Totally Concave Regression

Dohyeong Ki

Department of Statistics, UC Berkeley

Aug 5, 2025

Joint work with Aditya Guntuboyina

We propose a new multivariate generalization of univariate concave regression.

Our new method is based on total concavity.

Total concavity is a multivariate extension of univariate concavity, defined via max-order mixed (partial) derivative constraints.

Univariate Concave Regression

Data: $(x^{(1)}, y_1), \ldots, (x^{(n)}, y_n)$ where $x^{(i)} \in [0,1]$ and $y_i \in \mathbb{R}$

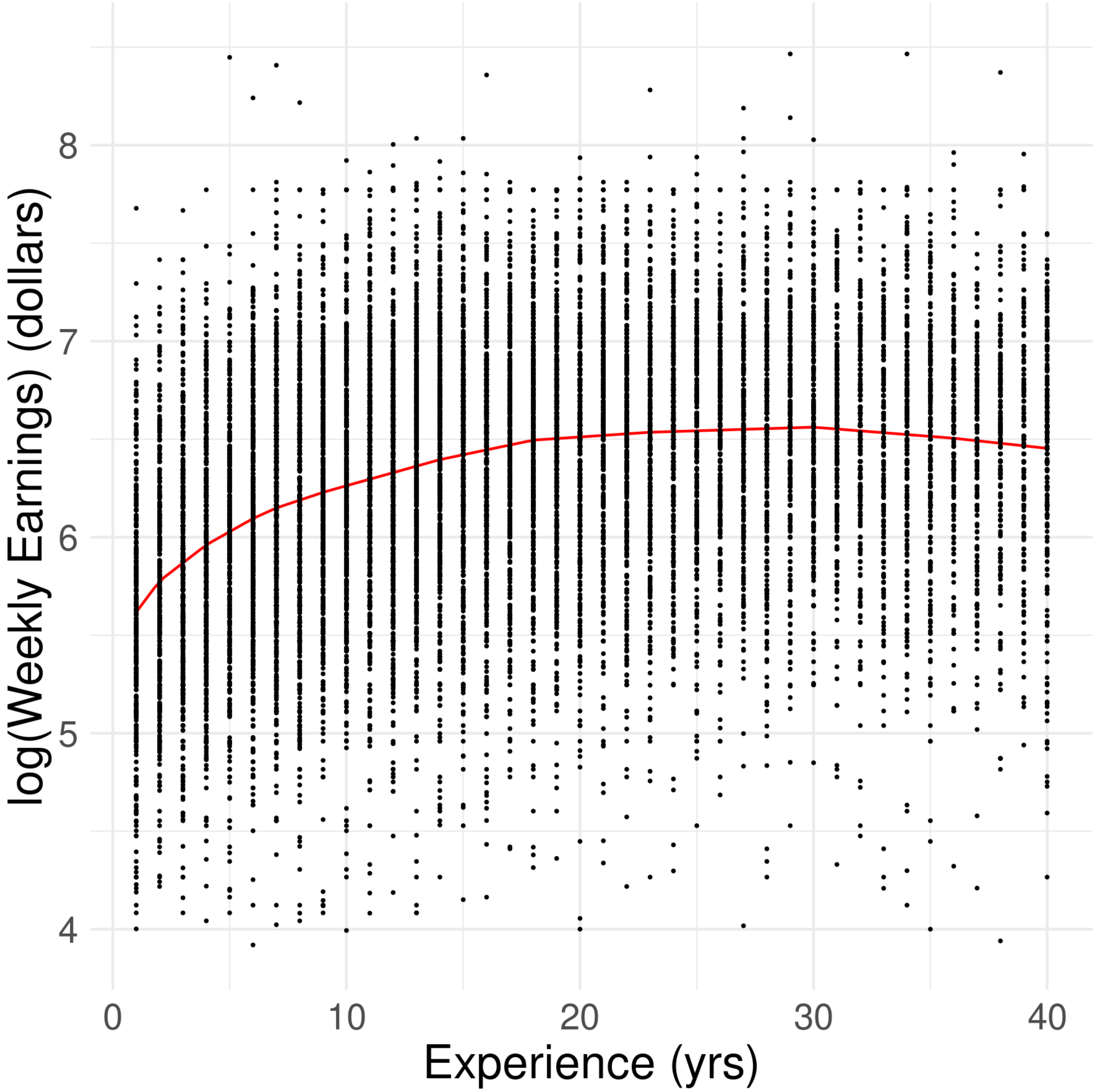
Suppose y has diminishing returns with respect to x.

