Some Applications of Mixed Volumes in Data Science

Eliza O'Reilly

Caltech

Application #1: Prediction with Random Tessellation Forests

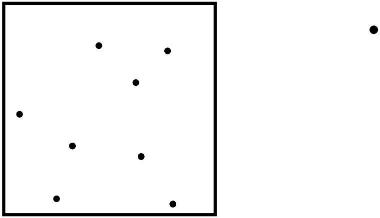
Joint work with Ngoc Mai Tran (UT Austin)

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^{n} \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Randomized Decision Trees:

- Recursively split data along random feature of input
- Induce a hierarchical axis-aligned partition of input space



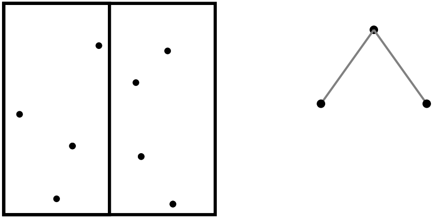
- **Random tree (regression) estimator**: average over cell/leaf

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n y_i \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}{\sum_{i=1}^n \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}$$

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Randomized Decision Trees:

- Recursively split data along random feature of input
- Induce a hierarchical axis-aligned partition of input space



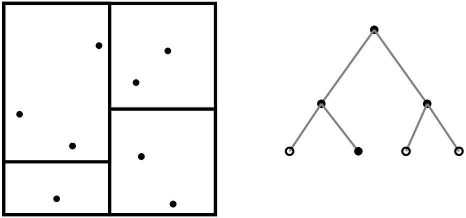
- **Random tree (regression) estimator**: average over cell/leaf

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n y_i \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}{\sum_{i=1}^n \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}$$

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Randomized Decision Trees:

- Recursively split data along random feature of input
- Induce a hierarchical axis-aligned partition of input space



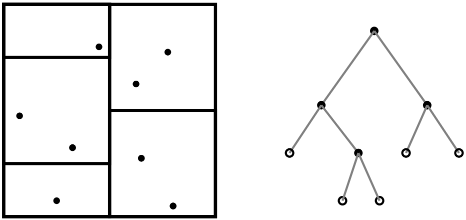
- **Random tree (regression) estimator**: average over cell/leaf

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n y_i \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}{\sum_{i=1}^n \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}$$

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Randomized Decision Trees:

- Recursively split data along random feature of input
- Induce a hierarchical axis-aligned partition of input space



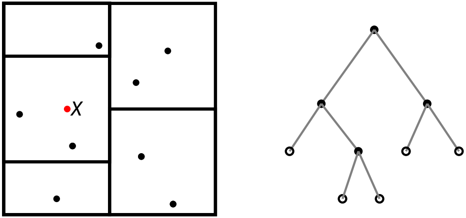
- **Random tree (regression) estimator**: average over cell/leaf

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n y_i \mathbb{1}\{x_i \text{ in same cell as } x\}}{\sum_{i=1}^n \mathbb{1}\{x_i \text{ in same cell as } x\}}$$

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Randomized Decision Trees:

- Recursively split data along random feature of input
- Induce a hierarchical axis-aligned partition of input space



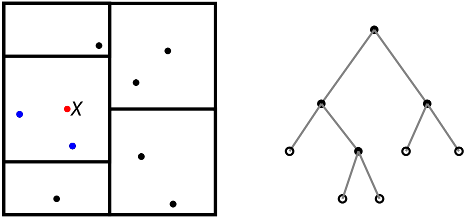
- **Random tree (regression) estimator**: average over cell/leaf

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n y_i \mathbb{1}\{x_i \text{ in same cell as } x\}}{\sum_{i=1}^n \mathbb{1}\{x_i \text{ in same cell as } x\}}$$

Goal: Given example input-output pairs $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$, obtain estimator $\hat{f}_n$ to **predict** output from new input x: $\hat{y} = \hat{f}_n(x)$

Randomized Decision Trees:

- Recursively split data along random feature of input
- Induce a hierarchical axis-aligned partition of input space



- **Random tree (regression) estimator**: average over cell/leaf

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n y_i \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}{\sum_{i=1}^n \mathbb{1}_{\{x_i \text{ in same cell as } x\}}}$$

Random Forest (RF) Estimator

- Average of M i.i.d. tree estimators

---

[1] [Ho, 1995; 1998; Amit and Geman, 1997; Breiman, 2001]
[2] [Caruana and Niculescu-Mizil, 2006; Fernandez-Delgado et al., 2014]
[3] [Scornet et al., 2015; Wager and Athey, 2018; Chi et al., 2020; Klusowski and Tian, 2022]
[4] [Breiman, 2004; Genuer, 2012]

Random Forest (RF) Estimator

- Average of M i.i.d. tree estimators

  - Original RF algorithm[1]: splits ***dependent*** on data
  - State-of-the-art empirical performance for many tasks[2]

[1] [Ho, 1995; 1998; Amit and Geman, 1997; Breiman, 2001]
[2] [Caruana and Niculescu-Mizil, 2006; Fernandez-Delgado et al., 2014]
[3] [Scornet et al., 2015; Wager and Athey, 2018; Chi et al., 2020; Klusowski and Tian, 2022]
[4] [Breiman, 2004; Genuer, 2012]

Random Forest (RF) Estimator

- Average of M i.i.d. tree estimators

  - Original RF algorithm[1]: splits ***dependent*** on data
  - State-of-the-art empirical performance for many tasks[2]
  - Difficult to analyze; still a lack of theoretical understanding[3]

