Generative Trees: Adversarial and Copycat

Richard Nock

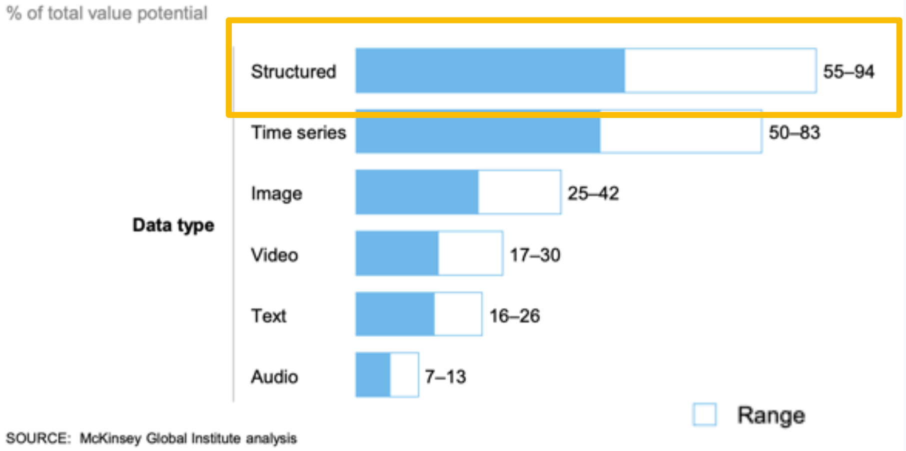
Mathieu Guillame-Bert

Google Research

{richardnock,gbm}@google.com

Why this work ?

Tabular data: important but scarce generative



Chui *et al.*, *Notes from the AI frontier*, McKinsey, 2018

- Modern generative techniques = Neural Networks (NN) / Deep Learning (DL) based
- On supervised learning side, the best techniques are (still) ***not*** DL-based but tree-based ; competing requires sophisticated+ DL techs
- “Lack of novelty” in state of the art (SOTA) modern generative approaches for tabular data
- Unconvincing results for DL + tabular pipelines

Camino *et al.*, ICBINB@NeurIPS’20

Google Research

Tabular data, *Supervised*

Losses: **proper**

Savage, JASA'71

Models: **tree-based**

Breiman *et al.* '84

Algorithms: **boosting**

Kearns & Mansour, STOC'96

Google Research

Tabular data, *Supervised*

Losses: **proper** →

Savage, JASA'71

Models: **tree-based** →

Breiman *et al.* '84

Algorithms: **boosting** →

Kearns & Mansour, STOC'96

This paper, *generative*, tabular data

Background: GAN game

Losses: designed from ***discriminator*** & in the **proper** framework

Models: ***tree-based***

Algorithms: **boosting**

↳ adversarial

↳ ***copycat***

Loss functions

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

GAN framework in a 

Measure-based loss, crafted from *generator*

(real) ↓ P, (fake) ↑ N

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

Google Research

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

GAN framework in a 

Measure-based loss, crafted from *generator*

(real) → P, (fake) → N

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}}\left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ **G**)

Nowozin *et al.*, NeurIPS'16

Google Research

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Our framework

Properness: $\mathbb{P}[\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=-1]$

↳ *Binary task* $\mathrm{B} \doteq (\pi, \mathrm{P}, \mathrm{N})$

↳ *Mixture* $\mathrm{M} \doteq \pi \cdot \mathrm{P} + (1-\pi) \cdot \mathrm{N}$

Google Research

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ **G**)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Our framework

Properness: $\mathbb{P}[\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=-1]$

↳ *Binary task* $\mathrm{B} \doteq (\pi, \mathrm{P}, \mathrm{N})$

↳ *Mixture* $\mathrm{M} \doteq \pi \cdot \mathrm{P} + (1-\pi)\cdot \mathrm{N}$

↳ discriminator learns a *posterior* $\eta : \mathcal{X} \rightarrow [0,1]$

$\hat{\mathbb{P}}[\mathrm{Y}=1|\mathrm{X}]$

↳ "*Ideal*" posterior computes $\mathbb{P}[\mathrm{Y}=1|\mathrm{X}]$

$$\eta^\star = \pi \cdot \frac{\mathrm{dP}}{\mathrm{dM}}$$ **Bayes posterior**

Google Research

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real) (fake)

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

f-divergence

"Information of Binary Task"

↳ variational formulation

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}}\left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

H + G

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Properness:

↳ *Binary task* $\mathrm{B} \doteq (\pi, \mathrm{P}, \mathrm{N})$ — with $\pi = \mathbb{P}[\mathrm{Y}=1]$, $\mathrm{P} = \mathbb{P}[\mathrm{X}|\mathrm{Y}=1]$, $\mathrm{N} = \mathbb{P}[\mathrm{X}|\mathrm{Y}=-1]$

↳ *Mixture* $\mathrm{M} \doteq \pi \cdot \mathrm{P} + (1-\pi)\cdot \mathrm{N}$

↳ discriminator learns a *posterior* $\eta : \mathcal{X} \rightarrow [0,1]$ ($\hat{\mathbb{P}}[\mathrm{Y}=1|\mathrm{X}]$)

↳ *"Ideal"* posterior computes $\mathbb{P}[\mathrm{Y}=1|\mathrm{X}]$

$$\eta^\star = \pi \cdot \frac{\mathrm{dP}}{\mathrm{dM}}$$ **Bayes posterior**

↳ a **loss** can be decomposed in *two partial losses*

$$\ell(y,u) \doteq [\![y=1]\!]\cdot \ell_1(u) + [\![y=-1]\!]\cdot \ell_{-1}(u)$$

u: estimated posterior in [0,1]

y: true label / class in {-1,1}

Google Research

GAN framework in a 🥜 → Our framework

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ **G**)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Properness: $\mathbb{P}[\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=-1]$

↳ *Binary task* $\mathrm{B} \doteq (\pi, \mathrm{P}, \mathrm{N})$

↳ *Mixture* $\mathrm{M} \doteq \pi\cdot\mathrm{P} + (1-\pi)\cdot\mathrm{N}$

↳ discriminator learns a *posterior* $\eta : \mathcal{X} \rightarrow [0,1]$

$\hat{\mathbb{P}}[\mathrm{Y}=1|\mathrm{X}]$

↳ "*Ideal*" posterior computes $\mathbb{P}[\mathrm{Y}=1|\mathrm{X}]$

$$\eta^\star = \pi\cdot\frac{\mathrm{dP}}{\mathrm{dM}}$$ **Bayes posterior**

↳ a **loss** can be decomposed in *two partial losses*

$$\ell(y,u) \doteq [\![y=1]\!]\cdot\ell_1(u) + [\![y=-1]\!]\cdot\ell_{-1}(u)$$

estimated posterior in [0,1]

true label / class in {-1,1}

↳ a loss is *symmetric* iff $\ell_1(u) = \ell_{-1}(1-u)$

Google Research

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Our framework

Properness: $\mathbb{P}[\mathsf{Y} = 1]$ $\mathbb{P}[\mathsf{X}|\mathsf{Y} = 1]$ $\mathbb{P}[\mathsf{X}|\mathsf{Y} = -1]$

↳ *Binary task* $\mathrm{B} \doteq (\pi, \mathrm{P}, \mathrm{N})$

↳ *Mixture* $\mathrm{M} \doteq \pi \cdot \mathrm{P} + (1 - \pi) \cdot \mathrm{N}$