In other words, the rate of change in y decreases as x increases.

$(= dy/dx)$

$$\hat{f}_{\text{concave}} \in \underset{f:\text{concave}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - f(x^{(i)})\right)^2$$

Example



log(Weekly Earnings) (dollars)
8
7
6
5
4
0
10
20
30
40
Experience (yrs)

Multivariate Concave Regression

What are multivariate versions of $\hat{f}_{\text{concave}}$?

For simplicity, we mainly focus on the bivariate case $d = 2$.

Data: $(x^{(1)}, y_1), \ldots, (x^{(n)}, y_n)$ where $x^{(i)} \in [0,1]^2$ and $y_i \in \mathbb{R}$

Example: $y = \log(\text{Earnings})$, $x_1 = \text{Education}$, and $x_2 = \text{Experience}$.

Existing Notions of Multivariate Concavity

- General Concavity

- Axial (= Coordinate-wise) Concavity

- Additive Concavity

Total concavity is different from them!

General Concavity

f is (generally) concave if and only if

$$f((1-\alpha)x + \alpha y) \geq (1-\alpha)f(x) + \alpha f(y)$$

for all $\alpha \in [0,1]$ and $x, y \in [0,1]^2$.

Studied in, for example,

[Balázs 2016], [Kuosmanen 2008], [Lim & Glynn 2012],

[Seijo & Sen 2011], [Kur et al 2024], ...

Weakness:

It requires concavity on every line $\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 = 0$.

