Faster Algorithms for Learning Convex Functions

Ali Siahkamari, Durmus Alp Emre Acar, Christopher Liao, Kelly Geyer, Venkatesh Saligrama, Brian Kulis

ICML 2022, Baltimore, USA

Faster Algorithms for Learning Convex functions

Problem setup

- Input:

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n) \quad | \quad \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R} \,.$$

- Model:

$$\mathcal{F} \triangleq \{f : \mathbb{R}^d \rightarrow \mathbb{R} \mid f \text{ is convex}\} \,.$$

- Estimation rule:

$$\hat{f} \triangleq \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\mathbf{x}_i))^2 + \lambda \|f\|,$$

where, $\|f\| \triangleq \sum_{l=1}^{d} \sup_{\mathbf{x}} \sup_{v \in \partial_{x_l} f(\mathbf{x})} |v|$

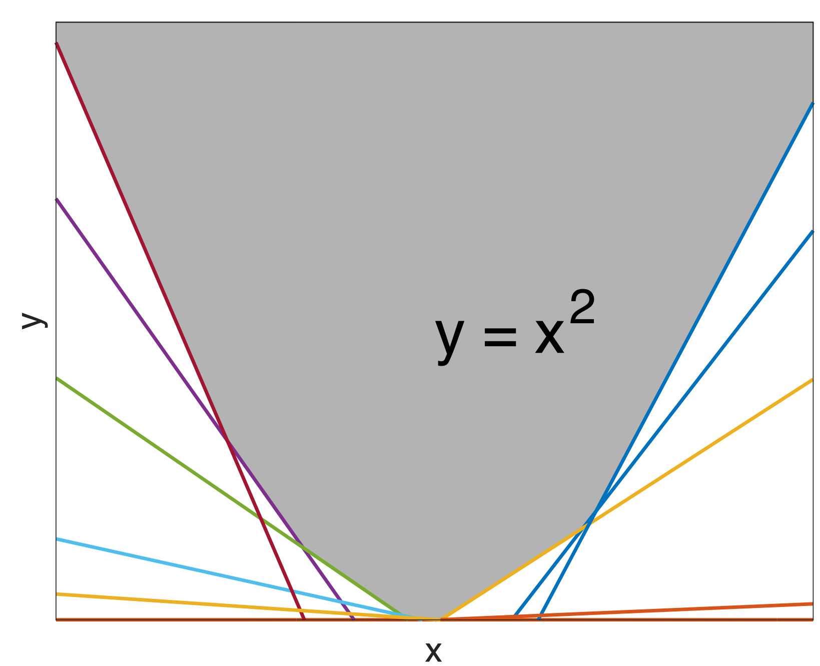
Faster Algorithms for Learning Convex functions

Convex formulation

Using [Boyd & Vandenberghe (2004)], the estimation rule for $\hat{f}$ is equivalent to solving:

$$\min_{\hat{y}_i, \mathbf{a_i}} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 + \lambda \sum_{l=1}^{d} \max_{i=1}^{n} |a_{i,l}|$$

$$s.t: \hat{y}_i - \hat{y}_j \geq \langle \mathbf{a}_j, \mathbf{x}_i - \mathbf{x}_j \rangle, \quad i, j \in [n].$$



And then constructing $\hat{f}$ as:

$$\hat{f}(\mathbf{x}) \triangleq \max_{i=1}^{n} \langle \mathbf{a}_i, \mathbf{x} - \mathbf{x}_i \rangle + \hat{y}_i$$

Faster Algorithms for Learning Convex functions

Convergence

Theorem 1: [from He & Yuan (2012)] Consider the separable convex optimization problem,

$$\min_{\mathbf{b}^1 \in \mathcal{S}_1, \mathbf{b}^2 \in \mathcal{S}_2} \left[\psi(\mathbf{b}^1, \mathbf{b}^2) = \psi_1(\mathbf{b}^1) + \psi_2(\mathbf{b}^2)\right] \qquad s.t: \mathbf{A}\mathbf{b}^1 + \mathbf{B}\mathbf{b}^2 + \mathbf{b} = \mathbf{0},$$

Let $\mathbf{b}_t^1$ and $\mathbf{b}_t^2$ be solutions at iteration t of a two block ADMM procedure with learning rate ρ.

