The Geometry of Losses

Robert Williamson
Machine Learning is always done for a purpose
• Machine Learning is always done for a purpose
• Even if you think you are just seeking “information”, ultimately you will use the information
Machine Learning is always done for a **purpose**

Even if you think you are just seeking “information”, ultimately you will **use** the information

And that use will incur a **loss**
Machine Learning is always done for a purpose
Even if you think you are just seeking “information”, ultimately you will use the information
And that use will incur a loss
Thus loss functions are central to machine learning
Machine Learning is always done for a purpose
Even if you think you are just seeking “information”, ultimately you will use the information
And that use will incur a loss
Thus loss functions are central to machine learning
Restrict scope: Given labels $i \in [n]$ predict $p \in \Delta^n$. 
But Machine Learning only pays Lip-Service to Losses

\[ \ell(i, p) \]

Perhaps \( \ell \) is convex and/or Lipschitz in \( p \)

No structure.
No principles.
No guidance.
No understanding.
But Machine Learning only pays Lip-Service to Losses

\[ \ell(i, p) \]

- Perhaps \( \ell \) is convex and/or Lipschitz in \( p \)
But Machine Learning only pays Lip-Service to Losses

\[ \ell(i, p) \]

- Perhaps \( \ell \) is convex and/or Lipschitz in \( p \)
- No structure.
But Machine Learning only pays Lip-Service to Losses

\[ \ell(i, p) \]

- Perhaps \( \ell \) is convex and/or Lipschitz in \( p \)
- No structure. No principles.
But Machine Learning only pays Lip-Service to Losses

\[ \ell(i, p) \]

- Perhaps \( \ell \) is convex and/or Lipschitz in \( p \)
- No structure. No principles. No guidance.
But Machine Learning only pays Lip-Service to Losses

$$\ell(i, p)$$

- Perhaps $\ell$ is convex and/or Lipschitz in $p$
- No structure. No principles. No guidance. No understanding.
However there **is** a theory of loss functions. . .

- Loss $\ell(i, p) \in \mathbb{R}_+$ incurred when predict $p \in \Delta^n$ and $i \in [n]$ occurs.
However there is a theory of loss functions.

- Loss $\ell(i, p) \in \mathbb{R}_+$ incurred when predict $p \in \Delta^n$ and $i \in [n]$ occurs.
- View as $\ell : \Delta^n \rightarrow \mathbb{R}^n_+$,

$$
\ell(p) = \begin{pmatrix}
\ell(1, p) \\
\ell(2, p) \\
\vdots \\
\ell(n, p)
\end{pmatrix}
$$

Loss set: $\ell(\Delta^n) \subset \mathbb{R}^n_+$ Superprediction set to the North-East of $\ell(\Delta^n)$. . .
However there is a theory of loss functions...

- Loss $\ell(i, p) \in \mathbb{R}_+$ incurred when predict $p \in \Delta^n$ and $i \in [n]$ occurs.
- View as $\ell: \Delta^n \rightarrow \mathbb{R}_n^+$,

$$\ell(p) = \begin{pmatrix}
\ell(1, p) \\
\ell(2, p) \\
\vdots \\
\ell(n, p)
\end{pmatrix}$$

- $\ell$ is proper if

$$p \in \arg \min_{q \in \Delta^n} \mathbb{E}_{i \sim p} \ell(i, q) = \arg \min_{q \in \Delta^n} \langle p, \ell(q) \rangle$$
However, there is a theory of loss functions...

- Loss $\ell(i, p) \in \mathbb{R}_+$ incurred when predict $p \in \Delta^n$ and $i \in [n]$ occurs.
- View as $\ell: \Delta^n \to \mathbb{R}_n^n$,

$$
\ell(p) = \begin{pmatrix}
\ell(1, p) \\
\ell(2, p) \\
\vdots \\
\ell(n, p)
\end{pmatrix}
$$

- $\ell$ is proper if

$$
p \in \arg\min_{q \in \Delta^n} E_i p \ell(i, q) = \arg\min_{q \in \Delta^n} \langle p, \ell(q) \rangle
$$

- Loss set: $\ell(\Delta^n) \subset \mathbb{R}_n^n$
However there is a theory of loss functions. . .