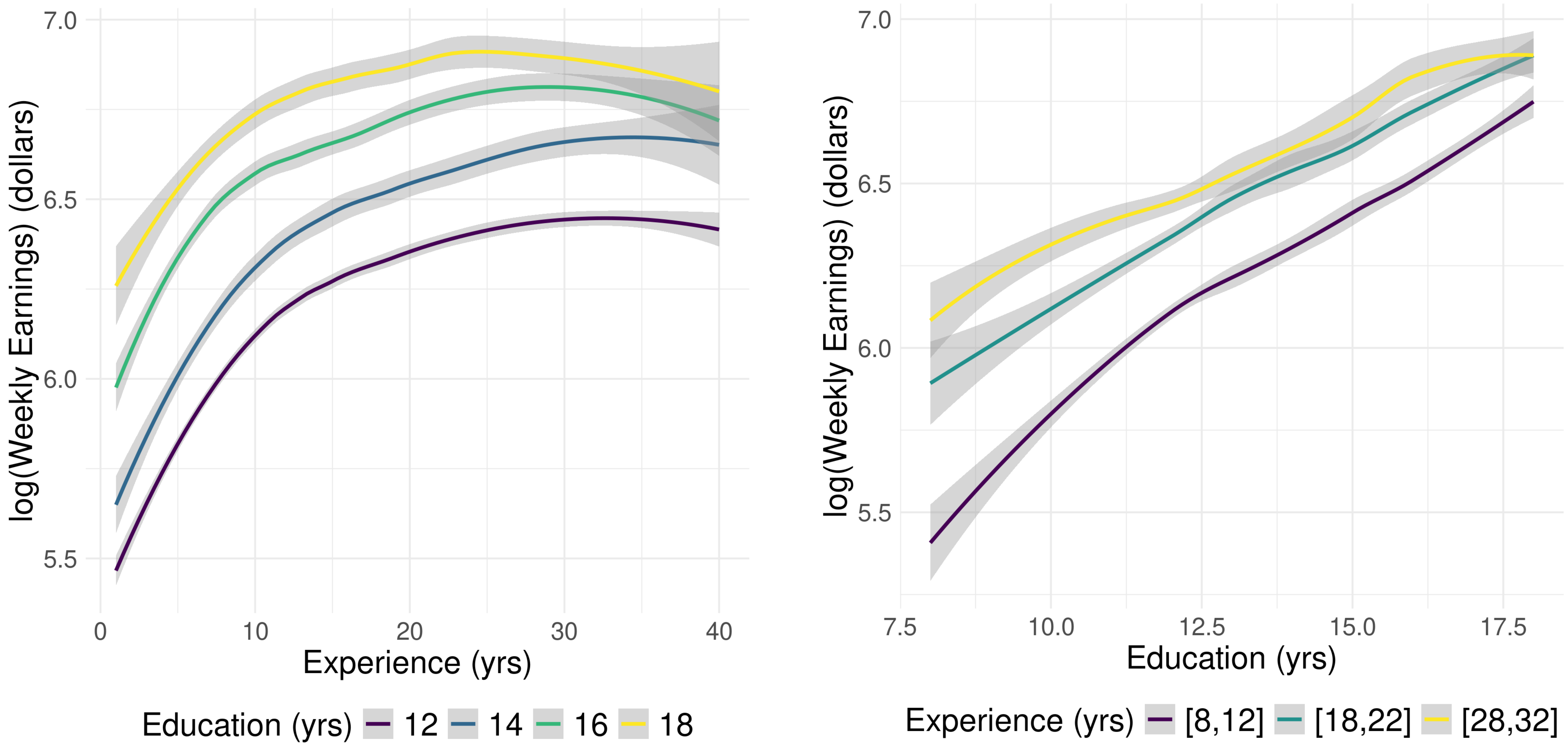
Axial Concavity

f is axially (coordinate-wise) concave if and only if

$f(\cdot, x_2)$ and $f(x_1, \cdot)$ are concave for each $x_1, x_2 \in [0,1]$.

It is often justifiable from domain knowledge or data.

Example



7.0
6.5
6.0
5.5
log(Weekly Earnings) (dollars)
0
10
20
30
40
Experience (yrs)
Education (yrs) 12 14 16 18
7.0
6.5
6.0
5.5
log(Weekly Earnings) (dollars)
7.5
10.0
12.5
15.0
17.5
Education (yrs)
Experience (yrs) [8,12] [18,22] [28,32]