---

[1] [Ho, 1995; 1998; Amit and Geman, 1997; Breiman, 2001]
[2] [Caruana and Niculescu-Mizil, 2006; Fernandez-Delgado et al., 2014]
[3] [Scornet et al., 2015; Wager and Athey, 2018; Chi et al., 2020; Klusowski and Tian, 2022]
[4] [Breiman, 2004; Genuer, 2012]

Random Forest (RF) Estimator

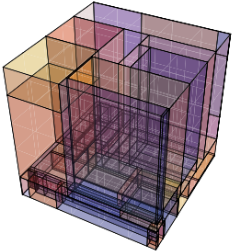
- Average of M i.i.d. tree estimators

  - Original RF algorithm[1]: splits ***dependent*** on data
  - State-of-the-art empirical performance for many tasks[2]
  - Difficult to analyze; still a lack of theoretical understanding[3]
  - Purely RF variants[4]: splits ***independent*** of data

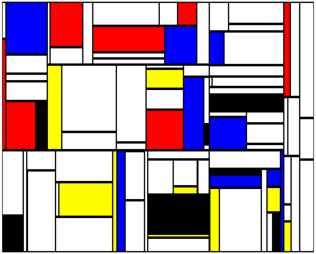
---

[1] [Ho, 1995; 1998; Amit and Geman, 1997; Breiman, 2001]
[2] [Caruana and Niculescu-Mizil, 2006; Fernandez-Delgado et al., 2014]
[3] [Scornet et al., 2015; Wager and Athey, 2018; Chi et al., 2020; Klusowski and Tian, 2022]
[4] [Breiman, 2004; Genuer, 2012]

Mondrian process



- Introduced by Roy and Teh in 2008
- Stochastic process that recursively builds an axis-aligned hierarchical partition in $\mathbb{R}^d$



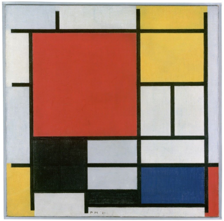


Figure: Piet Mondrian (1921).

Mondrian process construction in $\mathbb{R}^d$

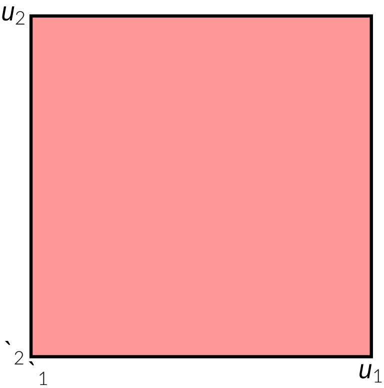
1. Fix **lifetime parameter** > 0

2. Draw

$$Exp\left[\sum_{i=1}^{d}(u_i \quad \ell_i)\right]$$

3. IF $>$ stop,
   ELSE sample a split:
   - *Dimension*: j with probability proportional to $u_j \quad \ell_j$
   - *Location*: uniform on $[\ell_j; u_j]$

4. Recurse independently on each subrectangle with lifetime $\quad 4$



Mondrian process construction in $\mathbb{R}^d$

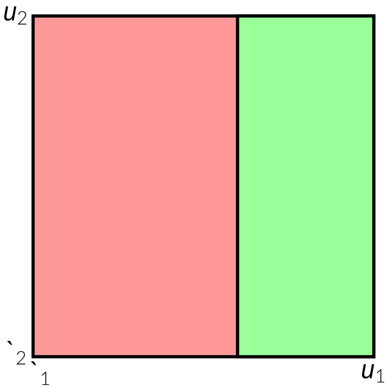
1. Fix **lifetime parameter** > 0

2. Draw

$$Exp\left[\sum_{i=1}^{d}(u_i \quad `_i)\right]$$

3. IF $>$ stop,
   ELSE sample a split:
   - *Dimension*: j with probability proportional to $u_j \quad `_j$
   - *Location*: uniform on $[`_j; u_j]$

4. Recurse independently on each subrectangle with lifetime 4



Mondrian process construction in $\mathbf{R}^d$

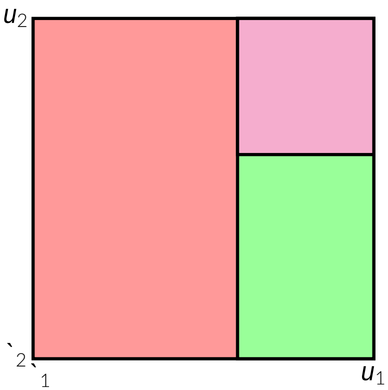
1. Fix **lifetime parameter** > 0

2. Draw

$$Exp\left[\sum_{i=1}^{d}(u_i \quad \ell_i)\right]$$

3. IF $>$ stop,
   ELSE sample a split:
   - *Dimension*: j with probability proportional to $u_j \quad \ell_j$
   - *Location*: uniform on $[\ell_j; u_j]$

4. Recurse independently on each subrectangle with lifetime $\quad 4$



Mondrian process construction in $\mathbf{R}^d$

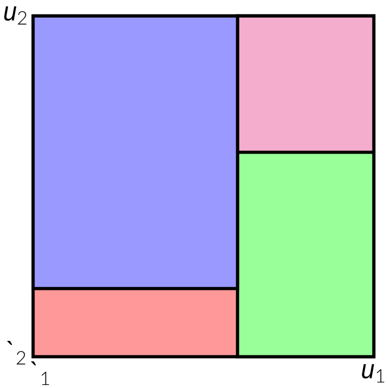
1. Fix **lifetime parameter** $\quad > 0$