- Loss $\ell(i, p) \in \mathbb{R}_+$ incurred when predict $p \in \Delta^n$ and $i \in [n]$ occurs.
- View as $\ell : \Delta^n \rightarrow \mathbb{R}_+^n$,

$$
\ell(p) = \begin{pmatrix}
\ell(1, p) \\
\ell(2, p) \\
\vdots \\
\ell(n, p)
\end{pmatrix}
$$

- $\ell$ is proper if

$$
p \in \arg \min_{q \in \Delta^n} \mathbb{E}_{i \sim p} \ell(i, q) = \arg \min_{q \in \Delta^n} \langle p, \ell(q) \rangle
$$

- Loss set: $\ell(\Delta^n) \subset \mathbb{R}_+^n$
- Superprediction set to the North-East of $\ell(\Delta^n)$ . . .
Superprediction Set $S_\ell := \{x \in \mathbb{R}^n : \exists y \in \text{dom } \ell, \ x \geq \ell(y)\}$

- Every proper loss $\ell$ induces a convex $S_\ell$.

---

$x = \ell(v)$

$\ell_1(v)$

$\ell_2(v)$

$S_\ell$

$\ell(V)$

$q$

Every proper loss $\ell$ induces a convex $S_\ell$. 
Superprediction Set \( S_\ell := \{ x \in \mathbb{R}^n : \exists y \in \text{dom } \ell, \ x \geq \ell(y) \} \)

Every proper loss \( \ell \) induces a convex \( S_\ell \).

Any loss with non-convex \( S_\ell \) is inadmissible.
Every proper loss $\ell$ induces a convex $S_\ell$.

Any loss with non-convex $S_\ell$ is inadmissible.

$S_\ell$ convex $\Rightarrow \ell$ can be reparametrised to be proper.
Superprediction Set $S_\ell := \{x \in \mathbb{R}^n : \exists y \in \text{dom } \ell, x \geq \ell(y)\}$

- Every proper loss $\ell$ induces a convex $S_\ell$.
- Any loss with non-convex $S_\ell$ is inadmissible.
- $S_\ell$ convex $\Rightarrow$ $\ell$ can be reparametrised to be proper.
- $\Rightarrow$ Nothing to lose by only considering convex $S_\ell$
Better understand the set of losses
Goal

Better understand the set of losses
Start with $S_\ell$ and recover $\ell$?
Better understand the set of losses

Start with $S_\ell$ and recover $\ell$?

Build an algebra and calculus of losses
Goal

- Better understand the set of losses
- Start with $S_\ell$ and recover $\ell$?
- Build an algebra and calculus of losses
- Develop an inverse loss
Background

Minkowski sum \( S + T := \{ s + t : s \in S, t \in T \} \)

The recession cone \( 0 + C := \{ y \in \mathbb{R}^n : C \text{ recedes in direction } y \} \).
- **Minkowski sum**
  \[
  S + T := \{s + t : s \in S, \ t \in T\}
  \]

- \(C \subset \mathbb{K}^n\) **recedes** in the direction \(y \in \mathbb{R}^n\) if
  \[
  x + \lambda y \in C, \ \forall \lambda \geq 0, \ \forall x \in C.
  \]

- **The recession cone**
  \[
  0^+ C := \{y \in \mathbb{R}^n : C \text{ recedes in direction } y\}.\]
Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is convex. Its subdifferential is

$$\bar{\partial} \phi(x) = \{ x^* : \phi(x) + \langle x^*, y-x \rangle \leq \phi(y) \ \forall y \in \mathbb{R}^n \}.$$ 

Suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ is concave. Its superdifferential is

$$\hat{\partial} \phi(x) = \{ x^* : \phi(x) + \langle x^*, y-x \rangle \geq \phi(y) \ \forall y \in \mathbb{R}^n \}.$$
A set $C$ is of **negative type** if $C \in \mathcal{K}^n$, $C$ is closed, $0 \in \text{int} \ C$ and $0^+ C = \mathbb{R}_-^n$.

A set $C$ is of **positive type** if $C \in \mathcal{K}^n$, $C$ is closed, $0 \notin \text{int} \ C$ and $0^+ C = \mathbb{R}_+^n$. 
Support Functions

- The support function of a set $C$ is

$$\tilde{\sigma}_C(x) = \sup_{y \in C} \langle x, y \rangle.$$
The support function of a set $C$ is

$$\tilde{\sigma}_C(x) = \sup_{y \in C} \langle x, y \rangle.$$ 

This support function corresponds naturally to proper gains; when working with proper losses we will use

$$\hat{\sigma}_C(x) = \inf_{y \in C} \langle x, y \rangle.$$
The support function of a set $C$ is

$$\tilde{\sigma}_C(x) = \sup_{y \in C} \langle x, y \rangle.$$ 

This support function corresponds naturally to proper gains; when working with proper losses we will use

$$\hat{\sigma}_C(x) = \inf_{y \in C} \langle x, y \rangle.$$ 

The support functions are closed, 1-homogeneous and $\tilde{\sigma}_C$ is convex and $\hat{\sigma}_C$ is concave.
The **support function** of a set $C$ is

$$\tilde{\sigma}_C(x) = \sup_{y \in C} \langle x, y \rangle.$$  

This support function corresponds naturally to proper **gains**; when working with proper **losses** we will use

$$\hat{\sigma}_C(x) = \inf_{y \in C} \langle x, y \rangle.$$  

The support functions are closed, 1-homogeneous and $\tilde{\sigma}_C$ is convex and $\hat{\sigma}_C$ is concave.

