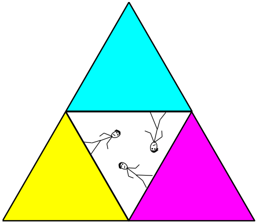
Information Elicitation and Design of Surrogate Losses



Bo Waggoner
University of Colorado, Boulder

Peking University
October 30, 2020

Based on joint work with Jessie Finocchiaro and Rafael Frongillo (U. Colorado, Boulder).

Motivation 1: Design of surrogate loss functions

- We measure how well an algorithm predicts using a loss function.

Motivation 1: Design of surrogate loss functions

- We measure how well an algorithm predicts using a loss function.
- Often, natural or popular losses are **intractable.**

Motivation 1: Design of surrogate loss functions

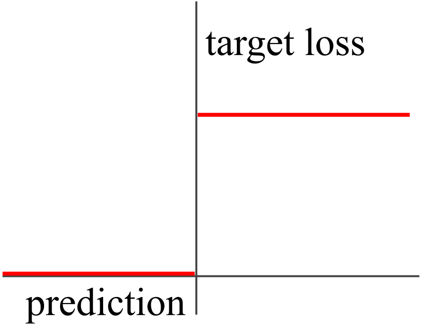
- We measure how well an algorithm predicts using a loss function.
- Often, natural or popular losses are **intractable.**
- So, we minimize another one – the **surrogate**.

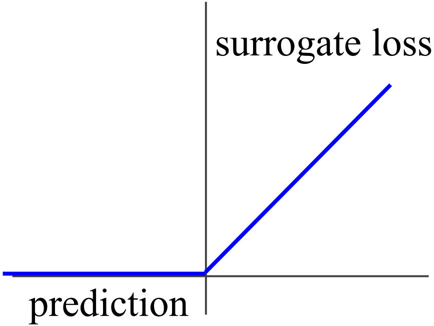
Motivation 1: Design of surrogate loss functions

- We measure how well an algorithm predicts using a loss function.
- Often, natural or popular losses are **intractable.**
- So, we minimize another one – the **surrogate**.

Motivation 1: Design of surrogate loss functions

- We measure how well an algorithm predicts using a loss function.
- Often, natural or popular losses are **intractable.**
- So, we minimize another one – the **surrogate**.





Key tool and motivation 2: Information elicitation

- A forecaster is asked to make a prediction about a future event.

Key tool and motivation 2: Information elicitation

- A forecaster is asked to make a prediction about a future event.
- She is assigned a **loss** based on the outcome.

Key tool and motivation 2: Information elicitation

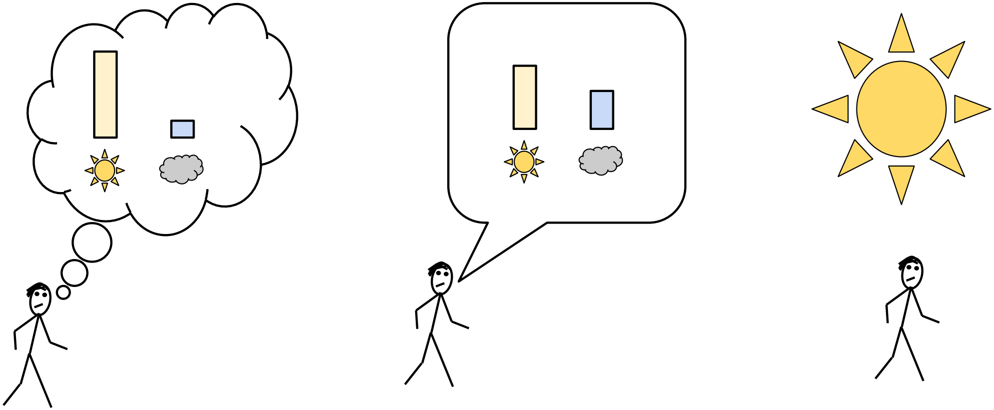
- A forecaster is asked to make a prediction about a future event.
- She is assigned a **loss** based on the outcome.
- She minimizes expected loss.

Key tool and motivation 2: Information elicitation

- A forecaster is asked to make a prediction about a future event.
- She is assigned a **loss** based on the outcome.
- She minimizes expected loss.

Key tool and motivation 2: Information elicitation

- A forecaster is asked to make a prediction about a future event.
- She is assigned a **loss** based on the outcome.
- She minimizes expected loss.