Axial Concavity

Studied in, for example, [Iwanaga et al 2016]

Weakness:

Axial concavity is a weak assumption.

It may require more data to avoid overfitting than others

Additive Concavity

f is additive concave if and only if

$f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ for some univariate concave functions f_1 and f_2.

Studied, for example, in

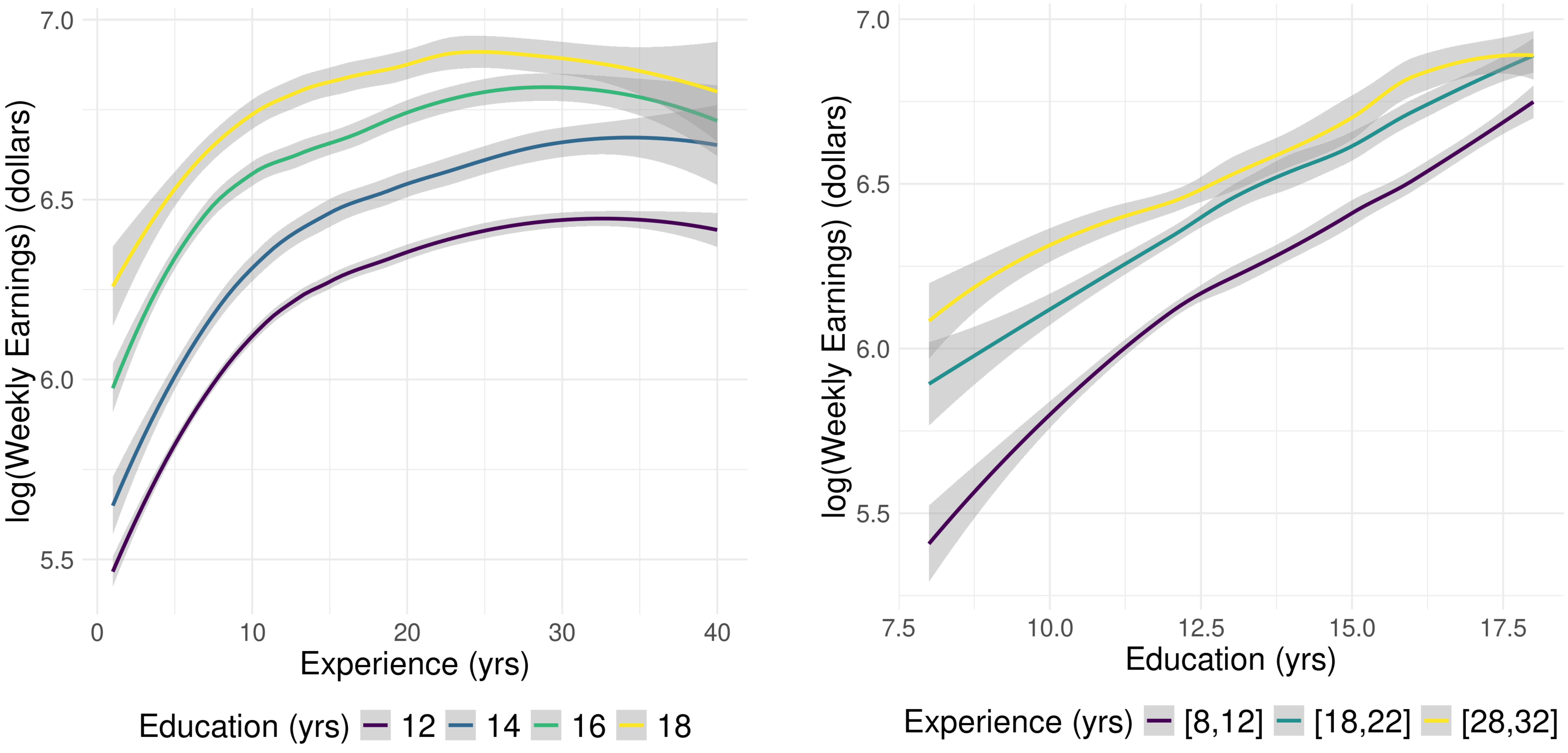
[Chen & Samworth 2016], [Meyer 2013], [Meyer 2018], [Pya & Wood 2015], ...

