statistics, a generalized additive model (GAM) is a generalized linear model in which the linear response variable depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions.*

$$y = f_1(x_1) + \ldots + f_D(x_D)$$

GAMs - Interpretability

GAMs are called explainable-by-design, why?

Output/target variable:

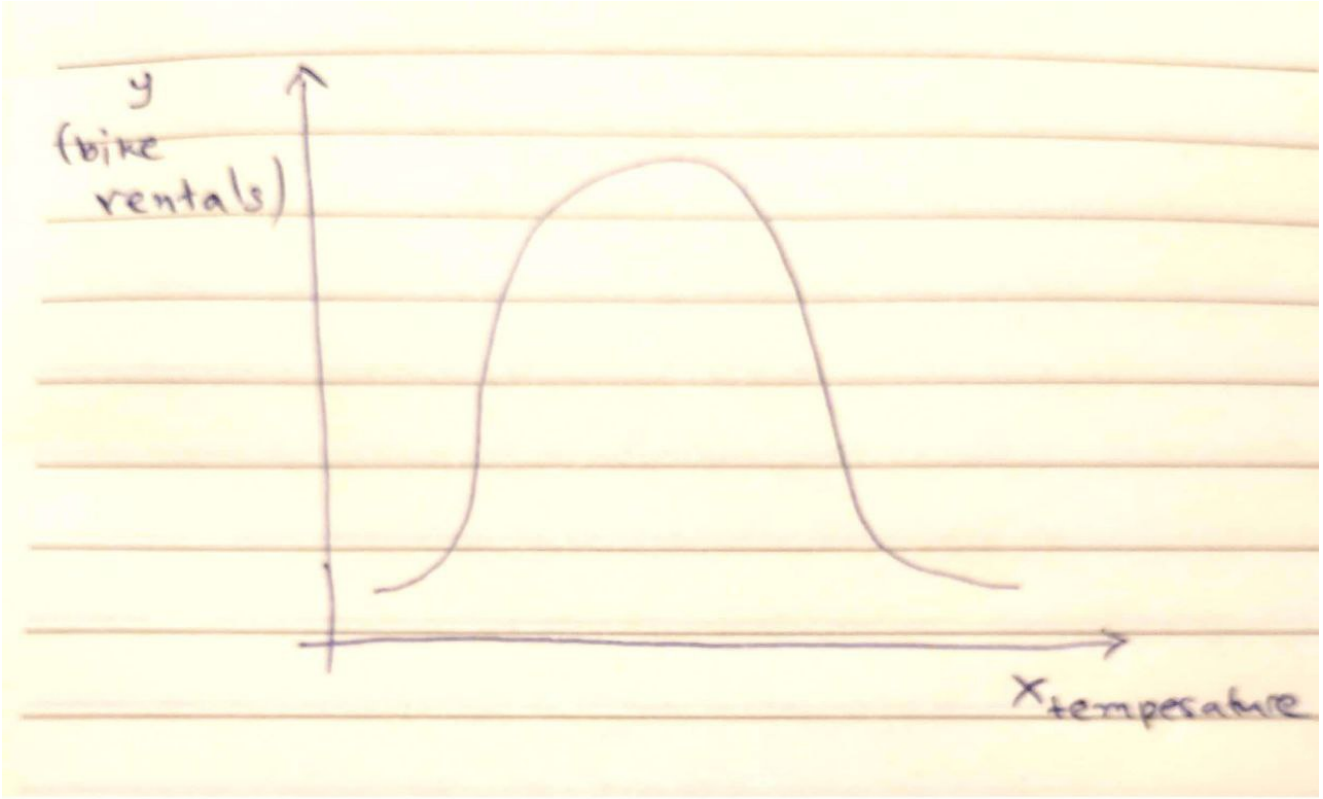
- $y_{\texttt{bike\_rentals}}$ the expected number of bike rentals

Input/covariates:

- $x_{\texttt{temperature}}$
- $x_{\texttt{humidity}}$
- $x_{\texttt{is\_weekday}}$