Connection between problems

$$\underset{r \in \mathcal{R}}{\operatorname{argmin}} \underset{Y \sim p}{\mathbb{E}} \ell(r, Y)$$

Outline

1. Concepts and definitions from information elicitation
   what do you get when you minimize a loss?
2. Surrogate loss functions for machine learning
3. The embedding approach; our contributions

Part 1: Concepts and definitions from information elicitation

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \underset{r \in \mathcal{R}}{\operatorname{argmin}} \underset{Y \sim p}{\mathbb{E}} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$ — $\Gamma(p) = \mathbb{E}_{Y \sim p} Y$ *(mean)*

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \underset{r \in \mathcal{R}}{\operatorname{argmin}} \underset{Y \sim p}{\mathbb{E}} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$ $\Gamma(p) = \mathbb{E}_{Y \sim p} Y$ *(mean)*
- $\ell(r, y) = |r - y|$

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$ $\Gamma(p) = \mathbb{E}_{Y \sim p} Y$ *(mean)*
- $\ell(r, y) = |r - y|$ *median*

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$ — $\Gamma(p) = \mathbb{E}_{Y \sim p} Y$ *(mean)*
- $\ell(r, y) = |r - y|$ — *median*
- $\ell(r, y) = \begin{cases} 0 & r = y \\ 1 & \text{otherwise} \end{cases}$

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$ — $\Gamma(p) = \mathbb{E}_{Y \sim p} Y$ *(mean)*
- $\ell(r, y) = |r - y|$ — *median*
- $\ell(r, y) = \begin{cases} 0 & r = y \\ 1 & \text{otherwise} \end{cases}$ — *mode*

Information elicitation

What do you get when you minimize a loss?

$$\Gamma(p) := \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \tag{1}$$

Examples:

- $\ell(r, y) = (r - y)^2$ — $\Gamma(p) = \mathbb{E}_{Y \sim p} Y$ *(mean)*
- $\ell(r, y) = |r - y|$ — *median*
- $\ell(r, y) = \begin{cases} 0 & r = y \\ 1 & \text{otherwise} \end{cases}$ — *mode*
- $\ell(r, y) = ??$ — *variance*

Information elicitation

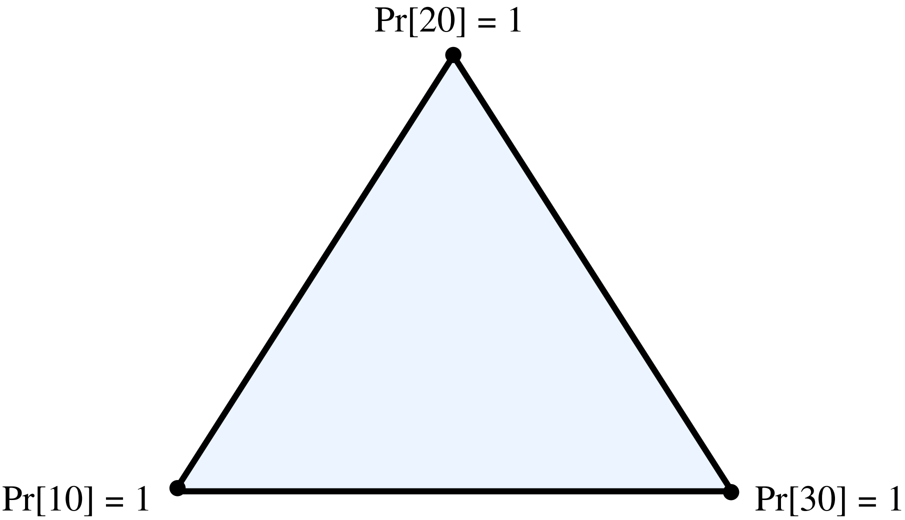
What do you get when you minimize a loss?

$$\Gamma(p) := \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \tag{1}$$

- $\Gamma : \Delta_{\mathcal{Y}} \to 2^{\mathcal{R}}$ is a **property** of the distribution p.
- Γ is **elicitable** if there exists ℓ such that (1) holds.