Denote: $(\tilde{\mathbf{b}}_T^1, \tilde{\mathbf{b}}_T^2) = \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{b}_t^1, \frac{1}{T}\sum_{t=1}^{T} \mathbf{b}_t^2\right)$ and $(\mathbf{b}_1^*, \mathbf{b}_2^*)$ the optimal solution. For all κ we have:

$$\psi(\tilde{\mathbf{b}}_T^1, \tilde{\mathbf{b}}_T^2) - \psi(\mathbf{b}_*^1, \mathbf{b}_*^2) - \kappa^T(\mathbf{A}\tilde{\mathbf{b}}_T^1 + \mathbf{B}\tilde{\mathbf{b}}_T^2 + \mathbf{b}) \leq \frac{\mathbf{1}}{\mathbf{T}}\left(\frac{\rho}{\mathbf{2}}\|\mathbf{B}\mathbf{b}_*^{\mathbf{2}}\|^{\mathbf{2}} + \frac{\mathbf{1}}{\mathbf{2}\rho}\|\kappa\|^{\mathbf{2}}\right),$$

Faster Algorithms for Learning Convex functions

Convergence

Theorem 2: Let $\{\hat{y}_i^t, \mathbf{a}_i^t\}_{i=1}^n$ be the output of convex regression ADMM at t^{th} iteration,

Set $\tilde{y}_i \triangleq \frac{1}{T}\sum_{t=1}^{T} \hat{y}_i^t$, $\tilde{\mathbf{a}}_i \triangleq \frac{1}{T}\sum_{t=1}^{T} \mathbf{a}_i^t$ and $\rho = \frac{\sqrt{d}\lambda^2}{n}$

Denote. $\tilde{f}_T(\mathbf{x}) \triangleq \max_i \langle \tilde{\mathbf{a}}_i, \mathbf{x} - \mathbf{x}_i \rangle + \tilde{y}_i$.

Assume $\max_{i,l} |x_{i,l}| \leq 1$ and $Var(\{y_i\}_{i=1}^n) \leq 1$. For $\lambda \geq \frac{3}{\sqrt{2nd}}$ and $T \geq n\sqrt{d}$ we have:

$$\frac{1}{n}\sum_{i=1}^{n} (\tilde{f}_T(\mathbf{x}_i) - y_i)^2 + \lambda \|\tilde{f}_T\| \leq \min_{f \in \mathcal{F}} \left(\frac{1}{n}\sum_{i=1}^{n} \big(f(\mathbf{x}_i) - y_i\big)^2 + \lambda \|f\| \right) + \frac{6n\sqrt{d}}{T+1}$$

Faster Algorithms for Learning Convex functions

Computational complexity

SOTA computational complexity for Lipschitz convex regression:

$$O\left(\frac{n^5 d^2}{\epsilon}\right).$$

Out result:

$$O\left(\frac{n^3 d^{1.5} + n^2 d^{2.5} + n d^3}{\epsilon}\right).$$

Faster Algorithms for Learning Convex functions

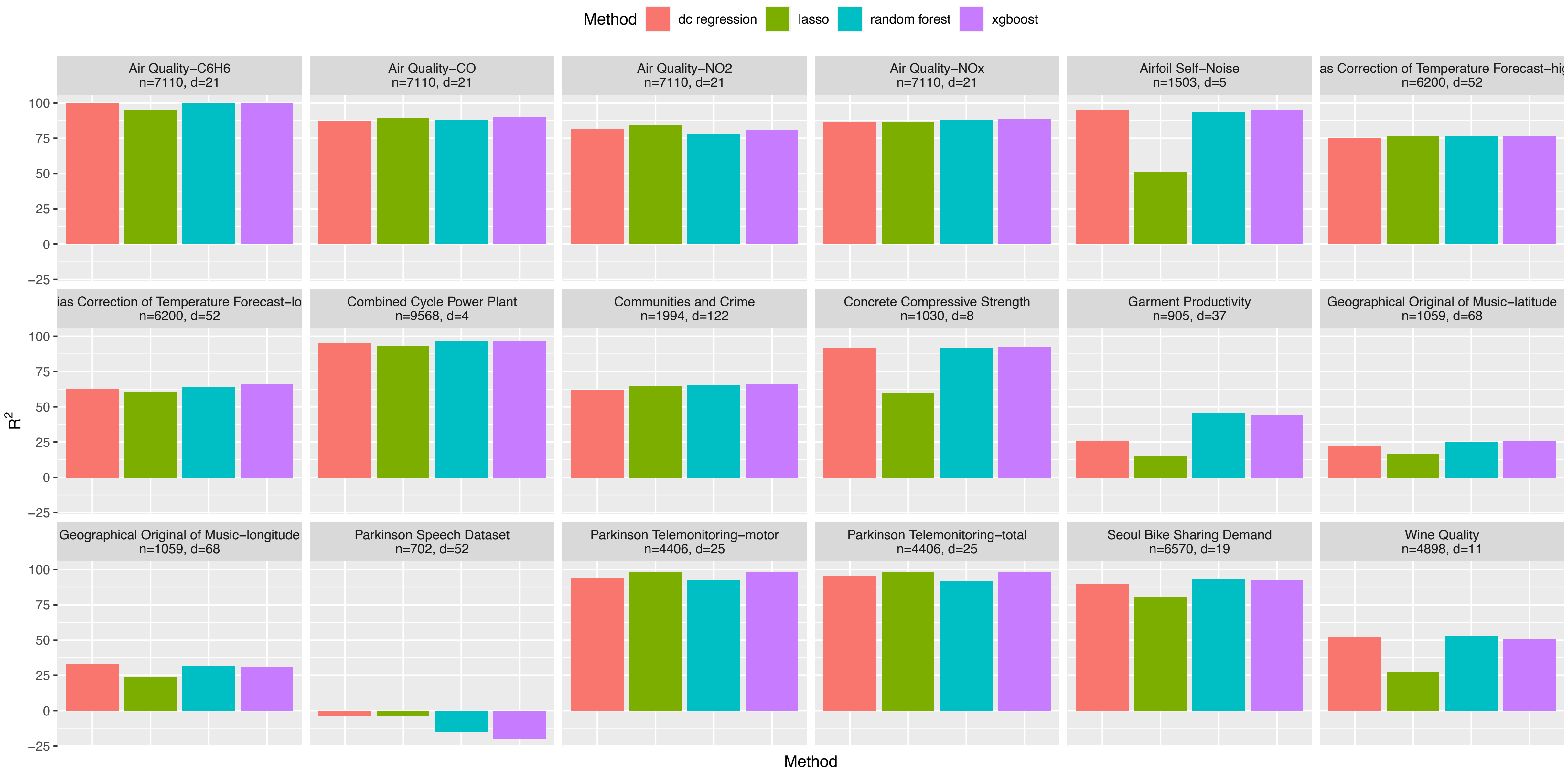
Experiments

Table 1. Comparison of Convex Regression run time against baseline on Synthetic Data

		Seconds	
n	d	Baseline (Siahkamari et al., 2020)	This Paper
1000	2	32.3	3.63
1000	4	30.8	3.75
1000	8	54.5	4.03
1000	16	112.3	4.12
1000	32	189.3	4.37

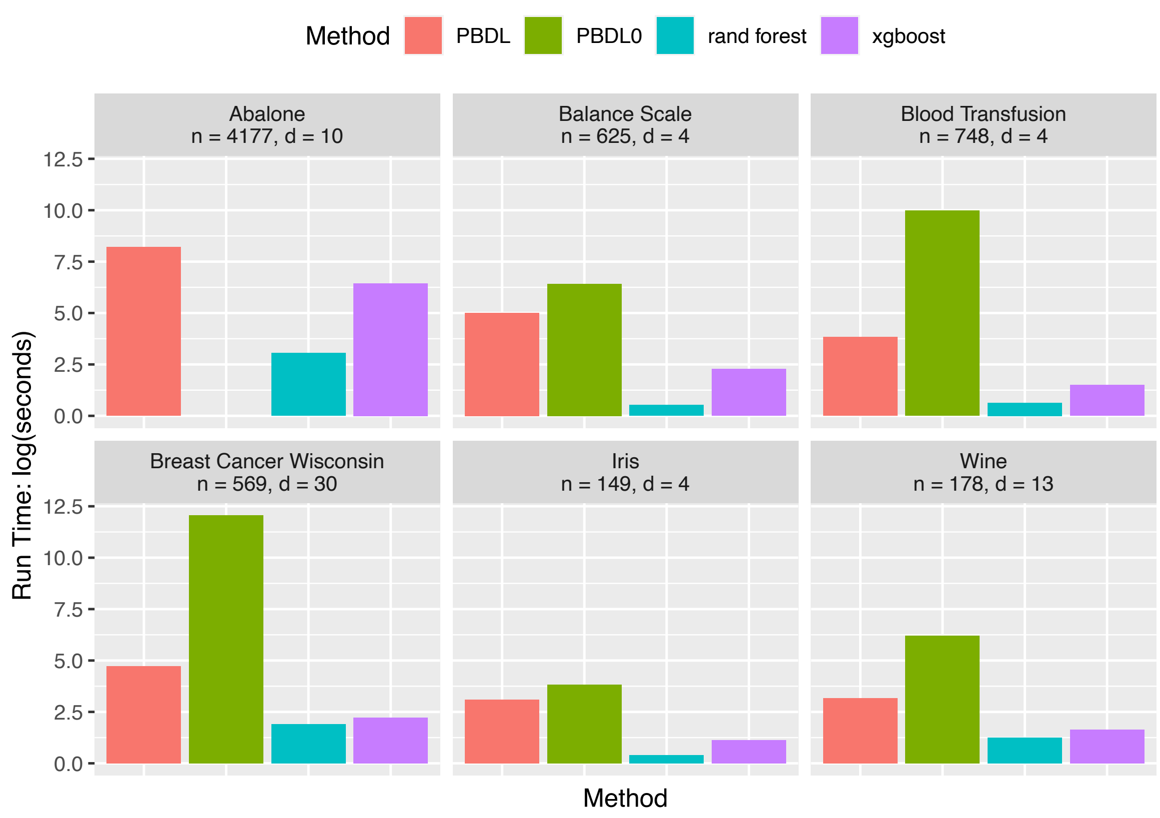
Faster Algorithms for Learning Convex functions