Weakness:

No interaction effects

Can be restrictive sometimes.

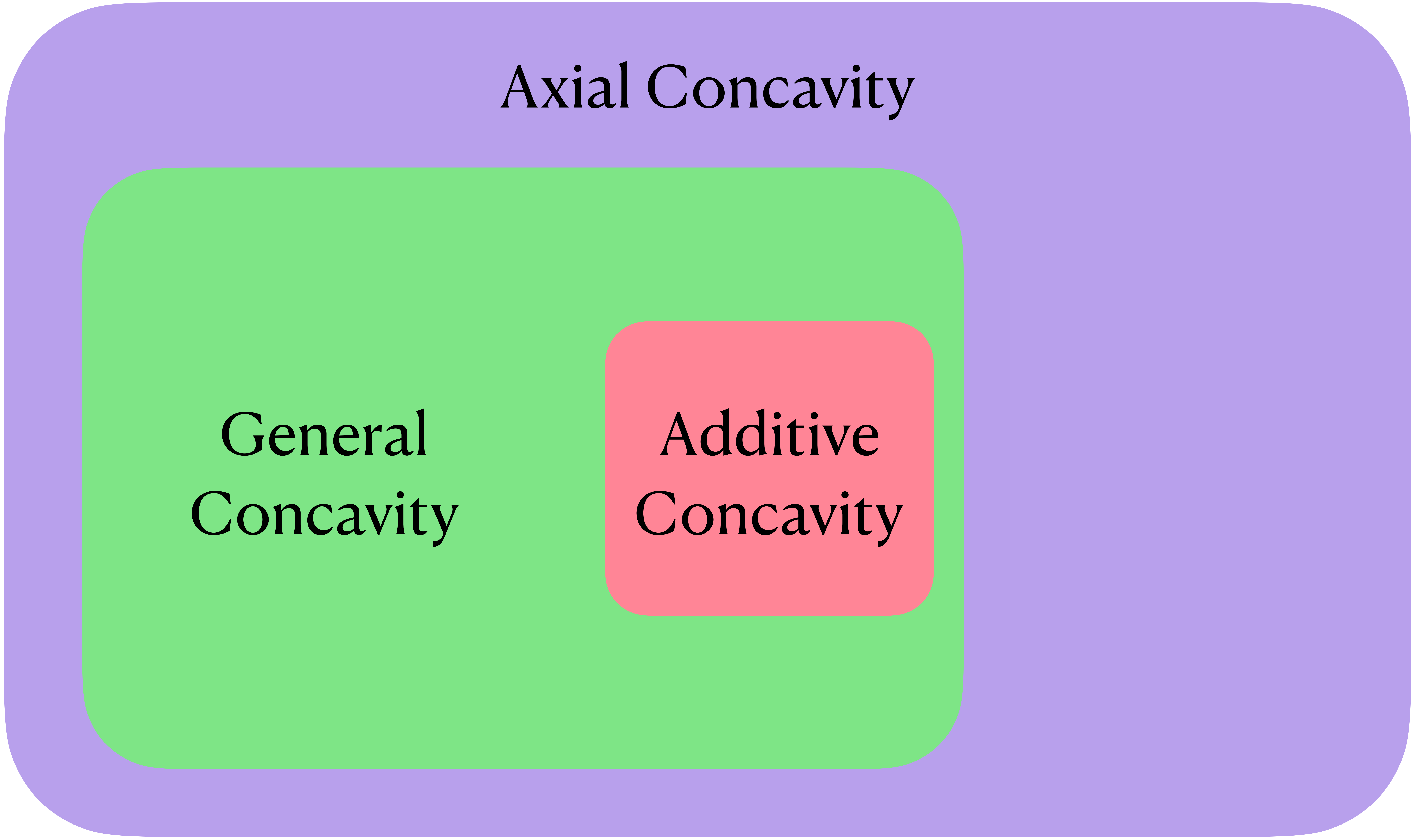
Example Again



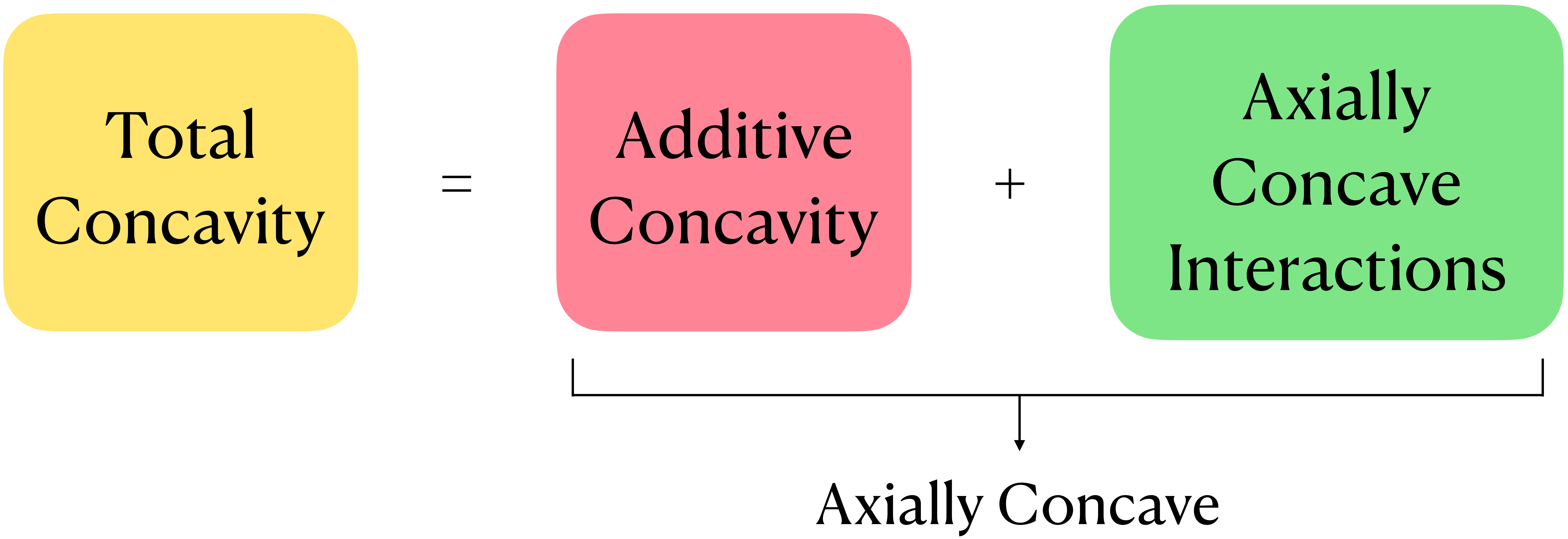
Characterization via Mixed (Partial) Derivatives

For smooth functions,

- Univariate Concavity: $f'' \leq 0$, i.e., $\dfrac{\partial^2 f}{\partial x_1^2} \leq 0$.
- General Concavity: $\dfrac{\partial^2 f}{\partial x_1^2} \leq 0$, $\dfrac{\partial^2 f}{\partial x_2^2} \leq 0$, and $\left(\dfrac{\partial^2 f}{\partial x_1 x_2}\right)^2 \leq \dfrac{\partial^2 f}{\partial x_1^2} \cdot \dfrac{\partial^2 f}{\partial x_2^2}$.
- Axial Concavity: $\dfrac{\partial^2 f}{\partial x_1^2} \leq 0$ and $\dfrac{\partial^2 f}{\partial x_2^2} \leq 0$.
- Additive Concavity: $\dfrac{\partial^2 f}{\partial x_1^2} \leq 0$, $\dfrac{\partial^2 f}{\partial x_2^2} \leq 0$, and $\dfrac{\partial^2 f}{\partial x_1 x_2} = 0$.