↳ discriminator learns a *posterior* $\eta : \mathcal{X} \rightarrow [0, 1]$

$\hat{\mathbb{P}}[\mathsf{Y} = 1|\mathsf{X}]$

↳ *"Ideal"* posterior computes $\mathbb{P}[\mathsf{Y} = 1|\mathsf{X}]$

$$\eta^\star = \pi \cdot \frac{\mathrm{dP}}{\mathrm{dM}}$$ **Bayes posterior**

↳ a **loss** can be decomposed in *two partial losses*

$$\ell(y, u) \doteq [\![y = 1]\!] \cdot \ell_1(u) + [\![y = -1]\!] \cdot \ell_{-1}(u)$$

estimated posterior in [0,1]

true label / class in {-1,1}

↳ a loss is *symmetric* iff $\ell_1(u) = \ell_{-1}(1 - u)$

↳ a loss is ***strictly proper*** iff Bayes posterior solely realises the **inf** of

$$\underline{L}(p) \doteq \inf_u \mathbb{E}_{\mathsf{Y} \sim \mathrm{B}(p)}[\ell(\mathsf{Y}, u)]$$

Bayes risk (concave)

Google Research

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Properness: $\mathbb{P}[\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=1]$ $\mathbb{P}[\mathrm{X}|\mathrm{Y}=-1]$

↳ *Binary task* $\mathrm{B} \doteq (\pi, \mathrm{P}, \mathrm{N})$

↳ *Mixture* $\mathrm{M} \doteq \pi \cdot \mathrm{P} + (1-\pi)\cdot \mathrm{N}$

↳ discriminator learns a *posterior* $\eta : \mathcal{X} \rightarrow [0,1]$

$\hat{\mathbb{P}}[\mathrm{Y}=1|\mathrm{X}]$

↳ *"Ideal"* posterior computes $\mathbb{P}[\mathrm{Y}=1|\mathrm{X}]$

$$\eta^\star = \pi \cdot \frac{\mathrm{dP}}{\mathrm{dM}}$$ **Bayes posterior**

↳ a **loss** can be decomposed in *two partial losses*

$$\ell(y,u) \doteq [\![y=1]\!]\cdot \ell_1(u) + [\![y=-1]\!]\cdot \ell_{-1}(u)$$

estimated posterior in [0,1]

true label / class in {-1,1}

↳ a loss is *symmetric* iff $\ell_1(u) = \ell_{-1}(1-u)$

↳ a loss is ***strictly proper*** iff Bayes posterior solely realises the **inf** of

$$\underline{L}(p) \doteq \inf_u \mathbb{E}_{\mathrm{Y}\sim\mathrm{B}(p)}[\ell(\mathrm{Y},u)]$$

Bayes risk (concave)

log-, square-, Matusita, etc.

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_\mathrm{P}[\tilde{h}(\mathrm{X})] - \mathbb{E}_\mathrm{N}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Google Research

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_\mathrm{P}[\tilde{h}(\mathrm{X})] - \mathbb{E}_\mathrm{N}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ **G**)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Our framework

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

↳ posterior $\tilde{\eta}$ is said **calibrated** iff satisfies

$$\tilde{\eta} = \pi \cdot \frac{\mathrm{dP}_{\tilde{\eta}}}{\mathrm{dM}_{\tilde{\eta}}}$$

σ-algebra coarsened to the level sets of $\tilde{\eta}$

↳ ex: the prior π, Bayes posterior $\eta^\star$ are calibrated

↳ **any decision tree** (w/ empirical posterior prediction at the leaves) **is calibrated**

Google Research

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

↳ posterior $\tilde{\eta}$ is said **calibrated** iff satisfies

$$\tilde{\eta} = \pi \cdot \frac{\mathrm{dP}_{\tilde{\eta}}}{\mathrm{dM}_{\tilde{\eta}}}$$

σ-algebra coarsened to the level sets of $\tilde{\eta}$

↳ ex: the prior π, Bayes posterior $\eta^\star$ are calibrated

↳ **any decision tree** (w/ empirical posterior prediction at the leaves) **is calibrated**

↳ for any calibrated $\tilde{\eta}$, its *statistical information* is

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim\mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

CART, C4.5, etc. (splitting criterion)

Google Research

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Our framework

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim\mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

Google Research

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real) ↓ (fake) ↑

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right)\mathrm{dN}$$

f-divergence

"Information of Binary Task"

↳ variational formulation

$$\mathbb{I}_f(\mathrm{P},\mathrm{N}) \geq \sup_{\tilde{h}} \left\{\underbrace{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})]}_{\mathrm{H}} + \underbrace{- \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]}_{\mathrm{G}}\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ **G**)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim \mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

Theorem: for any calibrated $\tilde{\eta}$ and any *strictly proper symmetric and differentiable* loss ℓ

$$\mathbb{I}_{f^\pi}(\mathrm{P}_{\tilde{\eta}}, \mathrm{N}_{\tilde{\eta}}) = \mathrm{H} + \mathrm{G} = \Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}})$$

Not shown for readability

+ *if* density ratio fct $\ell^{\mathrm{DR}}_{-1}(\rho) \doteq \ell_{-1}\left(\frac{1}{1+\rho}\right)$ cvx, *then*

$$\mathrm{G} \leq \underline{L}(\pi) - (1-\pi)\cdot \ell_{-1}\left(\frac{\pi}{1+(1-\pi)\cdot \chi^2(\mathrm{N}_{\tilde{\eta}} \| \mathrm{P}_{\tilde{\eta}})}\right)$$

chi square

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real) ↓

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

↑ (fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_\mathrm{P}[\tilde{h}(\mathrm{X})] - \mathbb{E}_\mathrm{N}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ **H** + **G**)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ **G**)

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X} \sim \mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

Theorem: for an[y] ...

proper symmetri[c] ...

$$\mathbb{I}_{f^\pi}(\mathrm{P}_{\tilde{\eta}}, \mathrm{N}_{\tilde{\eta}})$$

True for all tested losses; proof of partial ppty in general case

\+ *if* density ratio fct $\ell^{\mathrm{DR}}_{-1}(\rho) \doteq \ell_{-1}\left(\frac{1}{1+\rho}\right)$ cvx, *then*

$$\mathbf{G} \leq \underline{L}(\pi) - (1-\pi) \cdot \ell_{-1}\left(\frac{\pi}{1 + (1-\pi) \cdot \chi^2(\mathrm{N}_{\tilde{\eta}} \| \mathrm{P}_{\tilde{\eta}})}\right)$$

chi square

Google Research

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variatio

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase the IBT by discriminating real vs fake (w/ H + G)

↳ *generator* = N, seeks to decrease the IBT by generating fake data that looks like real (w/ G)

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim\mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

trictly

s ℓ

$$\underline{L}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}})$$

Not shown for readability

+ *if* density ratio fct $\ell^{\mathrm{DR}}_{-1}(\rho) \doteq \ell_{-1}\left(\frac{1}{1+\rho}\right)$ cvx, *then*

$$\mathrm{G} \leq \underline{L}(\pi) - (1-\pi)\cdot\ell_{-1}\left(\frac{\pi}{1+(1-\pi)\cdot\chi^2(\mathrm{N}_{\tilde{\eta}}\|\mathrm{P}_{\tilde{\eta}})}\right)$$

chi square

Summary

Google Research

Nowozin *et al.*, NeurIPS'16, Reid & Williamson, JMLR'11

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

H + G

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase

... ake (w/ **H** + **G**)