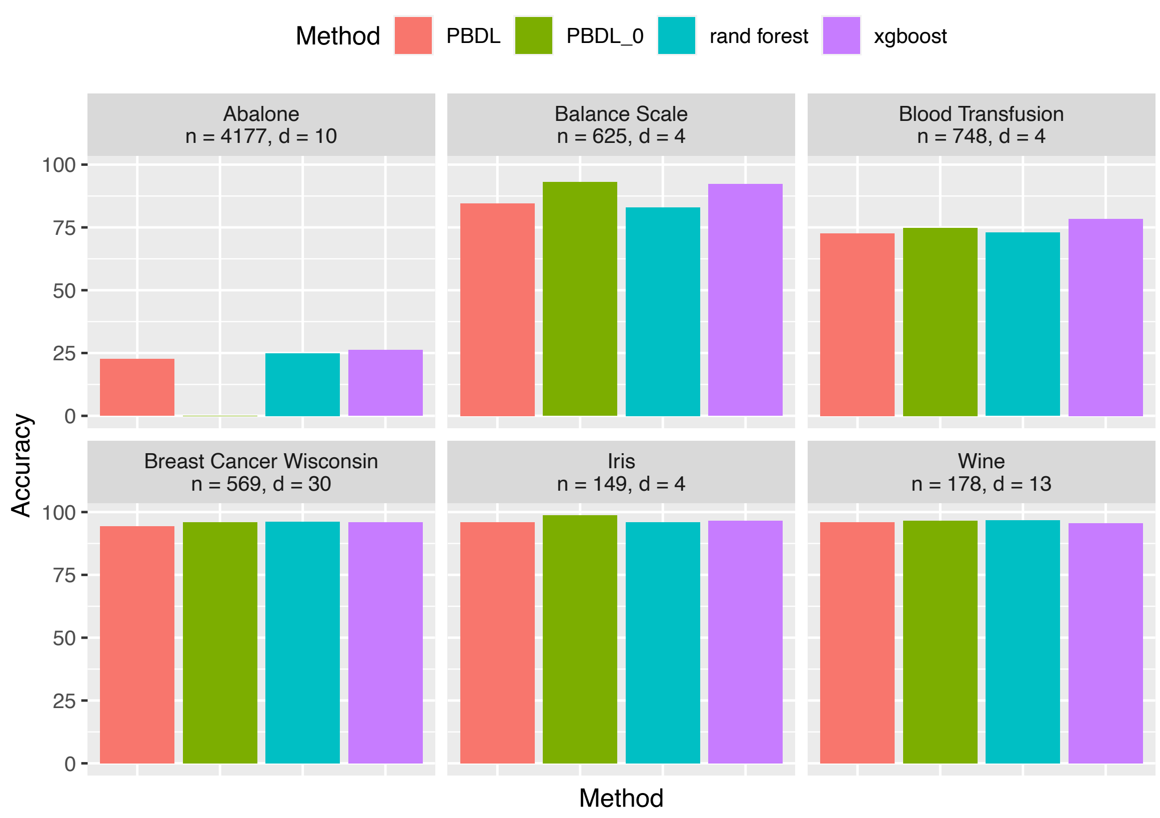
Experiments with DC regression - R^2



Faster Algorithms for Learning Convex functions

Experiments with Learning Bregman Divergences





Thank you

- Check out our paper: Faster Algorithms for Learning Convex function