Axial Concavity
General Concavity
Additive Concavity

Total Concavity



First introduced by Popoviciu in 1934 and recently described in [Gal 2010].

Representation Theorem for Univariate Concave Functions

Suppose $f : [0,1] \to \mathbb{R}$ is a concave function with $f'(0+) < +\infty$ and $f'(1-) > -\infty$.

Then, there exists a unique (Borel) measure ν on $(0,1)$ and $a_0, a_1 \in \mathbb{R}$ such that for all $x \in [0,1]$,

$$f(x) = a_0 + a_1 x - \int_{(0,1)} (x - t)_+ \, d\nu(t),$$

where $(x - t)_+ = \max\{x - t,\ 0\}$.

Concave functions are (infinite) linear combinations of basis functions $x \mapsto (x - t)_+$ with non-positive weights.

Additive Concave Functions

Additive concave functions can be written as

$$f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 - \int_{(0,1)} (x_1 - t_1)_+ \, d\nu_1(t_1) - \int_{(0,1)} (x_2 - t_2)_+ \, d\nu_2(t_2)$$

for some measures ν_1 and ν_2 on $(0,1)$.

Total Concavity = Additive Concavity + Axially Concave Interactions

How can we add interaction terms to these additive concave functions?

Axially Concave Interactions

How do we introduce interaction terms to linear regression?

Example:

$$y = \beta_1 x_1 + \beta_2 x_2 \;\rightarrow\; y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

Interactions can be introduced via

$$x_1 x_2,\; x_1(x_2 - t_2)_+,\; (x_1 - t_1)_+ x_2,\; \text{and}\; (x_1 - t_1)_+(x_2 - t_2)_+$$

Total Concavity = Additive Concavity + **Axially Concave** Interactions

When are they axially concave?

- $\beta x_1 x_2$: axially concave for all $\beta \in \mathbb{R}$.
- $\beta x_1 (x_2 - t_2)_+$, $\beta (x_1 - t_1)_+ x_2$, $\beta (x_1 - t_1)_+ (x_2 - t_2)_+$: axially concave iff $\beta \leq 0$.

Totally Concave Functions

$$f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 - \int (x_1 - t_1)_+ \, d\nu_1(t_1) - \int (x_2 - t_2)_+ \, d\nu_2(t_2)$$

$$+ a_{12} x_1 x_2 - x_1 \cdot \int (x_2 - t_2)_+ \, d\nu_{1=0,2} - x_2 \cdot \int (x_1 - t_1)_+ \, d\nu_{1,2=0}$$

$$- \int (x_1 - t_1)_+ (x_2 - t_2)_+ \, d\nu_{12}(t_1, t_2),$$

where $\nu_1, \nu_2, \nu_{1=0,2}, \nu_{1,2=0}$ are measures on $(0,1)$ and ν_{12} is a measure on $(0,1)^2$.