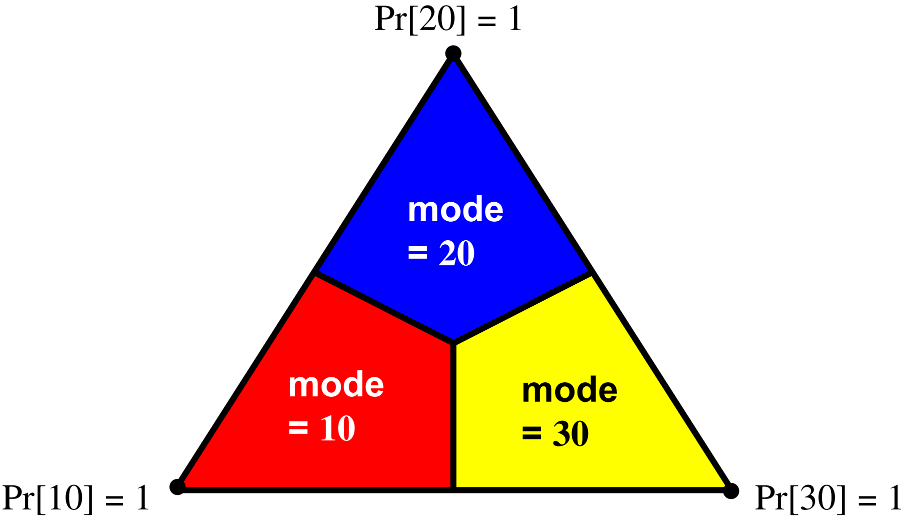
Information elicitation - the picture

The simplex $\Delta_{\mathcal{Y}}$ for $\mathcal{Y} = \{10, 20, 30\}$:



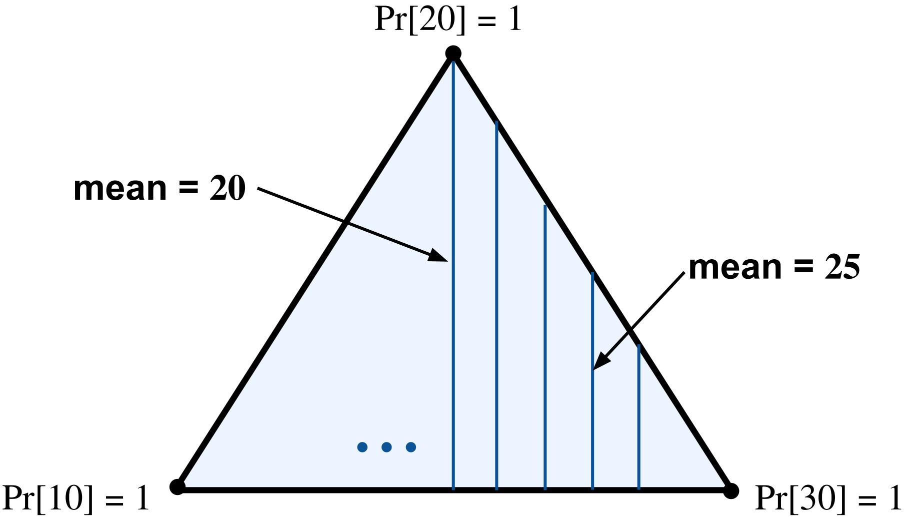
Information elicitation - the picture

- A property is a partition of the simplex.
- The **level set** of r is $\{p : \Gamma(p) = r\}$.



Information elicitation - the picture

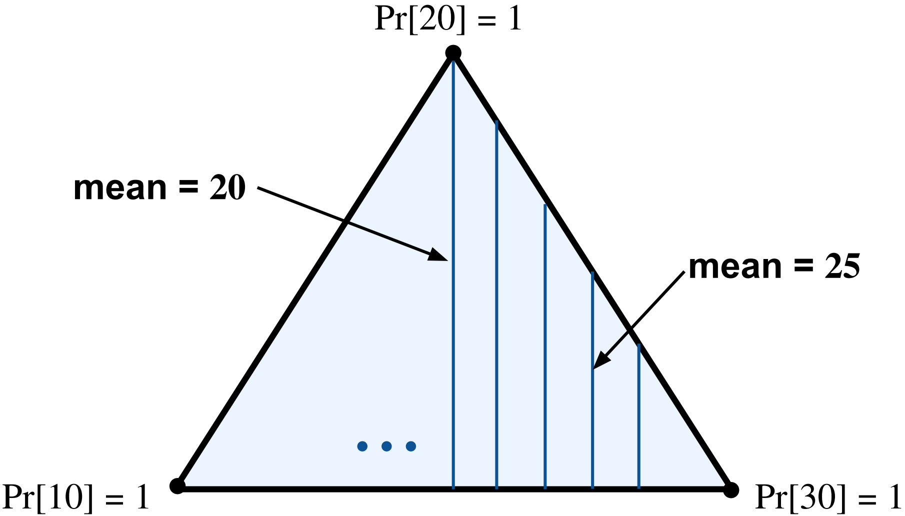
- A property is a partition of the simplex.
- The **level set** of r is $\{p : \Gamma(p) = r\}$.