... se the IBT by

... e real (w/ **G**)

... amson, JMLR'11

Gets the discriminator's loss from the generator's

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim\mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

Theorem: for any calibrated $\tilde{\eta}$ and any *strictly proper symmetric and differentiable* loss ℓ

$$\mathbb{I}_{f^\pi}(\mathrm{P}_{\tilde{\eta}}, \mathrm{N}_{\tilde{\eta}}) = \mathbf{H} + \mathbf{G} = \Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}})$$

Not shown for readability

+ *if* density ratio fct $\ell^{\mathrm{DR}}_{-1}(\rho) \doteq \ell_{-1}\left(\frac{1}{1+\rho}\right)$ cvx *then*

$$\ldots \frac{\pi}{\ldots} \ldots \pi) \cdot \chi^2(\mathrm{N}_{\tilde{\eta}} \| \mathrm{P}_{\tilde{\eta}})$$

chi square

Gets the generator's loss from the discriminator's

GAN framework in a  → Our framework

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{ \mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})] \right\}$$

Not an equa[…] in general

↳ *discrim[…]ator* hidden in $\tilde{h}$, seeks to increase

[…]ake (w/ **H** + **G**)

[…]se the IBT by

[…]e real (w/ **G**)

[…]amson, JMLR'11

Loose approximation (inequality) via variational formulation

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim\mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

Theorem: for any calibrated $\tilde{\eta}$ and any *strictly proper symmetric and differentiable* loss ℓ

$$\mathbb{I}_{f^\pi}(\mathrm{P}_{\tilde{\eta}}, \mathrm{N}_{\tilde{\eta}}) = \mathbf{H} + \mathbf{G} = \Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}})$$

N[…] shown for readab[…]y

+ *if* density ratio fct $\ell^{\mathrm{DR}}_{-1}(\rho) \doteq \ell_{-1}\left(\frac{1}{1+\rho}\right)$ cvx, *then*

π

[…]$\pi) \cdot \chi^2(\mathrm{N}_{\tilde{\eta}} \| \mathrm{P}_{\tilde{\eta}})$

chi square

Tight characterisation (all equalities), and one loss "to train against them all": the chi square

[…]ogle Research

[…]: Adversarial and Copycat

GAN framework in a 

Measure-based loss, crafted from *generator*

(real)

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \doteq \int f\left(\frac{\mathrm{dP}}{\mathrm{dN}}\right) \mathrm{dN}$$

(fake)

f-divergence

"Information of Binary Task"

↳ variational formulation

H + G

$$\mathbb{I}_f(\mathrm{P}, \mathrm{N}) \geq \sup_{\tilde{h}} \left\{\mathbb{E}_{\mathrm{P}}[\tilde{h}(\mathrm{X})] - \mathbb{E}_{\mathrm{N}}[f^\star \circ \tilde{h}(\mathrm{X})]\right\}$$

Not an equality in general

↳ *discriminator* hidden in $\tilde{h}$, seeks to increase

ake (w/ **H** + **G**)

se the IBT by

e real (w/ **G**)

"No" assumption necessary

amson, JMLR'11

Our framework

Partial losses ℓ_1, ℓ_{-1}, Bayes posterior $\eta^\star$ & risk $\underline{L}$

For any calibrated $\tilde{\eta}$, its *statistical information*:

$$\Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}}) = \underline{L}(\pi) - \mathbb{E}_{\mathrm{X}\sim\mathrm{M}_{\tilde{\eta}}}[\underline{L}(\tilde{\eta}(\mathrm{X}))]$$

Theorem: for any calibrated $\tilde{\eta}$ and any *strictly proper symmetric and differentiable* loss ℓ

$$\mathbb{I}_{f^\pi}(\mathrm{P}_{\tilde{\eta}}, \mathrm{N}_{\tilde{\eta}}) = \mathbf{H} + \mathbf{G} = \Delta\underline{\mathbb{L}}(\tilde{\eta}, \mathrm{M}_{\tilde{\eta}})$$

Not shown for readability

+ *if* density ratio fct $\ell^{\mathrm{DR}}_{-1}(\rho) \doteq \ell_{-1}\left(\frac{1}{1+\rho}\right)$ cvx, *then*

$$\pi \quad \pi) \cdot \chi^2(\mathrm{N}_{\tilde{\eta}} \| \mathrm{P}_{\tilde{\eta}})$$

chi square

Discriminator calibrated

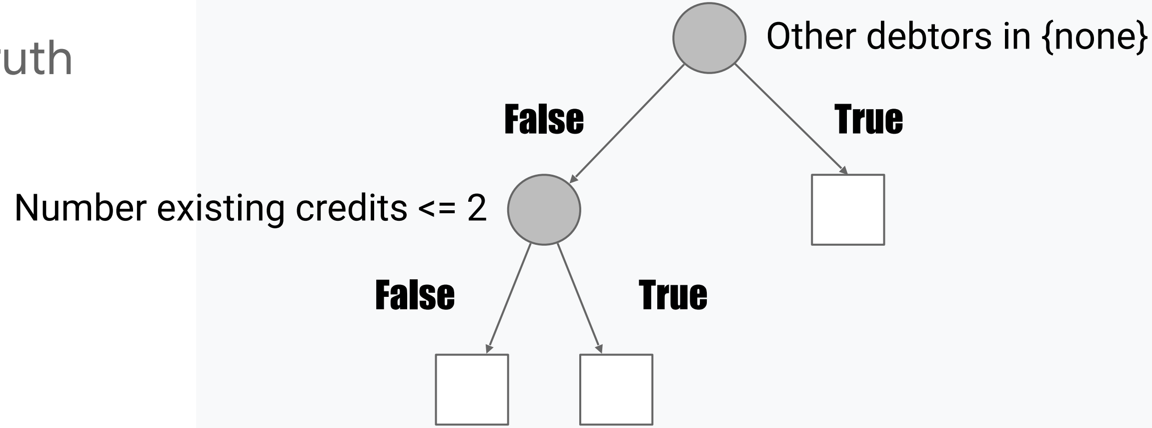
ogle Research

: Adversarial and Copycat

Models

Tree

A **tree** is a binary directed tree whose internal nodes are labeled with a test on an observation variable and outgoing arcs are labeled with truth values. Leaves are blank.

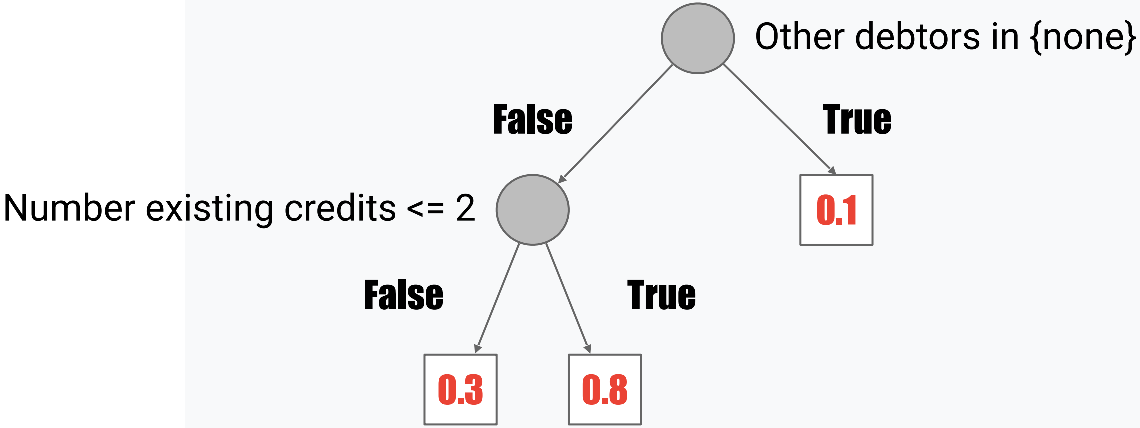


(Labelling from UCI German Credit)