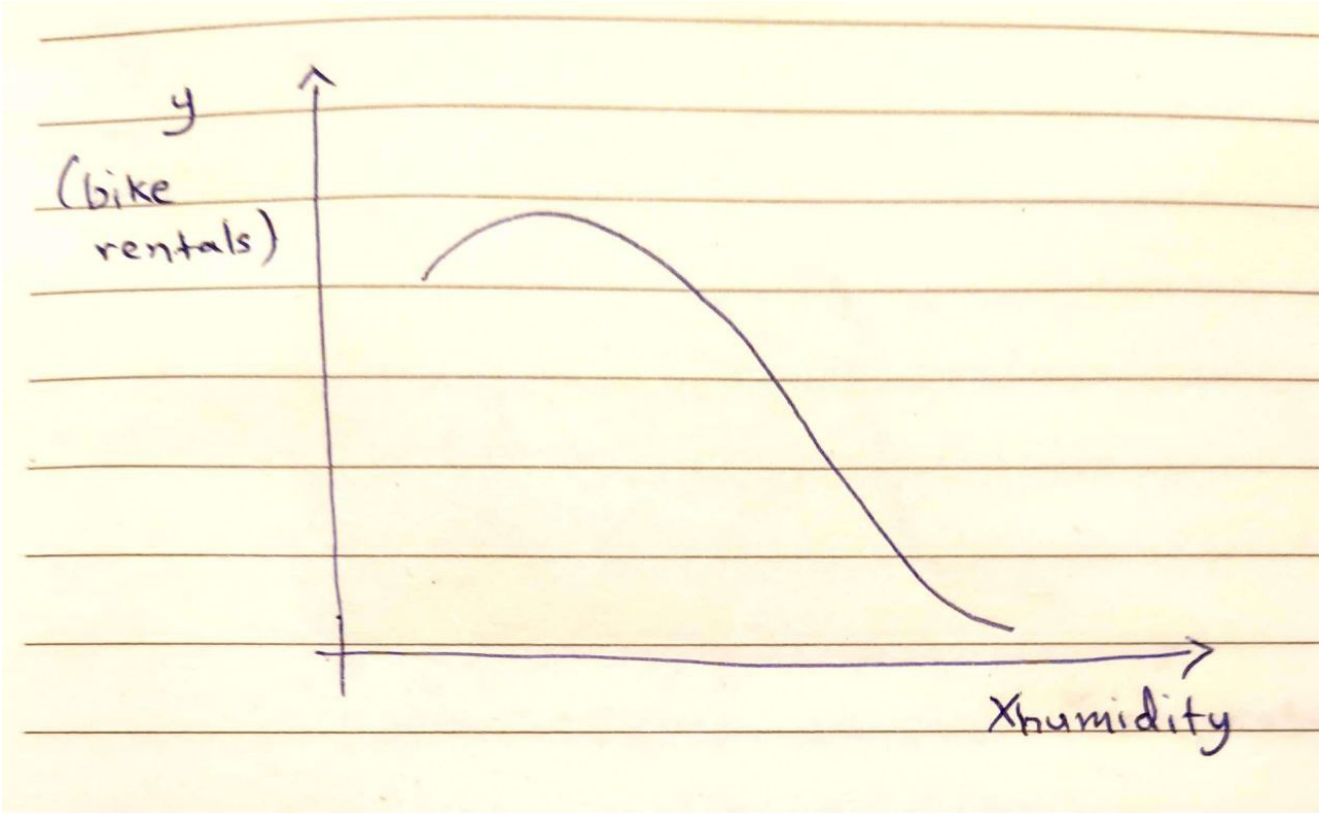
GAMs - Interpretability

GAMs are explainable-by-design, why?