Key basic fact

Theorem

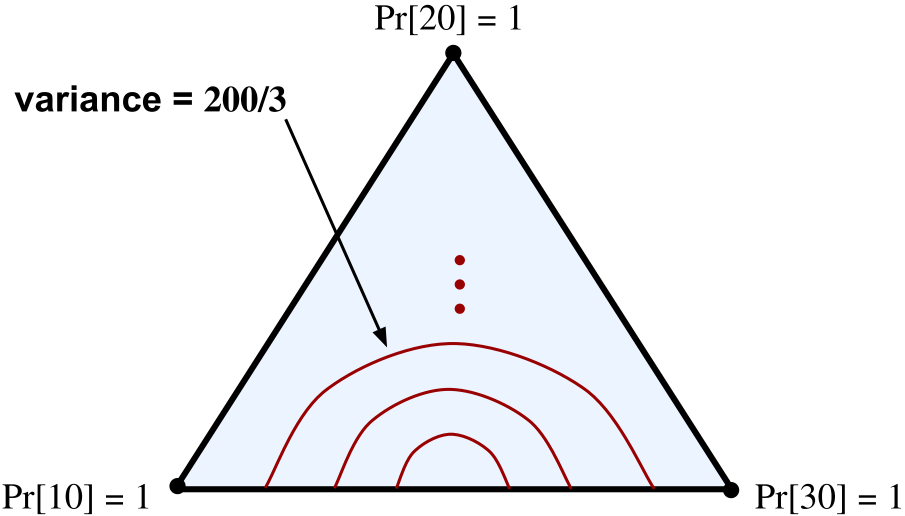
If a property is elicitable, then all of its level sets are convex sets.



Key basic fact

Theorem

If a property is elicitable, then all of its level sets are convex sets.



Dealing with non-elicitable properties

Indirect elicitation: Elicit some *other* properties, then compute $\Gamma(p)$.

[1] Frongillo and Kash, 2015.

Dealing with non-elicitable properties

Indirect elicitation: Elicit some *other* properties, then compute $\Gamma(p)$.

Example: elicit variance using mean, second moment.

[1] Frongillo and Kash, 2015.

Dealing with non-elicitable properties

Indirect elicitation: Elicit some *other* properties, then compute $\Gamma(p)$.

Example: elicit variance using mean, second moment.

$$L((\mu_1, \mu_2), y) = (\mu_1 - y)^2 + (\mu_2 - y^2)^2.$$

[1] Frongillo and Kash, 2015.

Dealing with non-elicitable properties

Indirect elicitation: Elicit some *other* properties, then compute $\Gamma(p)$.

Example: elicit variance using mean, second moment.

$$L((\mu_1, \mu_2), y) = (\mu_1 - y)^2 + (\mu_2 - y^2)^2.$$

Elicitation complexity[1] of Γ: fewest parameters needed to indirectly elicit Γ.
Variance: 2.

[1] Frongillo and Kash, 2015.

Part 2: Surrogate loss functions for machine learning

Supervised learning setting

- Space of features $\mathcal{X}$

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels
- Data point: $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels
- Data point: $(x, y) \in \mathcal{X} \times \mathcal{Y}$.
- Hypothesis $h : \mathcal{X} \to \mathcal{R}$.

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels
- Data point: $(x, y) \in \mathcal{X} \times \mathcal{Y}$.
- Hypothesis $h : \mathcal{X} \to \mathcal{R}$.
- Target loss $\ell : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$.

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels
- Data point: $(x, y) \in \mathcal{X} \times \mathcal{Y}$.
- Hypothesis $h : \mathcal{X} \to \mathcal{R}$.
- Target loss $\ell : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$.

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels
- Data point: $(x, y) \in \mathcal{X} \times \mathcal{Y}$.
- Hypothesis $h : \mathcal{X} \to \mathcal{R}$.
- Target loss $\ell : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$.