Google Research

Decision Tree (DT)

A *decision tree h* is a tree in which leaves are labeled by values in [0,1]

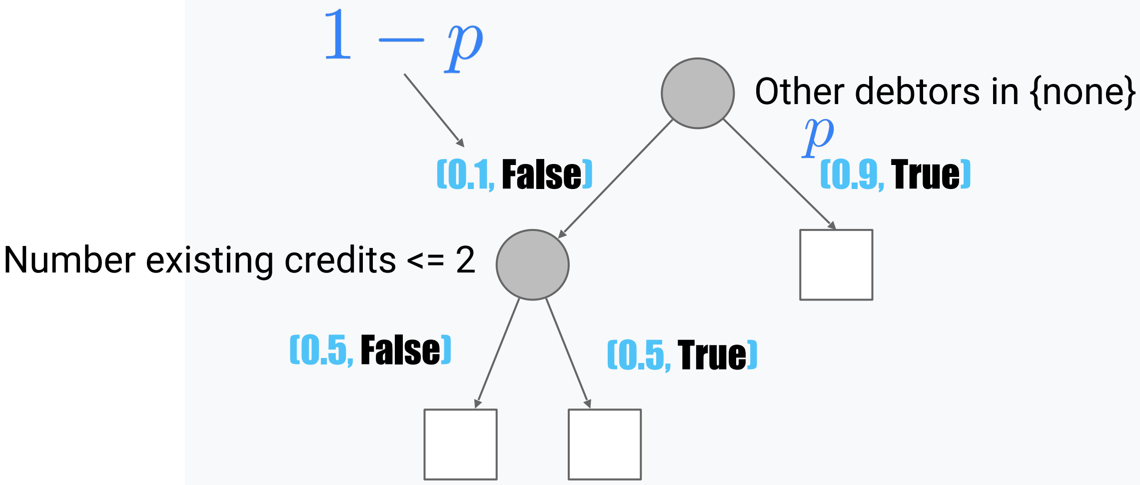


(Labelling from UCI German Credit)

Google Research

Generative Tree (GT)

A *generative tree G* is a tree in which outgoing arcs are labeled by Bernoulli trials $\mathrm{B}(p)$.



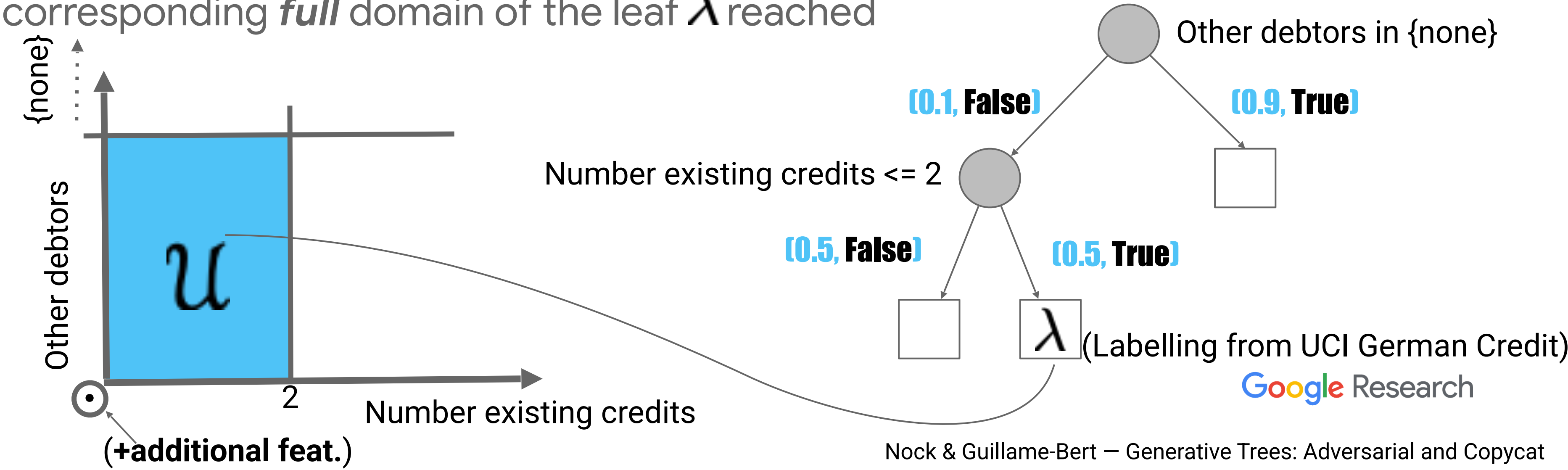
(Labelling from UCI German Credit)

Google Research

Key routines

For a decision tree **h**: for a given observation $x \in \mathcal{X}$, return the leaf $\lambda(x)$ whose path in the tree is satisfied by x

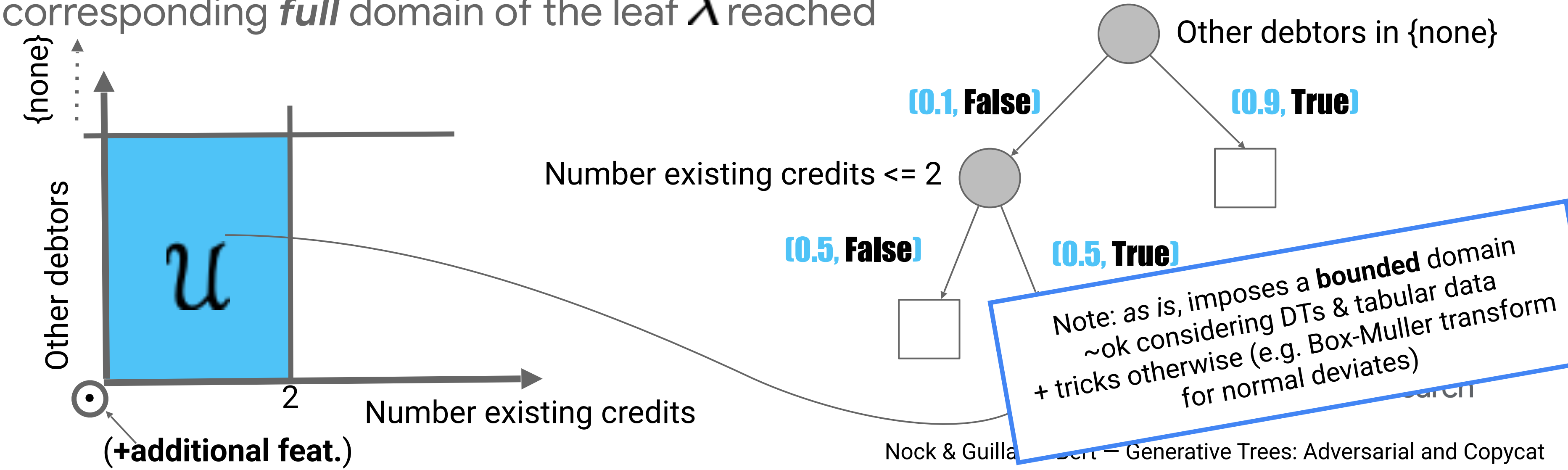
For a generative tree **G**: sample a path (wrt "Bernoullis") and sample uniformly in the corresponding ***full*** domain of the leaf λ reached