2. Draw

$$Exp\left[\sum_{i=1}^{d}(u_i \quad `_i)\right]$$

3. IF $\quad > \quad$ stop,
   ELSE sample a split:
   - *Dimension*: j with probability proportional to $u_j \quad `_j$
   - *Location*: uniform on $[`_j; u_j]$

4. Recurse independently on each subrectangle with lifetime $\qquad 4$

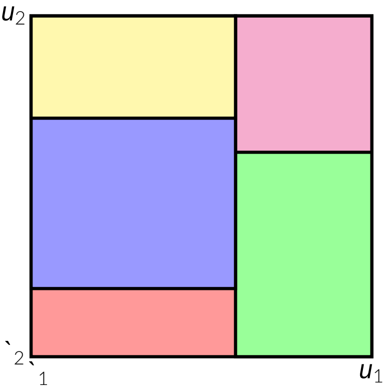


Mondrian process construction in $\mathbf{R}^d$

1. Fix **lifetime parameter** > 0
2. Draw

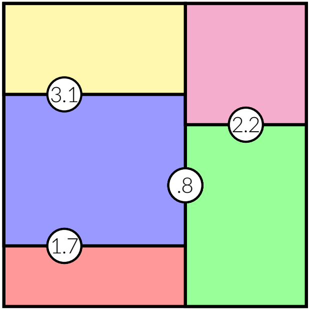
$$Exp \left[\sum_{i=1}^{d} (u_i \quad \ell_i) \right]$$

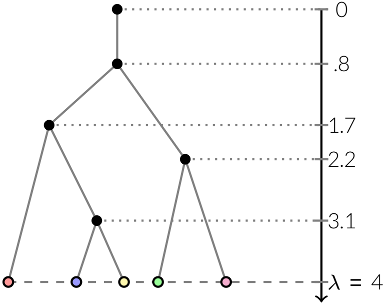
3. IF > stop,
   ELSE sample a split:
   - ***Dimension***: ***j*** with probability proportional to $u_j \quad \ell_j$
   - ***Location***: uniform on $[\ell_j; u_j]$
4. Recurse independently on each subrectangle with lifetime *4*



Mondrian Random Forests

- Comparable empirical performance to RF for some tasks[5]
- Minimax rates under nonparametric assumptions in arbitrary dimension[6]





[5] [Lakshminarayanan, Roy, and Teh, 2014]
[6] [Mourtada, Gaïffas, Scornet, 2020]

Beyond axis-aligned partitions

- Non-axis-aligned splits can capture dependencies between features

- Non-axis-aligned RF variants show **improved empirical performance**[7]
- Lack of theoretical analysis, computational efficiency

[7] [Breiman, 2001; Fan, Li, and Sisson, 2019; Tomita et al., 2020]

Beyond axis-aligned partitions

- Non-axis-aligned splits can capture dependencies between features

- Non-axis-aligned RF variants show **improved empirical performance**[7]

- Lack of theoretical analysis, computational efficiency

Question: Is there a generalization of the Mondrian process with non-axis-aligned cuts?

[7] [Breiman, 2001; Fan, Li, and Sisson, 2019; Tomita et al., 2020]

Stable Under Iteration Processes

- **Yes!** Mondrian is special case of the **STIT process** in stochastic geometry
- Introduced by Nagel and Weiss in 2003
- Indexed by a **directional distribution** on S^{d-1}

[8][Ge, Wang, Teh, Wang, and Elliott, 2019]

Stable Under Iteration Processes

- **Yes!** Mondrian is special case of the **STIT process** in stochastic geometry
- Introduced by Nagel and Weiss in 2003
- Indexed by a **directional distribution** on S^{d-1}
- Improved empirical performance with uniform STIT over Mondrian[8]
- General cell shapes introduce computational and **theoretical** challenges

[8] [Ge, Wang, Teh, Wang, and Elliott, 2019]

4HEORETICAL #HALLENGE -INIMAX 2ATES

4HEORETICAL #HALLENGE -INIMAX 2ATES

- !SSUMPTION $\{(X_i; Y_i)\}_{I=1}^{N}$ I I D SAMPLES OF $(8;9) \in 7 \times R$ SUCH THAT

$$9 = F(8) + ";$$

WHERE $7 \subset R^D$ IS A COMPACT AND CONVEX WINDOW

- $\hat{F}_{;N;-}$ 34)4 FOREST ESTIMATOR OF SIZE - LIFETIME PARAMETER

4HEORETICAL #HALLENGE -INIMAX 2ATES

- !SSUMPTION $f(X_i; Y_i)g_{i=}^{N}$ I I D SAMPLES OF $(8;9)$ 2 7 R SUCH THAT

$$9 = F(8) + ";$$

WHERE 7 R^D IS A COMPACT AND CONVEX WINDOW

- $\hat{F}_{;N;-}$ 34)4 FOREST ESTIMATOR OF SIZE - LIFETIME PARAMETER
- 4HE QUALITY OF THE ESTIMATOR $\hat{F}_{;N;-}$ IS MEASURED BY THE QUADRATIC RISK