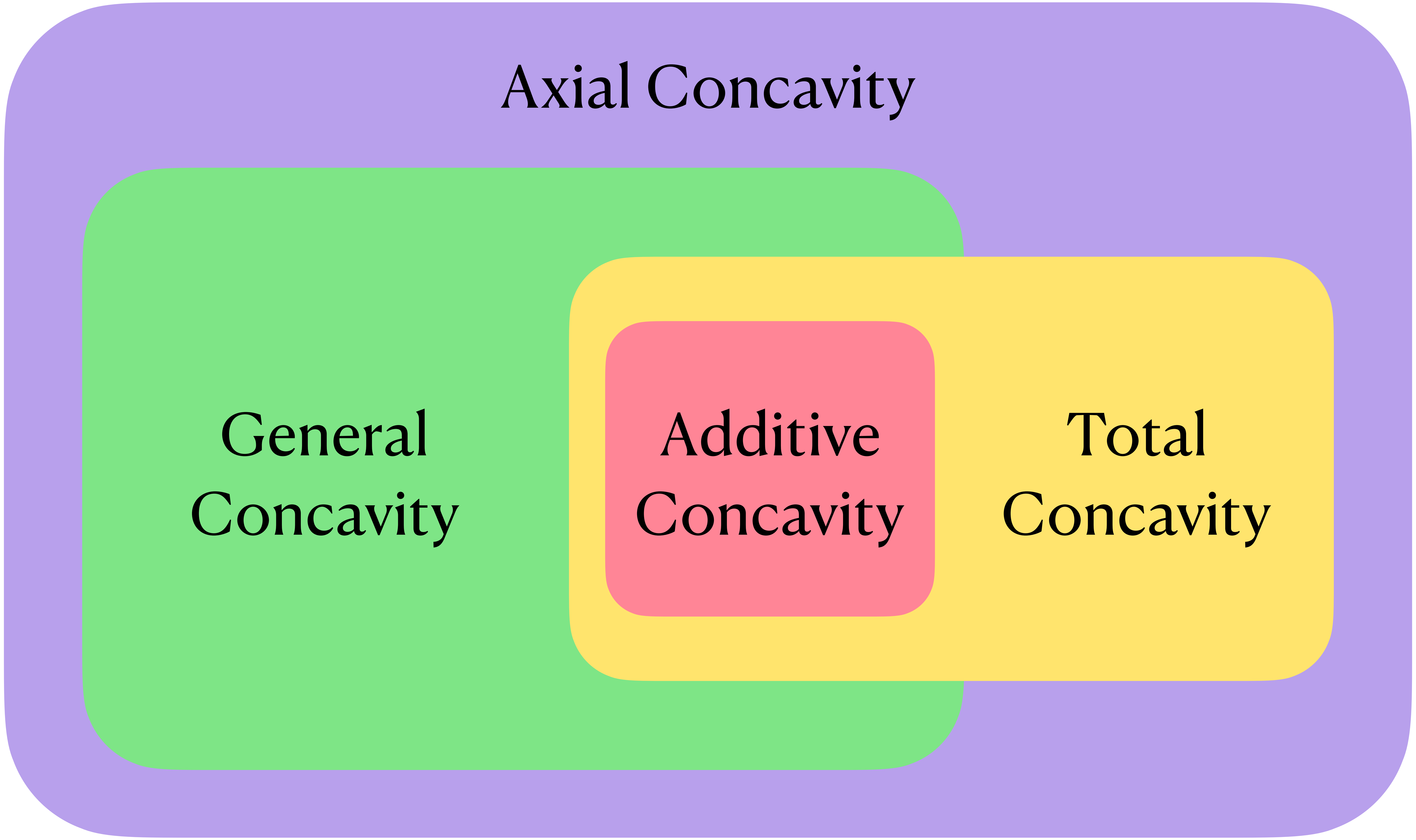
If f is a smooth totally concave function, then

$$\frac{\partial^2 f}{\partial x_1^2} \leq 0, \frac{\partial^2 f}{\partial x_2^2} \leq 0, \frac{\partial^3 f}{\partial x_1 x_2^2} \leq 0, \frac{\partial^3 f}{\partial x_1^2 x_2} \leq 0, \text{ and } \frac{\partial^4 f}{\partial x_1^2 x_2^2} \leq 0.$$

Characterization via Derivatives

- General Concavity: $\frac{\partial^2 f}{\partial x_1^2} \leq 0, \frac{\partial^2 f}{\partial x_2^2} \leq 0$, and $\left(\frac{\partial^2 f}{\partial x_1 x_2}\right)^2 \leq \frac{\partial^2 f}{\partial x_1^2} \cdot \frac{\partial^2 f}{\partial x_2^2}$.
- Axial Concavity: $\frac{\partial^2 f}{\partial x_1^2} \leq 0$ and $\frac{\partial^2 f}{\partial x_2^2} \leq 0$.
- Additive Concavity: $\frac{\partial^2 f}{\partial x_1^2} \leq 0, \frac{\partial^2 f}{\partial x_2^2} \leq 0$, and $\frac{\partial^2 f}{\partial x_1 x_2} = 0$.
- Total Concavity:

$$\frac{\partial^2 f}{\partial x_1^2} \leq 0, \frac{\partial^2 f}{\partial x_2^2} \leq 0, \frac{\partial^3 f}{\partial x_1 x_2^2} \leq 0, \frac{\partial^3 f}{\partial x_1^2 x_2} \leq 0, \text{ and } \frac{\partial^4 f}{\partial x_1^2 x_2^2} \leq 0.$$