Key routines

For a decision tree **h**: for a given observation $x \in \mathcal{X}$, return the leaf $\lambda(x)$ whose path in the tree is satisfied by x

For a generative tree **G**: sample a path (wrt "Bernoullis") and sample uniformly in the corresponding ***full*** domain of the leaf λ reached



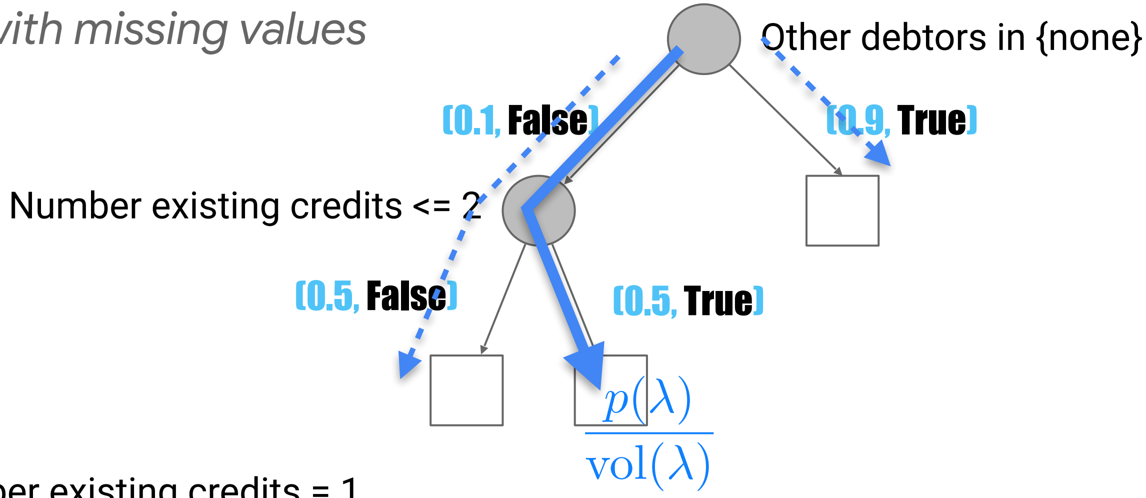
Additional conveniences of generative trees

↳ For any observation, *local density* computable in O(depth(G))

↳ If missing values, *likelihood* | observed values & generator G available in O(size(G))

↳ *XAI / fairness*: “as easy” to *interpret* as a decision tree

↳ Easily trainable from data *with missing values*



Other debtors = guarantor, Number existing credits = 1, …

Other debtors = **?**, Number existing credits = 3, …

Algorithms

Adversarial

↳ GAN-style (for “vs” training)
↳ simple (leaf→feature→split→*p* in $\mathrm{B}(p)$→repeat)

Boosting compliance in generative framework:
↳ a weak *generative* assumption = non total independence between data generation (**G**) and classification (**h**)
↳ most “expensive" computational bit = the computation of Bernoulli *p*’s
↳ geometric convergence of the chi square

$$\chi^2\left(\mathrm{N}^{\mathrm{new}}_{\tilde{\eta}} || \mathrm{P}_{\tilde{\eta}}\right) \leq \frac{1}{1+Q} \cdot \chi^2\left(\mathrm{N}^{\mathrm{old}}_{\tilde{\eta}} || \mathrm{P}_{\tilde{\eta}}\right)$$

↑
(details in paper)

Google Research

Adversarial

↳ GAN-style (for "vs" training)
↳ simple (leaf→feature→split→*p* in $\mathrm{B}(p)$→repeat)

Boosting compliance in generative framework:
↳ a weak *generative* assumption = non total independence between data generation (**G**) and classification (**h**)
↳ most "expensive" computational bit = the computation of Bernoulli *p*'s
↳ geometric convergence of the chi square

$$\chi^2\left(\mathrm{N}^{\mathrm{new}}_{\tilde{\eta}}||\mathrm{P}_{\tilde{\eta}}\right) \leq \frac{1}{1+Q}\cdot\chi^2\left(\mathrm{N}^{\mathrm{old}}_{\tilde{\eta}}||\mathrm{P}_{\tilde{\eta}}\right)$$

Copycat

↳ Powerful boosting DT induction algorithms for discriminator **h**. Can we rely on them to train **G** ?

Google Research

Adversarial

↳ GAN-style (for “vs” training)
↳ simple (leaf→feature→split→*p* in $\mathrm{B}(p)$→repeat)

Boosting compliance in generative framework:
↳ a weak *generative* assumption = non total independence between data generation (**G**) and classification (**h**)
↳ most “expensive" computational bit = the computation of Bernoulli *p*’s
↳ geometric convergence of the chi square

$$\chi^2\left(\mathrm{N}^{\mathrm{new}}_{\tilde{\eta}}||\mathrm{P}_{\tilde{\eta}}\right) \leq \frac{1}{1+Q}\cdot\chi^2\left(\mathrm{N}^{\mathrm{old}}_{\tilde{\eta}}||\mathrm{P}_{\tilde{\eta}}\right)$$

Copycat

↳ Powerful boosting DT induction algorithms for discriminator **h**. Can we rely on them to train **G** ?

Train **G** at “0” additional cost & with guarantees
↳ GT **G** and DT **h** share a tree (graph)
↳ **G** copies **h**’s tree at induction time & completes it (*p*) for *hardest* current generator

↳ **G** = balanced distribution of the weak learning assumption in Kearns & Mansour, STOC’96.
↳ trivial computations for **G** + geometric convergence in density ratio loss *for free* from boosting

(details in paper)

Experiments

Summary

Experiments carried out with Copycat training (fast, simple, little hyperparameter tuning required, ...), using Kearns and Mansour's optimal top-down algorithm;

1 classical toy generative problem + 4 more experiments against SOTA

↳ **Toy**: 2D heatmaps of densities (vs CTGAN)
↳ **Missing data imputation**: predict missing values in a dataset (vs MICE)
↳ **Gen-discrim**: discriminate between fake and real examples (vs CTGAN)
↳ **Train-gen** (*supervised data*): train model over fake data, test over real (vs CTGAN)
↳ **Gen-aug** (*supervised data*): augment real with generated + Train-gen (vs CTGAN)

(details in paper)

Google Research

CTGAN: Xu, Skoularidou, Cuesta-Infante & Veeramachaneni, NeurIPS'19
MICE: van Buuren, "Flexible imputation of missing data", Chapman & Hall / CRC, 2018