$$E[(\hat{F}_{;N;-}(8) \quad F(8)) \;]$$

4HEORETICAL #HALLENGE -INIMAX 2ATES

- !SSUMPTION $f(X_i; Y_i)g_{I=}^{N}$ I I D SAMPLES OF $(8;9)$ 2 7 R SUCH THAT

$$9 = F(8) + ";$$

WHERE 7 R^D IS A COMPACT AND CONVEX WINDOW

- $\hat{F}_{;N;-}$ 34)4 FOREST ESTIMATOR OF SIZE - LIFETIME PARAMETER
- 4HE QUALITY OF THE ESTIMATOR $\hat{F}_{;N;-}$ IS MEASURED BY THE QUADRATIC RISK

$$E[(\hat{F}_{;N;-}(8) \quad F(8)) \quad]$$

- 4HE MINIMAX RISK FOR A FUNCTION CLASS F IS

$$\min_{\hat{F}_N} \max_{F 2 F} E[(\hat{F}_N(8) \quad F(8)) \quad]$$

4HEORETICAL #HALLENGE -INIMAX 2ATES

4HEOREM 4RAN AND /

-)F F IS ,IPSCHITZ LETTING $N_N \asymp N^{=(D+\ )}$ GIVES

$$E[(\hat{F}_{N;N;-}(8) \; F(8))\ ] \;/\; N^{=(D+\ )}$$

4HEORETICAL #HALLENGE -INIMAX 2ATES

4HEOREM 4RAN AND /

I)F F IS ,IPSCHITZ LETTING N N =(D+) GIVES

$$E[(\hat{F}_{N;N;-}(8) \quad F(8))\ ] \quad / \quad N^{=(D+\ )}$$

II)F F IS C AND 8 HAS POSITIVE AND ,IPSCHITZ DENSITY LETTING N N =(D+) AND - N & N =(D+) GIVES

$$E[(\hat{F}_{N;N;-_N}(8) \quad F(8))\ ] \quad / \quad N^{=(D+\ )}$$

4HEORETICAL #HALLENGE -INIMAX 2ATES

4HEOREM 4RAN AND /

I)F F IS ,IPSCHITZ LETTING $N_N = N^{=(D+\ )}$ GIVES

$$E[(\hat{F}_{N;N;-}(8) \quad F(8)) \;] \quad / \quad N^{=(D+\ )}$$

II)F F IS C AND 8 HAS POSITIVE AND ,IPSCHITZ DENSITY LETTING $N_N = N^{=(D+\ )}$ AND $-_N \;\&\; N^{=(D+\ )}$ GIVES

$$E[(\hat{F}_{N;N;-_N}(8) \quad F(8)) \;] \quad / \quad N^{=(D+\ )}$$

- 34)4 RANDOM FORESTS ARE MINIMAX OPTIMAL FOR ,IPSCHITZ AND C FUNCTIONS

2ATES FOR MULTI INDEX MODELS

- 3UPPOSE FOR G: R^S ! R AND A_i 2 R^D I = ; : : : ; S

$$F(X) = G(hA\ ;Xi; : : : ; hA_S;Xi); \quad X2 \text{"}_D(\ ;2):$$

;,I &UKUMIZU ET AL $ALALYAN ET AL =

2ATES FOR MULTI INDEX MODELS

- 3UPPOSE FOR G:R^{S} ! R AND A$_{I}$ 2 R^{D} I = ;:::;S

$$F(X) = G(hA ;Xi;:::;hA_{S};Xi); \quad X2 "_{D}(;2):$$

- 3 := SPAN(A ;:::;A$_{S}$) R^{D} IS THE S DIMENSIONAL RELEVANT FEATURE SUBS

;,I &UKUMIZU ET AL $ALALYAN ET AL =

2ATES FOR MULTI INDEX MODELS

- 3UPPOSE FOR $G : R^S \rightarrow$ R AND A_I 2 R^D I = ; : : : ; S

$$F(X) = G(hA ; Xi ; : : : ; hA_S; Xi); \quad X2 \; "_D(; 2):$$

- 3 := SPAN(A ; : : : ; A_S) R^D IS THE S-DIMENSIONAL RELEVANT FEATURE SUBS
- ,ET $\hat{F}_{; N; -}$ BE A 34)4 FOREST ESTIMATOR WITH DIRECTIONAL DISTRIBU

$$_N = (\quad "_N) \; _3 + \; "_N \; _{3^?} \; ;$$

FOR $"_N$ 2 (;) WHERE SUPP($_3$) = 3 \ S^D SUPP($_3$) = $3^?$ \ S^D

;,I &UKUMIZU ET AL $ALALYAN ET AL =

2ATES FOR MULTI INDEX MODELS

- 3UPPOSE FOR $G : R^S \to R$ AND $A_I \in R^D$ I = ; : : : ; S

$$F(X) = G(hA ; Xi ; : : : ; hA_S ; Xi); \quad X \in " _D(; 2):$$

- 3 := SPAN(A ; : : : ; A_S) R^D IS THE S DIMENSIONAL RELEVANT FEATURE SUBS
- ,ET $\hat{F}_{; N; -}$ BE A 34)4 FOREST ESTIMATOR WITH DIRECTIONAL DISTRIBU

$$_N = (\quad "_N) \; _3 + "_N \; _{3^?} ;$$

FOR $"_N \in (\; ; \;)$ WHERE SUPP($_3$) = 3 \ S^D SUPP($_{3^?}$) = $3^?$ \ S^D

4HEOREM

- I)F G IS ,IPSCHITZ LETTING N N =(S+) AND "N N =(S+) GIVES

$$E[(\hat{F}_{N;N;-} (8) \quad F(8)) \;] \quad / \quad N^{=(S+ \;)}$$

- II ! SIMILAR EXTENSION HOLDS IF ,SOS

;,I &UKUMIZU ET AL $ALALYAN ET AL =

0ROOF)DEA "IAS 6ARIANCE $ECOMPOSITION OF TI

- 7E HAVE THE FOLLOWING BIAS-VARIANCE DECOMPOSITION

$$E[(\hat{F}_{;N;}(8) \quad F(8))\ ] = E[(F(8) \quad F(8))\ ] + E[(F(8) \quad \hat{F}_{;N}(8))\ ];$$