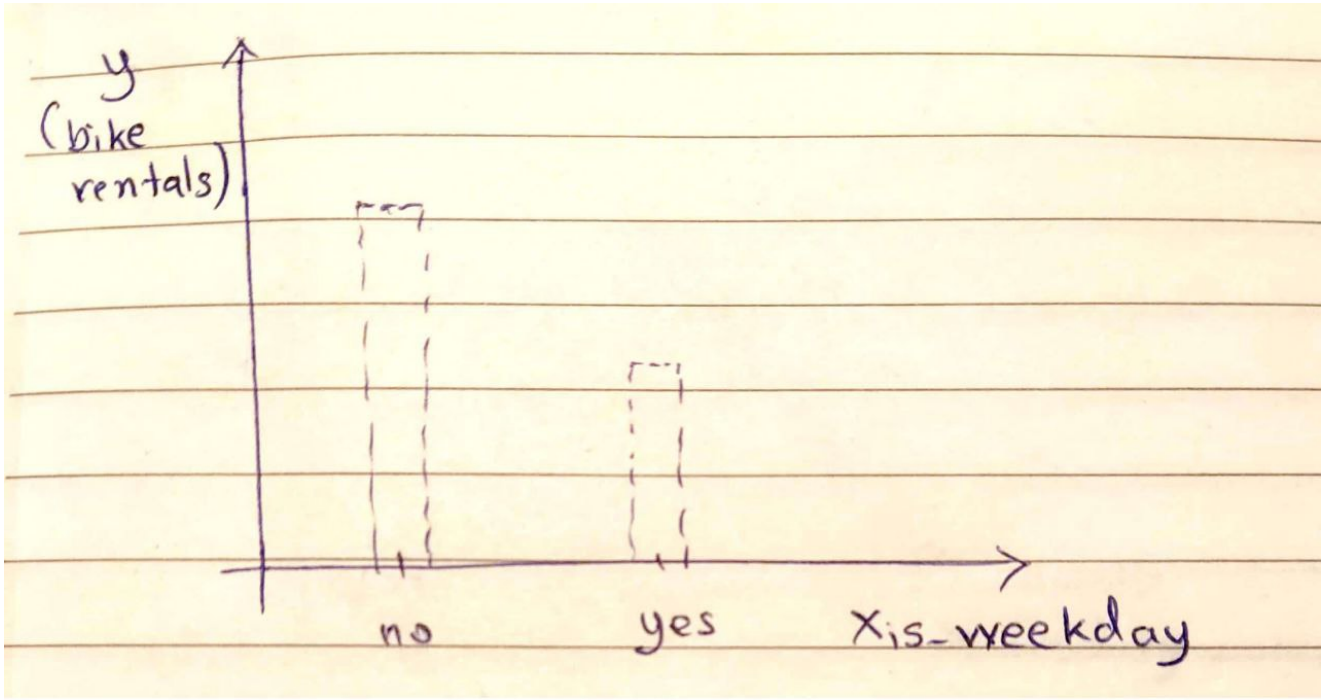
GAMs - Interpretability

GAMs are explainable-by-design, why?



GAMs - Interpretability

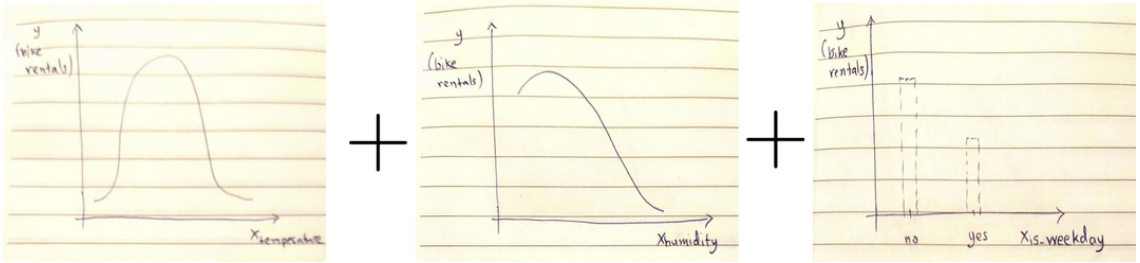
GAMs are explainable-by-design, why?



GAMs - Interpretability

GAMs are explainable-by-design, why?

$$y = f_1(x_{\texttt{temperature}}) + f_2(x_{\texttt{humidity}}) + f_3(x_{\texttt{is\_weekday}})$$



GAMs - Limitations

What if:

- temperature has different effect on week-days vs weekends

GAMs - Limitations

What if:

- temperature has different effect on week-days vs weekends
- Cause: go to work vs go sightseeing

GAMs - Limitations

What if:

- temperature has different effect on week-days vs weekends
- Cause: go to work vs go sightseeing
- Solution 1: $f(x_{\texttt{temperature}}, x_{\texttt{is\_weekday}})$

GAMs - Limitations

What if:

- temperature has different effect on week-days vs weekends
- Cause: go to work vs go sightseeing
- Solution 1: $f(x_{\texttt{temperature}}, x_{\texttt{is\_weekday}})$
- Solution 2: $f(x_{\texttt{temperature}}|\mathit{weekday})$ vs $f(x_{\texttt{temperature}}|\mathit{weekend})$

GAMs - Limitations

What if:

- temperature has different effect on week-days vs weekends
- Cause: go to work vs go sightseeing
- Solution 1: $f(x_{\texttt{temperature}}, x_{\texttt{is\_weekday}})$
- Solution 2: $f(x_{\texttt{temperature}}|\mathit{weekday})$ vs $f(x_{\texttt{temperature}}|\mathit{weekend})$

- Solution 1: GA^2M

GAMs - Limitations

What if:

- temperature has different effect on week-days vs weekends
- Cause: go to work vs go sightseeing
- Solution 1: $f(x_{\texttt{temperature}}, x_{\texttt{is\_weekday}})$
- Solution 2: $f(x_{\texttt{temperature}}|\mathit{weekday})$ vs $f(x_{\texttt{temperature}}|\mathit{weekend})$

- Solution 1: GA^2M
- Solution 2: RAM

GAMs - Limitations

What if:

- Have you ever rided a bike with cold and humidity?