XAI in a 

```
[#1:root]
|-[0.0489, [ lng (CONTINUOUS) in [-76.8665, -72.7167]; |-1|-1| ] ]--[#2]
| |-[0.0366, [ search_vehicle (NOMINAL) in {TRUE}; |-1|-1| ] ]--[#100]
| | |-[0.1212, [ lat (CONTINUOUS) in [40.7067, 41.7329]; |-1|-1| ] ]--[#3010 (sampling)]
| | \-[0.8788, [ lat (CONTINUOUS) in [41.7329, 42.3426]; |-1|-1| ] ]--[#3011]
| |   |-[0.8276, [ lat (CONTINUOUS) in [41.7329, 41.8060]; |-1|-1| ] ]--[#3338]
| |   | |-[0.8333, [ warning_issued (NOMINAL) in {FALSE}; |-1|-1| ] ]--[#3756]
| |   | | |-[0.9500, [ raw_subject_race_code (NOMINAL) in {W, B}; |-1|-1| ] ]--[#4184]
| |   | | | |-[0.2632, [ reason_for_stop (NOMINAL) in {TrafficControlSignal, Other}; |-1|-1| ] ]--[#4380 (sampling)]
| |   | | | \-[0.7368, [ reason_for_stop (NOMINAL) in {StopSign, DefectiveLights, CellPhone, SuspendedLicense, Registration,
| |   | | |   |-[0.5000, [ district (NOMINAL) in {BARRYSQUARE, NORTHMEADOWS}; |-1|-1| ] ]--[#5788]
| |   | | |   | |-[0.4286, [ subject_age (INTEGER) in {14, 15, ..., 29}; |-1|-1| ] ]--[#9118 (sampling)]
| |   | | |   | \-[0.5714, [ subject_age (INTEGER) in {30, 31, ..., 94}; |-1|-1| ] ]--[#9119 (sampling)]
| |   | | |   \-[0.5000, [ district (NOMINAL) in {SOUTHWEST, ASYLUMHILL, PARKVILLE, FROGHOLLOW, BEHINDTHEROCKS, SOUTHGREEN,
| |   | | \-[0.0500, [ raw_subject_race_code (NOMINAL) in {A, I}; |-1|-1| ] ]--[#4185 (sampling)]
| |   | \-[0.1667, [ warning_issued (NOMINAL) in {TRUE}; |-1|-1| ] ]--[#3757 (sampling)]
| |   \-[0.1724, [ lat (CONTINUOUS) in [41.8060, 42.3426]; |-1|-1| ] ]--[#3339 (sampling)]
| \-[0.9634, [ search_vehicle (NOMINAL) in {FALSE}; |-1|-1| ] ]--[#101]
|   |-[0.0265, [ raw_search_authorization_code (NOMINAL) in {C, I}; |-1|-1
|   | |-[0.0870, [ lat (CONTINUOUS) in [40.7067, 41.6730]; |-1|-1| ] ]--[#
|   | \-[0.9130, [ lat (CONTINUOUS) in [41.6730, 42.3426]; |-1|-1| ] ]--[#4929]
```

↑↑ disparities in density young vs not-young
on "car/driver based search" in specific area

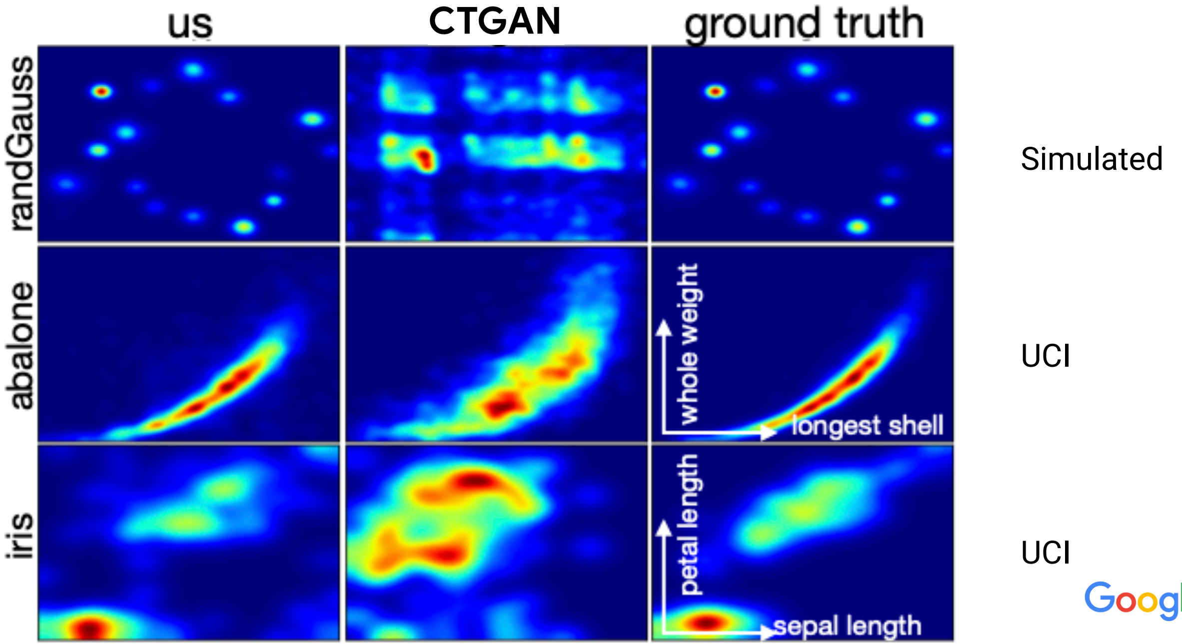
Example of generative tree learned on Stanford Open Policing / Hartford (more examples in paper)

Google Research

Toy 2D heat maps

↳**Setup**: generate data, compare with ground truth (10 000 nodes GT, 1K epoch CTGANs)

↳Some results:

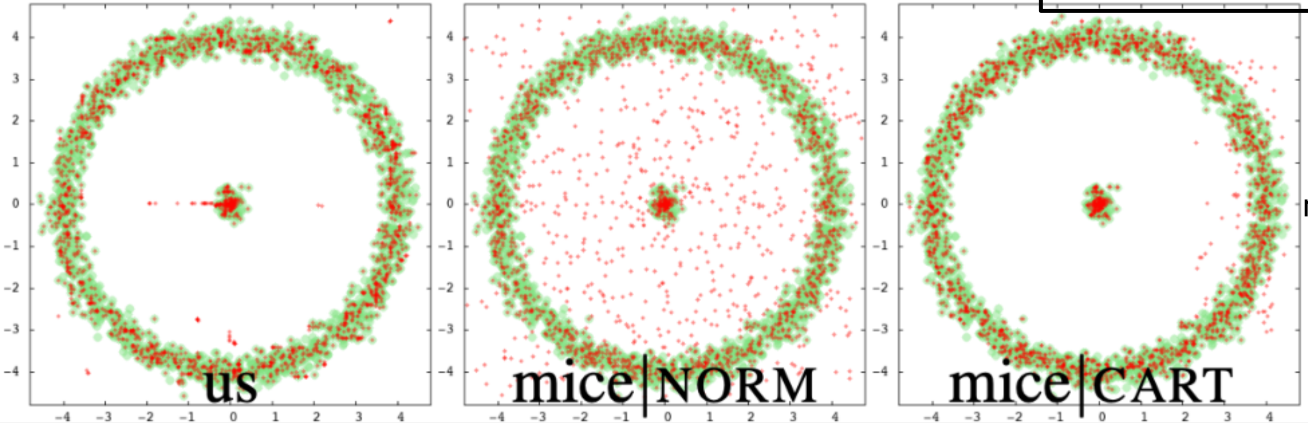


Google Research

Missing Data Imputation

↳**Summary**: synthetic data, remove *q*% features (Missing Completely At Random), impute w/ GT vs SOTA = mice (CART: use decision trees to predict missing in one column given the others, cycle through all columns several times)

In green: domain; red: imputed data



We use 1 tree

mice|CART uses 10 trees

(more examples & more details in paper)