WHERE

$$F(X) = E_8[F(8) j 8 2 : _X]; \quad X 2 7 ;$$

IS THE ORTHOGONAL PROJECTION OF $F 2 ,(7;)$ ONTO THE SUBSPACE OF FUN
THAT ARE CONSTANT WITHIN THE CELL OF THE 34)4 TESSELLATION

0ROOF)DEA "IAS 6ARIANCE $ECOMPOSITION OF T

- 7E HAVE THE FOLLOWING BIAS VARIANCE DECOMPOSITION

$$E[(\hat{F}_{;N}(8) \quad F(8))\ ] = E[(F(8) \quad \bar{F}(8))\ ] + E[(\bar{F}(8) \quad \hat{F}_{;N}(8))\ ];$$

WHERE

$$\bar{F}(X) = E_8[F(8) j 8 2 : X]; \quad X 2 7 ;$$

IS THE ORTHOGONAL PROJECTION OF F ONTO THE SUBSPACE OF FUN THAT ARE CONSTANT WITHIN THE CELL OF THE 34)4 TESSELLATION

- "IAS IS CONTROLLED BY THE DIAMETER OF THE CELL CONTAINING 8
- 6ARIANCE IS CONTROLLED BY THE EXPECTED NUMBER OF CELLS IN 7

3TATIONARY 34)4 4ESSELLATION ON R^D

- Y(; 7) 34)4 IN COMPACT AND CONVEX 7 R^D WITH LIFETIME
- #ONSISTENCY &OR 7 7

$$Y(; 7) \stackrel{D}{=} Y(; 7) \setminus 7$$

;.AGEL AND 7EISS =

3TATIONARY 34)4 4ESSELLATION ON R^D

- Y(; 7) 34)4 IN COMPACT AND CONVEX 7 R^D WITH LIFETIME
- #ONSISTENCY &OR 7 7

$$Y(\; ; \; 7 \;) \overset{D}{=} Y(\; ; \; 7 \;) \setminus \; 7$$

- 4HERE EXISTS A STATIONARY 34)4 TESSELLATION Y( ) ON R^D SUCH THAT
  - Y( ) \ 7 $\overset{D}{=}$ Y(; 7) FOR ALL COMPACT 7

;.AGEL AND 7EISS =

3TATIONARY 34)4 4ESSELLATION ON R^D

- Y(; 7) 34)4 IN COMPACT AND CONVEX 7 R^D WITH LIFETIME

- #ONSISTENCY &OR 7 7

$$Y(; 7) \overset{D}{=} Y(; 7) \setminus 7$$

- 4HERE EXISTS A STATIONARY 34)4 TESSELLATION Y( ) ON R^D SUCH THAT
  - Y( ) \ 7 $\overset{D}{=}$ Y(; 7) FOR ALL COMPACT 7
  - 3TABLE 5NDER)TERATION FOR ALL ≥

$$Y( ) \overset{D}{=} N(Y( ) \quad Y( )) ;$$

S TY Y := Y [S C2 CELLS(Y) (Y(C) \ C)
WHERE fY (C) : C2 Y g ARE I I D COPIES OF Y

  - 3CALING PROPERTY Y( ) $\overset{D}{=}$ Y( )

;.AGEL AND 7EISS =

#ELLS OF STATIONARY RANDOM TESSELLATIONS

- #ELLS OF Y( ) FORM A STATIONARY POINT PROCESS ON SPACE K OF COMPACT CONVEX POLYTOPES

#ELLS OF STATIONARY RANDOM TESSELLATIONS

- #ELLS OF Y( ) FORM A STATIONARY POINT PROCESS ON SPACE OF K COMPACT CONVEX POLYTOPES
- 4YPICAL CELL IS A CENTERED RANDOM POLYTOPE : SUCH THAT FOR ALL ! 2 B(K)

$$E\left[\sum_{\# 2 \, CELLS(Y)} f \, \# \; 2 \; ! \; g\right] = \frac{}{E[vol_D(:)]} E \int_{R^D} f \; : \; + \; Y 2 \; ! \; g \, dY$$

0ROOF)DEA 6ARIANCE "OUND

- 2ECALL 6ARIANCE IS CONTROLLED BY THE EXPECTED NUMBER OF CE
- ,ET7 R^D BE A COMPACT AND CONVEX SET
- ,ETY( ) BE A 34)4 TESSELLATION IN R^D WITH LIFETIME

,EMMA

,ET : BE THE TYPICAL CELL OF Y( ) 4HEN

$$E\left[\sum_{\#2 \text{ cells}(Y(\;))} f\#\backslash 7 \; 6 \; ;g\right] = \sum_{K=}^{D} \binom{D}{K} \frac{E[6(7 [K]; : [D \; K])]}{E[\mathrm{vol}_D(:)]} ;$$