Goal: for a distribution $\mathcal{D}$ on $\mathcal{X} \times \mathcal{Y}$,
find a hypothesis minimizing expected loss:

$$\min_h \mathop{\mathbb{E}}_{X,Y \sim \mathcal{D}} \ell(h(x), y).$$

Supervised learning setting

- Space of features $\mathcal{X}$
- Space of outcomes $\mathcal{Y}$
  finite, e.g. labels
- Data point: $(x, y) \in \mathcal{X} \times \mathcal{Y}$.
- Hypothesis $h : \mathcal{X} \to \mathcal{R}$.
- Target loss $\ell : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$.

Goal: for a distribution $\mathcal{D}$ on $\mathcal{X} \times \mathcal{Y}$,
find a hypothesis minimizing expected loss:

$$\min_h \mathop{\mathbb{E}}_{X,Y \sim \mathcal{D}} \ell(h(x), y).$$

Let p_x = conditional distribution of Y given $X = x$.
Bayes optimal: $h(x) = \gamma(p_x)$
where γ is the property elicited by ℓ.

Why surrogate losses?

Problem 1: wish to learn $h : x \mapsto \gamma(p_x)$, but γ is not elicitable.

Why surrogate losses?

Problem 1: wish to learn $h : x \mapsto \gamma(p_x)$, but γ is not elicitable.

Problem 2: ℓ is **intractable.**

Why surrogate losses?

Problem 1: wish to learn $h : x \mapsto \gamma(p_x)$, but γ is not elicitable.

Problem 2: ℓ is **intractable.**

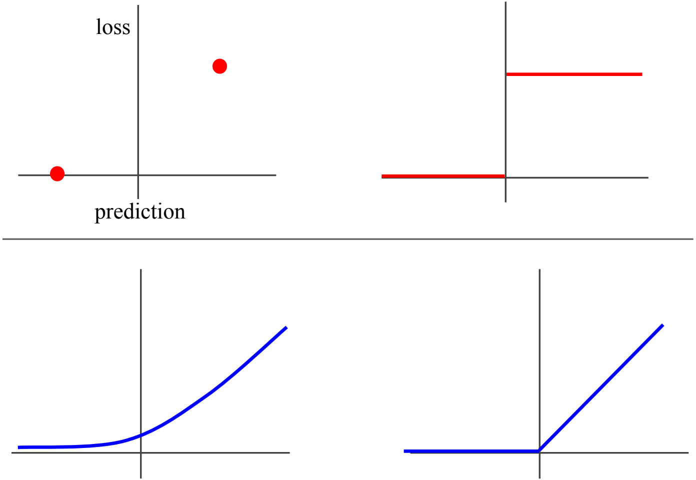
Solutions: use a surrogate loss L.

Why surrogate losses?

Problem 1: wish to learn $h : x \mapsto \gamma(p_x)$, but γ is not elicitable.

Problem 2: ℓ is **intractable.**

Solutions: use a surrogate loss L.



What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex
- d is small[2]

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex
- d is small[2]

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex
- d is small[2]

(2) Relationship to ℓ:

- Note: need combination of L and **link** $\psi : \mathbb{R}^d \to \mathcal{R}$.

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex
- d is small[2]

(2) Relationship to ℓ:

- Note: need combination of L and **link** $\psi : \mathbb{R}^d \to \mathcal{R}$.
- **Consistency:** optimizing L, then applying ψ, leads to optimizing ℓ.

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex
- d is small[2]

(2) Relationship to ℓ:

- Note: need combination of L and **link** $\psi : \mathbb{R}^d \to \mathcal{R}$.
- **Consistency:** optimizing L, then applying ψ, leads to optimizing ℓ.
- **Calibration:** if $\psi(u)$ is suboptimal for ℓ, then u is suboptimal for L (for each p_x).

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

What makes a surrogate loss good?

(1) tractability, (2) correct relationship to ℓ.

(1) Tractability of a surrogate loss $L : \mathbb{R}^d \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$

- Convex
- d is small[2]

(2) Relationship to ℓ:

- Note: need combination of L and **link** $\psi : \mathbb{R}^d \to \mathcal{R}$.
- **Consistency:** optimizing L, then applying ψ, leads to optimizing ℓ.
- **Calibration:** if $\psi(u)$ is suboptimal for ℓ, then u is suboptimal for L (for each p_x).
- **Our point (new work):** it is necessary and almost sufficient for L, ψ to **indirectly elicit** γ.
  Lower bounds, etc.