Axial Concavity
General Concavity
Additive Concavity
Total Concavity

Extension to Higher Dimensions

Recall that when $d = 2$, f is totally concave if

$$\frac{\partial^2 f}{\partial x_1^2} \leq 0, \frac{\partial^2 f}{\partial x_2^2} \leq 0, \frac{\partial^3 f}{\partial x_1 x_2^2} \leq 0, \frac{\partial^3 f}{\partial x_1^2 x_2} \leq 0, \text{ and } \frac{\partial^4 f}{\partial x_1^2 x_2^2} \leq 0;$$

that is,

$$\frac{\partial^{p_1+p_2} f}{\partial x_1^{p_1} \partial x_2^{p_2}} \leq 0$$

for every $(p_1, p_2) \in \{0,1,2\}^2$ with $\max\{p_1, p_2\} = 2$.

For general d, f is totally concave if

$$\frac{\partial^{p_1+\cdots+p_d} f}{\partial x_1^{p_1} \cdots \partial x_d^{p_d}} \leq 0$$

for every $(p_1, \ldots, p_d) \in \{0,1,2\}^d$ with $\max_k p_k = 2$

Totally Concave Regression

Data: $(x^{(1)}, y_1), \ldots, (x^{(n)}, y_n)$ where $x^{(i)} \in [0,1]^d$ and $y_i \in \mathbb{R}$

$$\hat{f}_{\mathrm{TC}} \in \underset{f:\ \text{totally concave}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - f(x^{(i)})\right)^2$$

It can be computed via finite-dimensional convex optimization algorithms.

Rate of Convergence

Under the standard set of model assumptions:

(1) $y_i = f^*(x^{(i)}) + \epsilon_i$ where f^* is totally concave and $\epsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$

(2) $x^{(1)}, \ldots, x^{(n)}$ form an equally-spaced lattice,

we have

multiplicative constant depends on d, σ, and f^*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_{\mathrm{TC}}(x^{(i)}) - f^*(x^{(i)})\right)^2\right] = O\left(n^{-\frac{4}{5}}(\log n)^{\frac{3(2d-1)}{5}}\right).$$

$\hat{f}_{TC}$ can avoid the curse of dimensionality to some extent.

Extensions

- We can restrict the order of interactions.

This leads to a faster rate of convergence.

- We can impose total concavity only on a subset of covariates.

The other covariates only appear via linear terms.

$$f(x_1, \ldots, x_d) = f_1(x_1, \ldots, x_p) + a_{p+1}x_{p+1} + \cdots + a_d x_d$$

where f_1 is a totally concave function and $a_{p+1}, \ldots, a_d \in \mathbb{R}$.

All of them are implemented in the R package regmdc, available at

https://github.com/DohyeongKi/regmdc.

Thank you for your attention!

https://arxiv.org/abs/2501.04360

References

Balázs, G. (2016). Convex Regression: Theory, Practice, and Applications. Ph. D. thesis, University of Alberta.

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. Econom. J. 11 (2), 308–325.

Lim, E. and P. W. Glynn (2012). Consistency of multidimensional convex regression. Ops Res. 60 (1), 196–208.

Seijo, E. and B. Sen (2011). Nonparametric least squares estimation of a multivariate convex regression function. Ann. Statist. 39 (3), 1633–1657.

Kur, G., F. Gao, A. Guntuboyina, and B. Sen (2024). Convex regression in multidimensions: Suboptimality of least squares estimators. Ann. Statist. 52 (6), 2791–2815.

Iwanaga, J., N. Nishimura, N. Sukegawa, and Y. Takano (2016). Estimating product-choice probabilities from recency and frequency of page views. Knowledge-Based Systems 99, 157–167.

Chen, Y. and R. J. Samworth (2016). Generalized additive and index models with shape constraints. J. R. Statist. Soc. B 78 (4), 729–754.

Meyer, M. C. (2013). Semi-parametric additive constrained regression. J. Nonparam. Stat. 25 (3), 715–730.

Meyer, M. C. (2018). A framework for estimation and inference in generalized additive models with shape and order restrictions. Statist. Sci. 33 (4), 595–614.

Pya, N. and S. N. Wood (2015). Shape constrained additive models. Statist. Comput. 25, 543–559.

Gal, S. G. (2010). Shape-Preserving Approximation by Real and Complex Polynomials. Boston, MA:Birkhäuser.