WHERE E[6(7 [K]; : [D K])] := E[6(7 ; : : : ; 7 ; : : : : : : :)]

$$\underbrace{\qquad}_{K} \quad \underbrace{\qquad}_{D \; K}$$

0ROOF)DEA 6ARIANCE "OUND

- 2ECALL 6ARIANCE IS CONTROLLED BY THE EXPECTED NUMBER OF CE
- ,ET7 R^D BE A COMPACT AND CONVEX SET
- ,ETY( ) BE A 34)4 TESSELLATION IN R^D WITH LIFETIME

,EMMA

,ET : BE THE TYPICAL CELL OF Y( ). 4HEN

$$E\left[\sum_{\#2\ \mathrm{cells}(Y(\ ))} f\#\backslash 7 \ne ;g\right] = \sum_{k=}^{D} \binom{D}{K} {}_K\frac{E[6(7\ [K];:\ [D\ \ K])]}{E[\mathrm{vol}_D(:)]};$$

WHERE $E[6(7\ [K];:\ [D\ \ K])] := E[6(7\ \underbrace{;\ldots;7}_{K};\ \underbrace{:\ ;\ldots;:}_{D\ K})]$

- 0ROOF APPLY 3TEINERåS FORMULA TO $E\left[\sum_{\#2P(\ )} f\#\backslash 7 \ne ;g\right] = \frac{E[\mathrm{vol}_D(7\ \ :\ )]}{E[\mathrm{vol}_D(:\ )]}$

0ROOF)DEA 2ISK "OUND

- ,ET $\hat{F}_{;N;-}$ BE THE FOREST ESTIMATOR CORRESPONDING TO THE 34)4 (E 3)
- ,ET: THE ZERO CELL CELL CONTAINING THE ORIGIN AND ( ) OF :ET THE TYPICAL CELL

!SSUME F: 7 ! R IS, ,IPSCHITZ 4HEN

$$E[(\hat{F}_{;N;-}(8) \quad F(8))\ ]$$

$$\frac{, E[\mathrm{diam}(: \quad)\ ]}{} + \frac{(\quad k F k_1 \quad + \quad)}{N} \sum_{K=}^{D} \frac{D}{K} \; {}_K \frac{E[6(7\ [K]; :\ [D \quad K])]}{E[\mathrm{vol}_D(:\ )]}$$

- ,ETTING = $_N$ N $^{=(D+ \ )}$ GIVES THE MINIMAX RATE FOR ,IPSCHITZ FUN

0ROOF)DEA 2ISK "OUND FOR -ULTI)NDEX -ODEL

- 3UPPOSE FOR $G : R^S \to R$ AND $A_I \in R^D$ $I = ; : : : ; S$

$$F(X) := G(\langle A ; X \rangle ; : : : ; \langle A_S ; X \rangle)$$

- $3 := SPAN(A ; : : : ; A_S)$ R^D IS THE S-DIMENSIONAL RELEVANT FEATURE SUBS
- ,ET $\hat{F}_{; N}$ BE A 34)4 FOREST ESTIMATOR WITH DIRECTIONAL DISTRIBUT

$$_N = (\quad "_N) \; _3 + "_N \; _{3^?}$$

0ROOF)DEA 2ISK "OUND FOR -ULTI)NDEX -ODEL

- 3UPPOSE FOR G: R^S ! R AND A$_I$ 2 R^D I = ; : : : ; S

$$F(X) := G(hA ; Xi ; : : : ; hA_S ; Xi)$$

- 3 := SPAN(A ; : : : ; A$_S$) R^D IS THE S DIMENSIONAL RELEVANT FEATURE SUBS
- ,ET $\hat{F}_{; N}$ BE A 34)4 FOREST ESTIMATOR WITH DIRECTIONAL DISTRIBUT

$$_N = (\quad "_N) \; _3 + "_N \; _{3^?}$$

!SSUME G: R^S ! R IS , ,IPSCHITZ 4HEN

$$E[(\hat{F}_{; N}(8) \quad F(8))\ ] \quad \frac{,\ k!\ k_{OP}}{(\quad "_N)} E[diam(: \quad \backslash \ 3)\ ]$$

$$+ \frac{(\ kFk_1 \ + \quad)}{N} \quad {}^{D"D}_{N} \; {}^{S} \; \frac{D}{S} \; 2^D \; {}_D 6 (\; {}_3[S]; \; {}_{3^?} [D \quad S]) + O(\; {}^{D"D}_{N} \; {}^{S}) \quad :$$

- $_3$ $_{3^?}$ ARE CONVEX BODIES ZONOIDS DE NED BY $_3$ AND $_{3^?}$

3UMMARY AND FUTURE WORK

- 7E HAVE PROVED MINIMAX OPTIMAL RATES FOR A LARGE CLASS OF RANDO FOREST PARTITION ESTIMATORS WITH GENERAL SPLIT DIRECTIONS
- 4HEORY OF STATIONARY RANDOM TESSELLATIONS IS A POWERFUL AND FLEXIBLE FRAMEWORK FOR UNDERSTANDING AND DEVELOPING RANDOM PA
- 0ERFORMANCE DEPENDS ON GEOMETRY OF THE CELLS, E.G. MIXED VOLUMES OF TYPICAL CELL