Solutions:

GAMs - Limitations

What if:

- Have you ever rided a bike with cold and humidity?
- If it is weekend, cool! Netflix is here!

Solutions:

GAMs - Limitations

What if:

- Have you ever rided a bike with cold and humidity?
- If it is weekend, cool! Netflix is here!
- But what to do if the only option to go to work?

Solutions:

GAMs - Limitations

What if:

- Have you ever rided a bike with cold and humidity?
- If it is weekend, cool! Netflix is here!
- But what to do if the only option to go to work?

Solutions:

- $f(x_{\texttt{temperature}}, x_{\texttt{humidity}} | x_{\texttt{is\_weekday}})$ RA^2M

GAMs - Limitations

What if:

- Have you ever rided a bike with cold and humidity?
- If it is weekend, cool! Netflix is here!
- But what to do if the only option to go to work?

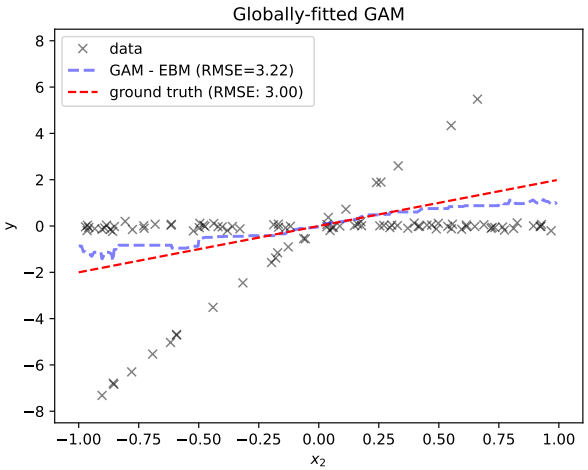
Solutions:

- $f(x_{\texttt{temperature}}, x_{\texttt{humidity}} | x_{\texttt{is\_weekday}})$ RA^2M
- $f(x_{\texttt{temperature}} | x_{\texttt{humidity}} = \{high, low\}, x_{\texttt{is\_weekday}})$ RAM with two conditions

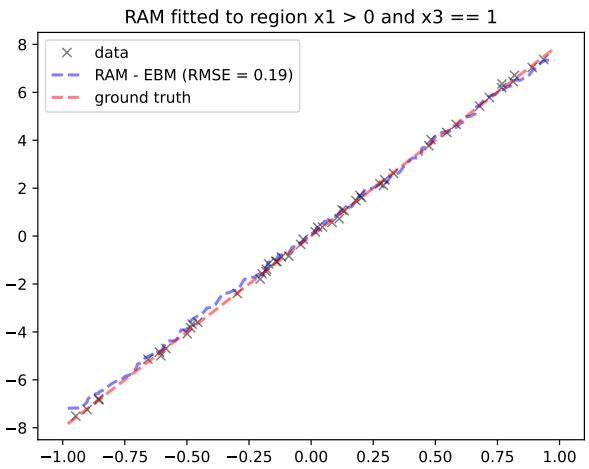
RAM on toy example

$$f(\boldsymbol{x}) = 8x_2\mathbb{1}_{x_1>0}\mathbb{1}_{x_3=0}$$

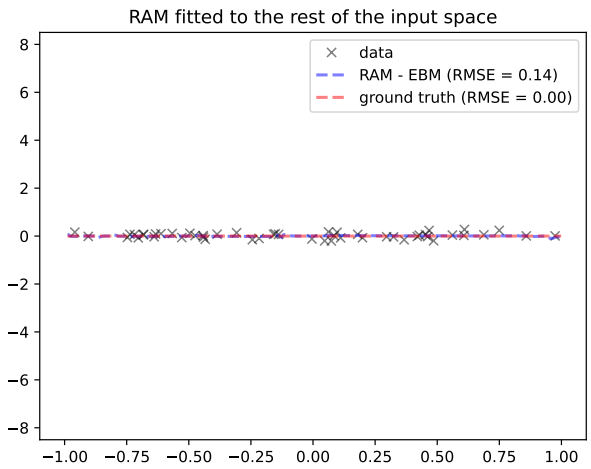
$$x_1, x_2 \sim \mathcal{U}(-1,1), x_3 \sim \mathit{Bernoulli}(0,1)$$