[2] *convex calibration dimension*, Ramaswamy & Agarwal 2016.

Part 3: The embedding approach; our contributions.

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, ...).

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, ...).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, $\ldots$).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points
- Find a convex surrogate L such that $u_r \in \operatorname{argmin} L(\cdot; p)$ if and only if $r \in \operatorname{argmin} \ell(\cdot; p)$.

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, $\ldots$).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points
- Find a convex surrogate L such that $u_r \in \operatorname{argmin} L(\cdot; p)$ if and only if $r \in \operatorname{argmin} \ell(\cdot; p)$.
- Set $\psi(u_r) = r$ for each r.

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, ...).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points
- Find a convex surrogate L such that $u_r \in \operatorname{argmin} L(\cdot; p)$ if and only if $r \in \operatorname{argmin} \ell(\cdot; p)$.
- Set $\psi(u_r) = r$ for each r.
- Choose $\psi(u)$ somehow on the rest of $\mathbb{R}^d$. (L, ψ) *is an embedding of* ℓ

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, ...).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points
- Find a convex surrogate L such that $u_r \in \operatorname{argmin} L(\cdot; p)$ if and only if $r \in \operatorname{argmin} \ell(\cdot; p)$.
- Set $\psi(u_r) = r$ for each r.
- Choose $\psi(u)$ somehow on the rest of $\mathbb{R}^d$. (L, ψ) *is an embedding of* ℓ

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, ...).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points
- Find a convex surrogate L such that $u_r \in \operatorname{argmin} L(\cdot; p)$ if and only if $r \in \operatorname{argmin} \ell(\cdot; p)$.
- Set $\psi(u_r) = r$ for each r.
- Choose $\psi(u)$ somehow on the rest of $\mathbb{R}^d$. (L, ψ) *is an embedding of* ℓ

Problems: when does such an L exist? how to find it?

Theorem

We can automatically, efficiently construct an embedding L *of* ℓ *from* ℓ *using dimensions* $d = |\mathcal{Y}| - 1$.

How do you find a good surrogate?

Suppose $\mathcal{R}$ is finite (e.g. classification, ranking, top-k, ...).

One idea: **embed** ℓ as follows:

- For each $r \in \mathcal{R}$, choose some $u_r \in \mathbb{R}^d$.
  embedding points
- Find a convex surrogate L such that $u_r \in \operatorname{argmin} L(\cdot; p)$ if and only if $r \in \operatorname{argmin} \ell(\cdot; p)$.
- Set $\psi(u_r) = r$ for each r.
- Choose $\psi(u)$ somehow on the rest of $\mathbb{R}^d$. (L, ψ) *is an embedding of* ℓ

Problems: when does such an L exist? how to find it?

Theorem

We can automatically, efficiently construct an embedding L *of* ℓ *from* ℓ *using dimensions* $d = |\mathcal{Y}| - 1$.

Proof idea: Use the convex conjugate of the Bayes risk of ℓ.
Not really a new construction; e.g. prediction markets!

Example 1: Mode

On $\mathcal{Y} = \{\pm 1\}$: hinge loss.

Example 1: Mode

On $\mathcal{Y} = \{\pm 1\}$: hinge loss.

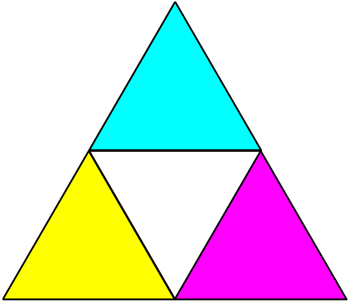
More generally: works!

Example 2: Classification with abstain

$$\ell(r, y) = \begin{cases} \frac{1}{2} & r = \text{ abstain} \\ 0 & r = y \\ 1 & \text{otherwise.} \end{cases}$$

Example 2: Classification with abstain

$$\ell(r,y) = \begin{cases} \frac{1}{2} & r = \text{ abstain} \\ 0 & r = y \\ 1 & \text{otherwise.} \end{cases}$$



Example 2: Classification with abstain

$$\ell(r, y) = \begin{cases} \frac{1}{2} & r = \text{ abstain} \\ 0 & r = y \\ 1 & \text{otherwise.} \end{cases}$$

Amazing embedding construction:[3] $d = \lceil \log_2 |\mathcal{Y}| \rceil$.