3UMMARY AND FUTURE WORK

- 7E HAVE PROVED MINIMAX OPTIMAL RATES FOR A LARGE CLASS OF RANDO
  FOREST PARTITION ESTIMATORS WITH GENERAL SPLIT DIRECTIONS
- 4HEORY OF STATIONARY RANDOM TESSELLATIONS IS A POWERFUL AND FLEXIBLE
  FRAMEWORK FOR UNDERSTANDING AND DEVELOPING RANDOM PA
- 0ERFORMANCE DEPENDS ON GEOMETRY OF THE CELLS E.G. MIXED VOLUMES OF
  TYPICAL CELL

- (OW TO LEARN DIRECTIONAL DISTRIBUTION FROM DATA
- /THER APPLICATIONS CLUSTERING RANDOM FEATURE MODELS

Application #2:
Optimal regularizers for a data source

Joint work with Oscar Leong (Caltech), Yong Sheng Soh (National University of Singapore), and Venkat Chandrasekaran (Caltech)

Inverse Problems and regularization

- Goal is to recover signal ***x*** from:

$$y = A(x) + ″$$

  where A is a known forward map and $″$ is observation noise

- Problem may be ***ill-posed***

Inverse Problems and regularization

- Goal is to recover signal *x* from:

$$y = A(x) + ″$$

  where A is a known forward map and ″ is observation noise

- Problem may be ***ill-posed***

Functional Analytic Regularization:

- Recover *x* with the following optimization problem:

$$\operatorname*{argmin}_x \text{loss}(A(x); y) + \text{regularizer}(x)$$

- Regularizer function promotes structure in the solution

Variety of regularizers

Hand-crafted:

Sparsity is promoted by ℓ_1 norm (convex) and by ℓ_p norm for $p \in [0, 1)$ (non-convex)

$$\operatorname*{argmin}_x \operatorname{loss}(A(x), y) + \|x\|_p$$

Variety of regularizers

Hand-crafted:

Sparsity is promoted by ℓ_1 norm (convex) and by ℓ_p norm for $p \in [0, 1)$ (non-convex)

$$\operatorname*{argmin}_x \text{loss}(A(x), y) + \|x\|_p$$

Data-driven:

Dictionary Learning (convex)

- Learn $A \in \mathbb{R}^{d \times p}$ such that $x \approx Az$, where z is sparse

$$\|A^T x\|_1 \quad (\;\;) \quad \|x\|_{A(B_{\ell_1})}$$

Variety of regularizers

Hand-crafted:

Sparsity is promoted by ℓ_1 norm (convex) and by ℓ_p norm for $p \in [0, 1)$ (non-convex)

$$\operatorname*{argmin}_x \operatorname{loss}(A(x), y) + \|x\|_p$$

Data-driven:

Dictionary Learning (convex)

- Learn $A \in \mathbb{R}^{d \times p}$ such that $x \approx Az$, where z is sparse

$$\|A^T x\|_1 \quad ( ) \quad \|x\|_{A(B_{\ell_1})}$$

Generative models (non-convex)

- Neural network based regularization

Main question considered in this work

Which regularizer should one choose?

Main question considered in this work

Which regularizer should one choose?

- What is the optimal regularizer to impose for a given data source?
- Convex versus nonconvex?

Set-up and assumptions

- Let P be a probability distribution on $\mathbb{R}^d$ modeling a data source
- Define optimal regularizer f from a family F as a solution to:

$$\underset{f \in F}{\operatorname{argmin}} \ \mathbb{E}_P[f(x)]$$

- Conditions on $f \in F$:
  - Positively homogenous: $f(\lambda x) = \lambda f(x), \quad \lambda \geq 0$
  - $f \geq 0$ and continuous

Set-up and assumptions

- Let P be a probability distribution on $\mathbb{R}^d$ modeling a data source
- Define optimal regularizer f from a family $\mathcal{F}$ as a solution to:

$$\operatorname*{argmin}_{f \in \mathcal{F}} \mathbb{E}_P[f(x)]$$

- Conditions on $f \in \mathcal{F}$:
  - Positively homogenous: $f(\lambda x) = \lambda f(x)$, $\lambda \geq 0$
  - $f \geq 0$ and continuous

$f \in \mathcal{F} \iff f = \|\cdot\|_K$ $\|\cdot\|_K$ is the Minkowski functional of a **star body** K

Minkowski functional of a compact set $C \subset \mathbb{R}^d$:

$$\|x\|_C := \inf\{t > 0 : x \in tC\}$$

Star bodies and radial functions

- The **radial function** of a compact set $K \subset \mathbb{R}^d$ is defined by

$$\rho_K(x) := \sup\{t > 0 : tx \in K\} = \|x\|_K^{-1}$$

- A compact set $K \subset \mathbb{R}^d$ is a **star body** if ρ_K is continuous and it is ***starshaped*** (with respect to the origin)

$$x \in K \Rightarrow [0, x] \subset K$$

- Star bodies are uniquely determined by their radial functions

Unique optimal regularizer

Theorem (Leong, Soh, Chandrasekaran, O., 2022+)

Let P be a distribution on $\mathbb{R}^d$ with density p and assume $\mathbb{E}_P[\|x\|_{\ell_2}] < \infty$.