(a) $f_2(x_2)$



(b) $f_2(x_2)\mathbb{1}_{x_1>0 \text{ and } x_3=1}$



(c) $f_2(x_2)\mathbb{1}_{x_1\leq 0 \text{ or } x_3\neq 1}$

Figure: (Left) GAM, (Middle and Right) RAM

How RAM works (1)

3-step approach:

- Fit a Black-box model to capture all complex structures
  - it should be differentiable
  - neural network
- Use a Regional Effect method to find important interactions
  - RHALE - Gkolemis et. al
  - Feature Interactions - Herbinger et. al
- Fit a univariate function on each detected subregion

Step 1

- Fit a Black-box model to capture all complex structures
  - it should be differentiable
  - A neural network is a good option

Step 2

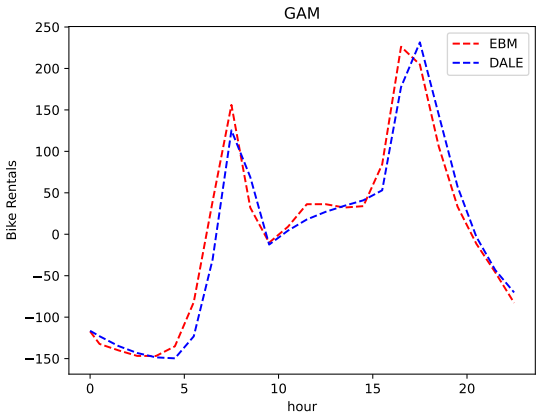
- Use a Regional Effect method to find important interactions
  - RHALE - Gkolemis et. al
  - Feature Interactions - Herbinger et. al

Step 3

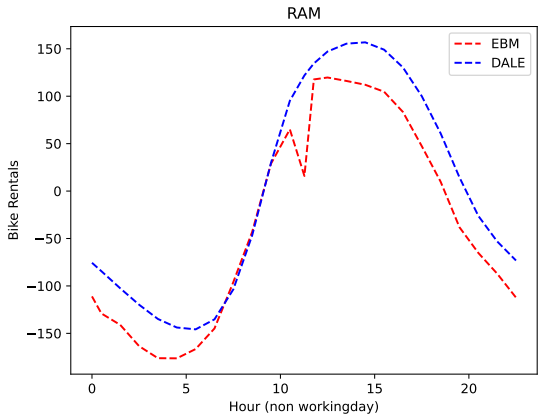
- Fit a univariate function on each detected subregion
  -

Bike Sharing dataset

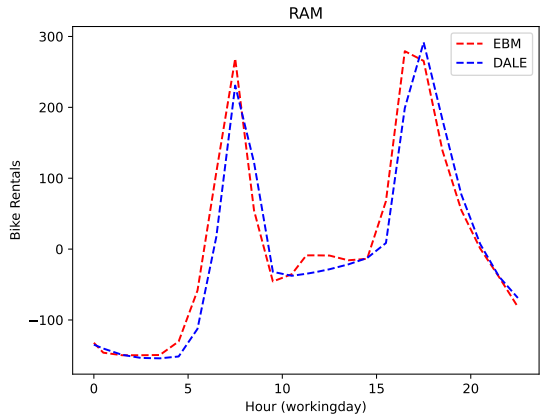
Predict bike-rentals (on daily basis) given some features of the day



(a) $f(X_{\texttt{hour}})$



(b) $f(X_{\texttt{hour}})\mathbb{1}_{X_{\texttt{workingday}}\neq 1}$



(c) $f(X_{\texttt{hour}})\mathbb{1}_{X_{\texttt{workingday}}=1}$

Experimental Results

	Black-box	**x-by-design**			
	all orders	1^{st} order		2^{nd} order	
	DNN	**GAM**	**RAM**	**GA^{2}M**	**RA^{2}M**
Bike (MAE)	0.254	0.549	0.430	0.298	0.278
Bike (RMSE)	0.389	0.734	0.563	0.438	0.412
Housing (MAE)	0.373	0.600	0.553	0.554	0.533
Housing (RMSE)	0.533	0.819	0.754	0.774	0.739

What next?

A lot has to be done!

- Results are only preliminary
  - Experiment with more datasets
  - Sometimes wide bins are needed
- Can we learn uncertain RAMs?
  - How do we model uncertainty?

Thank you for the attention

- For more discussion or future ideas on RAM, contact me:
  - vgkolemis@athenarc.gr
  - gkolemis@hua.gr
- Questions?