Google Research

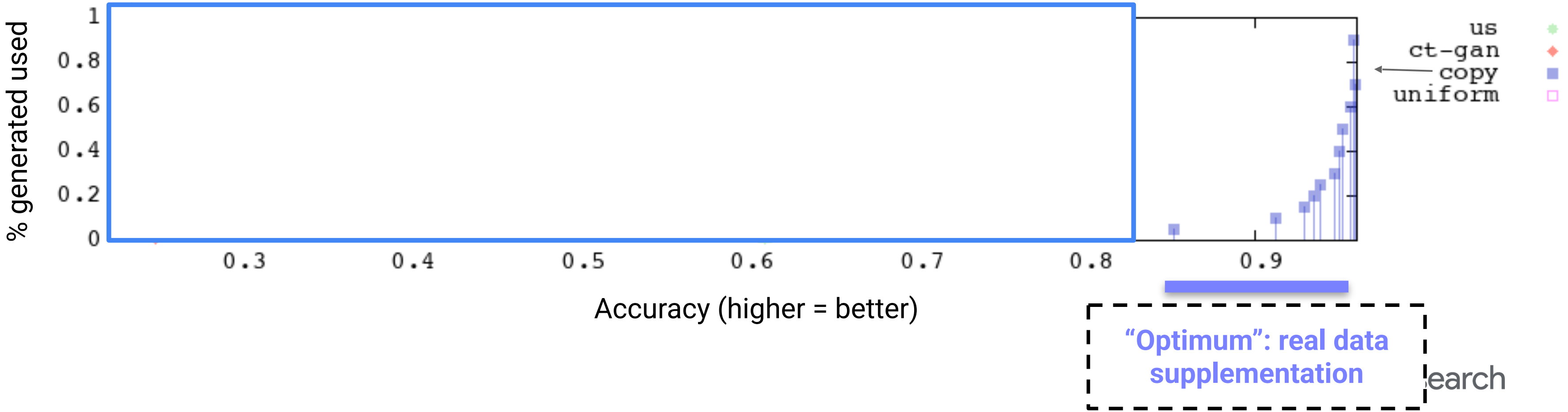
Gen-aug

↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data

Google Research

Gen-aug

↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data — example of UCI DNA, 181 binary features

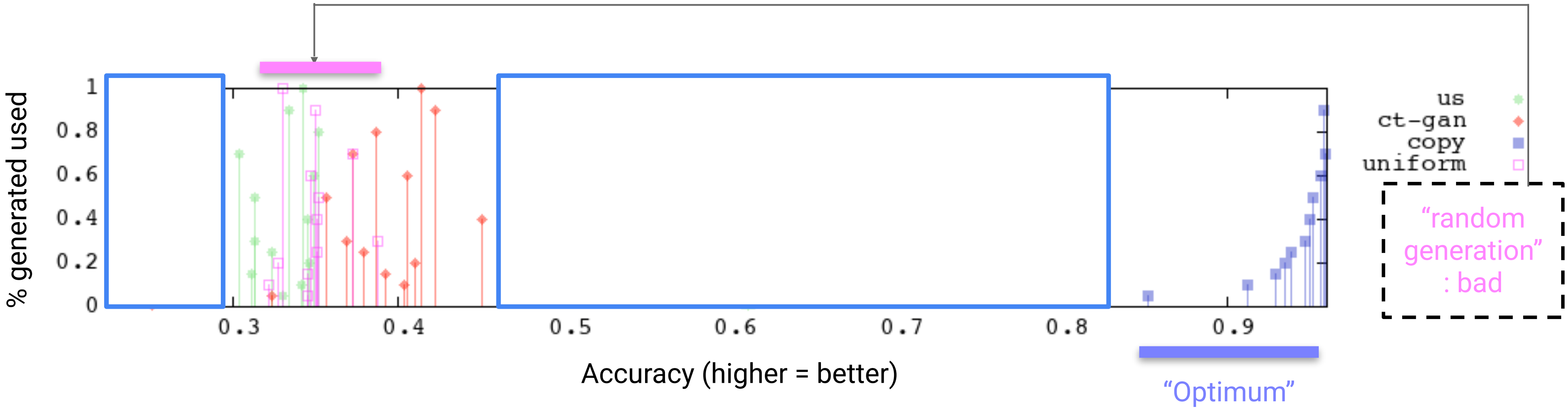


earch

Nock & Guillame-Bert — Generative Trees: Adversarial and Copycat

Gen-aug

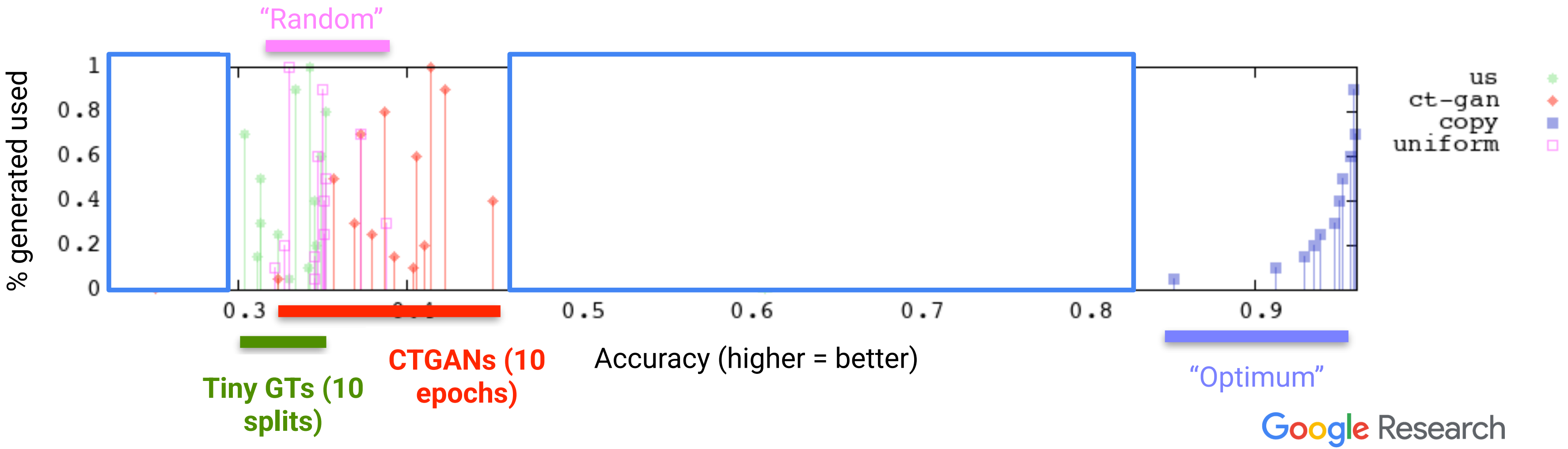
↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data — example of UCI DNA, 181 binary features