Unique optimal regularizer

Theorem (Leong, Soh, Chandrasekaran, O., 2022+)

Let P be a distribution on $\mathbb{R}^d$ with density p and assume $\mathbb{E}_P[\|x\|_{\ell_2}] < \infty$. Suppose the following function is continuous:

$$\rho_P(u) := \left(\int_0^\infty r^d p(ru) \mathrm{d}r \right)^{1/(d+1)}, \quad u \in \mathbb{S}^{d-1}. \tag{1}$$

Unique optimal regularizer

Theorem (Leong, Soh, Chandrasekaran, O., 2022+)

Let P be a distribution on $\mathbb{R}^d$ with density p and assume $\mathbb{E}_P[\|x\|_{\ell_2}] < \infty$. Suppose the following function is continuous:

$$\rho_P(u) := \left(\int_0^\infty r^d p(ru) \mathrm{d}r \right)^{1/(d+1)}, \quad u \in S^{d-1}. \tag{1}$$

Then $\exists$ a unique star body L_P with radial function ρ_P, and

$$K_* := \mathrm{vol}_d(L_P)^{-1/d} L_P$$

is the unique solution to

$$\operatorname*{argmin}_{K \in S^d : \mathrm{vol}_d(K) = 1} \mathbb{E}_P[\|x\|_K]$$

Unique optimal regularizer

Theorem (Leong, Soh, Chandrasekaran, O., 2022+)

Let P be a distribution on $\mathbb{R}^d$ with density p and assume $\mathbb{E}_P[\|x\|_{\ell_2}] < \infty$. Suppose the following function is continuous:

$$\rho_P(u) := \left(\int_0^\infty r^d p(ru) \mathrm{d}r \right)^{1/(d+1)}, \quad u \in S^{d-1}. \tag{1}$$

Then $\exists$ a unique star body L_P with radial function ρ_P, and

$$K_* := \mathrm{vol}_d(L_P)^{-1/d} L_P$$

is the unique solution to

$$\underset{K \in \mathcal{S}^d : \mathrm{vol}_d(K) = 1}{\operatorname{argmin}} \mathbb{E}_P[\|x\|_K]$$

- If L_P is convex, then the optimal regularizer is convex!

Examples

(i) Densities induced by star bodies:

$$p(x) = \quad (kxk_L) \quad) \quad L_P = c \quad L$$

Examples

(i) Densities induced by star bodies:

$$p(x) = \quad (kxk_L) \quad) \quad L_P = c \quad L$$

(ii) Gaussian Mixtures:

$$P = \frac{1}{2} N(0; \quad {}_1) + \frac{1}{2} N(0; \quad {}_2)$$

where $\quad_1 := [1; 0; 0; "] \; 2 \; \mathrm{R}^{2 \quad 2}$ and $\quad_2 := ["; 0; 0; 1] \; 2 \; \mathrm{R}^{2 \quad 2}$

Figure: Plots of L_P for $" = 0{:}1$ (left) and $" = 0{:}01$ (right).

Proof

Goal: Characterize unique solution to

$$\operatorname*{argmin}_{K \in S^d : \mathrm{vol}_d(K)=1} \mathbb{E}_P[\|x\|_K]$$

By change to polar coordinates,

$$\mathbb{E}_P[\|x\|_K] = \int_{\mathbb{R}^d} \|x\|_K p(x)\mathrm{d}x = \int_{\mathbb{S}^{d-1}} \|u\|_K \int_0^\infty r^d p(ru)\mathrm{d}r\mathrm{d}u$$

$$= \int_{\mathbb{S}^{d-1}} \rho_K(u)^{-1} \rho_P(u)^{d+1}\mathrm{d}u := d\tilde{V}_{-1}(K; L_P)$$

Proof

Goal: Characterize unique solution to

$$\operatorname*{argmin}_{K \in \mathcal{S}^d : \mathrm{vol}_d(K) = 1} \mathbb{E}_P[\|x\|_K]$$

By change to polar coordinates,

$$\begin{aligned}
\mathbb{E}_P[\|x\|_K] &= \int_{\mathbb{R}^d} \|x\|_K p(x) \mathrm{d}x = \int_{\mathbb{S}^{d-1}} \|u\|_K \int_0^\infty r^d p(ru) \mathrm{d}r \mathrm{d}u \\
&= \int_{\mathbb{S}^{d-1}} \rho_K(u)^{-1} \rho_P(u)^{d+1} \mathrm{d}u := d\tilde{V}_{-1}(K, L_P)
\end{aligned}$$

Theorem (Dual Mixed Volume Inequality (Lutwak, 1975))

For star bodies K and L,

$$\tilde{V}_{-1}(K, L)^d \geq \mathrm{vol}_d(K)^{-1} \mathrm{vol}_d(L)^{d+1},$$

and equality hold if and only if L and K are dilates, i.e. $L = \lambda K$ *for some* $\lambda > 0$

Summary and future work

- Dual Brunn-Minkowski theory provides tools for characterizing optimal functional for imposing structure on a dataset for inverse problems
- Other results: convergence of empirical minimizers and generalization error bounds

Summary and future work

- Dual Brunn-Minkowski theory provides tools for characterizing optimal functional for imposing structure on a dataset for inverse problems
- Other results: convergence of empirical minimizers and generalization error bounds

- How do optimal regularizers perform in downstream tasks?
- How to efficiently compute the optimal regularizer?

Papers

“Minimax Rates for High-Dimensional Random Tessellation Forests”
Joint with Ngoc Mai Tran. https://arxiv.org/abs/2109.10541

“Optimal Convex and Nonconvex Regularizers for a Data Source”
Joint with Oscar Leong, Yong Sheng Soh, and Venkat Chandrasekaran. In preparation.

Thank you!