[3] Ramaswamy et al. 2018

Back to the theorem

Theorem

(Restatement) If ℓ is a discrete loss, then it has an embedding L.

Furthermore, L is polyhedral.

polyhedral: piecewise-linear and convex.

Back to the theorem

Theorem

(Restatement) If ℓ is a discrete loss, then it has an embedding L.

Furthermore, L is polyhedral.

polyhedral: piecewise-linear and convex.

Theorem

If L is polyhedral, it embeds a discrete loss.

Back to the theorem

Theorem

(Restatement) If ℓ is a discrete loss, then it has an embedding L.

Furthermore, L is polyhedral.

polyhedral: piecewise-linear and convex.

Theorem

If L is polyhedral, it embeds a discrete loss.

Not just embedding!

Not yet answered: how to construct ψ outside of embedding points?

Not just embedding!

Not yet answered: how to construct ψ outside of embedding points?

Answer: the ϵ-thickened link $\psi_{L,\epsilon}$.

Not just embedding!

Not yet answered: how to construct ψ outside of embedding points?

Answer: the ϵ-thickened link $\psi_{L,\epsilon}$.

Theorem

Let L be polyhedral, embedding ℓ. Then there exists $\epsilon > 0$ such that $L, \psi_{L,\epsilon}$ is calibrated for ℓ.

Not just embedding!

Not yet answered: how to construct ψ outside of embedding points?

Answer: the ϵ-thickened link $\psi_{L,\epsilon}$.

Theorem

Let L be polyhedral, embedding ℓ. Then there exists $\epsilon > 0$ such that $L, \psi_{L,\epsilon}$ is calibrated for ℓ.

Theorem (Current work)

In fact, there exists $\epsilon > 0$ and $C > 0$ such that, for all u and p,

$$\ell(\psi_{L,\epsilon}(u); p) - \ell^*(p) \leq C \cdot (L(u;p) - L^*(p)) .$$

Not just embedding!

Not yet answered: how to construct ψ outside of embedding points?

Answer: the ϵ-thickened link $\psi_{L,\epsilon}$.

Theorem

Let L be polyhedral, embedding ℓ. Then there exists $\epsilon > 0$ such that $L, \psi_{L,\epsilon}$ is calibrated for ℓ.

Theorem (Current work)

In fact, there exists $\epsilon > 0$ and $C > 0$ such that, for all u and p,

$$\ell(\psi_{L,\epsilon}(u);p) - \ell^*(p) \leq C \cdot (L(u;p) - L^*(p)) .$$

Implication: Fast rates of convergence of L translate *linearly* to fast rates for ℓ.
Not generally true for smooth surrogates.

Summary and supplementary results

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Other results on tractability:

- Used elicitation to show inconsistency of proposed losses: Lovasz hinge, top-k

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Other results on tractability:

- Used elicitation to show inconsistency of proposed losses: Lovasz hinge, top-k
- Characterization of embeddability for $d = 1$ (ongoing work)

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Other results on tractability:

- Used elicitation to show inconsistency of proposed losses: Lovasz hinge, top-k
- Characterization of embeddability for $d = 1$ (ongoing work)
- Techniques for lower bounds on d-embeddability of given ℓ (ongoing work)

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Other results on tractability:

- Used elicitation to show inconsistency of proposed losses: Lovasz hinge, top-k
- Characterization of embeddability for $d = 1$ (ongoing work)
- Techniques for lower bounds on d-embeddability of given ℓ (ongoing work)

Summary and some other results

Summary:

- Introduced information elicitation and properties Γ
- Discussed "good" surrogates L for targets ℓ:
  - Tractable (convex, dimension d low)
  - Represents ℓ correctly (consistent, calibrated, etc)
- Results for discrete losses:
  - Formalized embedding approach; always possible
  - embeddings $\iff$ polyhedral surrogate losses
  - ϵ-thickened link $\implies$ calibration $\implies$ consistency

Other results on tractability:

- Used elicitation to show inconsistency of proposed losses: Lovasz hinge, top-k
- Characterization of embeddability for $d = 1$ (ongoing work)
- Techniques for lower bounds on d-embeddability of given ℓ (ongoing work)

Thanks!