Google Research

Gen-aug

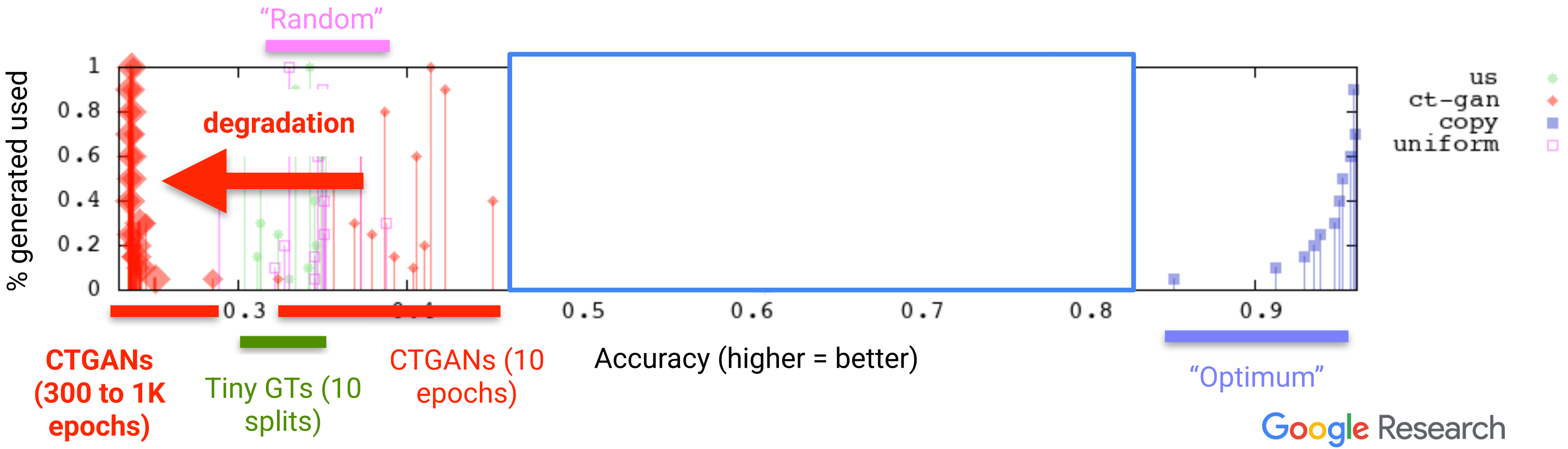
↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data — example of UCI DNA, 181 binary features



Google Research

Gen-aug

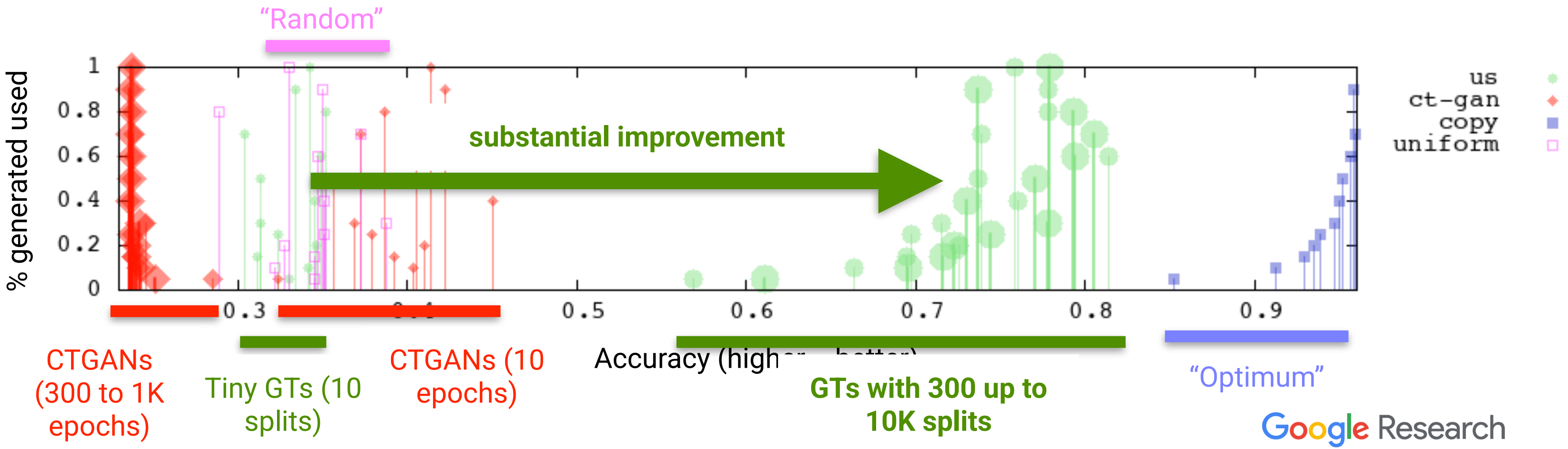
↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data — example of UCI DNA, 181 binary features



Google Research

Gen-aug

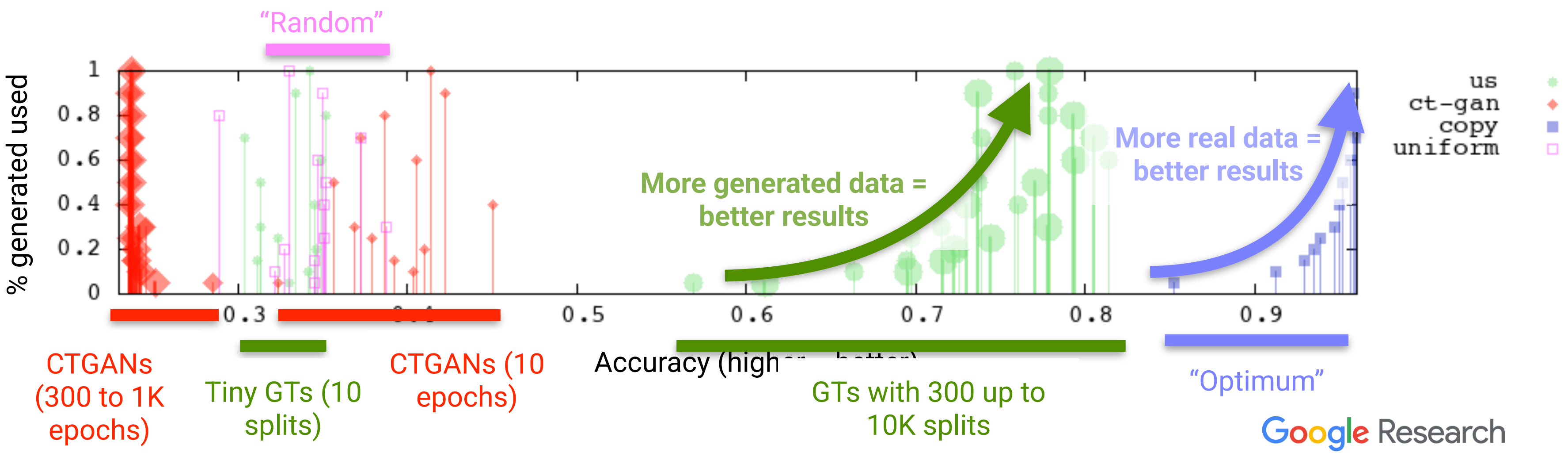
↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data — example of UCI DNA, 181 binary features



Google Research

Gen-aug

↳**Summary**: use part of real data to train generator, supplement remaining training data with varying % of generated data, train supervised classifier for the task, evaluate accuracy on test data — example of UCI DNA, 181 binary features



Google Research

↳See paper for more results

Conclusion / future work

Our contributions

↳ new *tight* formulation of the GAN losses *from the supervised side* (properness) if discriminator calibrated, gives the *chi square* as a “default” generator training loss

↳ new *generative models* & adversarial training w/ *boosting compliant convergence*

↳ new *cheap* training for generative models (*copycat*) + “boosting for free” convergence

Future work includes

↳ XAI / fairness: constrained induction of generative models

↳ privacy

↳ lots of formal questions (generalisation, pruning generators, ensembles of GTs, etc.)

Thank you !

{richardnock,gbm}@google.com

See also: **Yggdrasil** Decision Forests https://github.com/google/yggdrasil-decision-forests

Google Research