**1:1 relationship between support functions and convex bodies.**
The support plane of a set $C$ is

$$H_C^+(u) := \{x \in \mathbb{R}^n : \langle x, u \rangle = \sigma_C(u)\}$$

and the supporting halfspace for $C$ of negative type

$$\tilde{H}_C := \{x \in \mathbb{R}^n : \langle x, u \rangle \leq \tilde{\sigma}_C(u)\}$$

or for $C$ of positive type

$$\hat{H}_C := \{x \in \mathbb{R}^n : \langle x, u \rangle \geq \hat{\sigma}_C(u)\}.$$

The support set

$$F_C(x) := \tilde{H}_S(x) \cap S = \tilde{\partial} \tilde{\sigma}_C(x).$$

It is also true that

$$F_C(x) = \hat{H}_S(x) \cap S = \hat{\partial} \hat{\sigma}_C(x).$$
A set $A \subset \mathbb{R}^n$ is **star-shaped** if $0 \in A$ and $(0, 1] \cdot A \subset A$.

It is **shady** if $[1, \infty) \cdot A \subset A$.

The **convex-gauge** of closed star-shaped set $C \subset \mathbb{R}^n$ is defined by

$$\check{\gamma}_C(x) = \inf\{\mu \geq 0 : x \in \mu C\}.$$ 

The **concave-gauge** of a shady set $C$ is defined by

$$\hat{\gamma}_C(x) = \sup\{\mu \geq 0 : x \in \mu C\}.$$
The **convex-polar** of $C \in \mathcal{K}^n$ is

\[ C^\circ := \text{lev}_{\leq 1} \check{\sigma}_C \]

If $C$ is of **negative** type, then

\[ \check{\gamma}_C = \check{\sigma}_C \]
\[ \check{\gamma}_{C^\circ} = \check{\sigma}_C. \]
The concave-polar of $C \in \mathcal{K}^n$ is

$$C^\otimes := \text{lev}_{\geq 1} \hat{\sigma}_C$$

If $C$ is of positive type, then

$$\hat{\gamma}_C = \hat{\sigma}_{C^\otimes}$$
$$\hat{\gamma}_{C^\otimes} = \hat{\sigma}_C.$$
If $k$ is a convex-gauge (and thus non-negative, 1-homogeneous convex function with $k(0) = 0$) then the convex-polar of $k$ is defined as

$$k^\ominus(y) = \inf\{\mu \geq 0 : \langle x, y \rangle \leq \mu k(x) \forall x\}.$$ 

If $k$ is finite everywhere except the origin, one can instead write

$$k^\ominus(y) = \sup_{x \neq 0} \frac{\langle x, y \rangle}{k(x)}.$$ 

The notation $k^\ominus$ is justified since under the assumptions above,

$$(\check{\gamma}C)^\ominus = \check{\gamma}C^\ominus.$$
If $k$ is a concave-gauge (and thus non-negative, 1-homogeneous concave function with $k(0) = 0$) then the concave-polar of $k$ is defined as

$$k^\circ(y) = \sup\{\mu \geq 0 : \langle x, y \rangle \leq \mu k(x) \ \forall x\}.$$ 

If $k$ is finite everywhere except the origin, one can instead write

$$k^\circ(y) = \inf_{x \neq 0} \frac{\langle x, y \rangle}{k(x)}.$$ 

The notation $k^\circ$ is justified since under the assumptions above,

$$(\hat{\gamma}C)^\circ = \hat{\gamma}C^\circ.$$
Suppose that $S$ is of positive type and has concave support function $\hat{\sigma}_S$. Let

$$\ell_S := F_S = \hat{\partial} \hat{\sigma}_S.$$
Traditionally $L : \Delta^n \to \mathbb{R}$

Now we have $L : \mathbb{R}^n_+ \to \mathbb{R}$ but $L$ is 1-homogeneous

[A function $f : \mathbb{R}^n \to \mathbb{R}$ is 1-homogeneous (resp. 0-homogeneous) if for all $\alpha > 0$, $f(\alpha x) = \alpha f(x)$ (resp. $f(\alpha x) = f(x)$) for all $x \in \mathbb{R}^n$.]

Thus $\ell_S$ is 0-homogeneous

Losses well defined even for unnormalised probabilities

Losses are distorted probabilities

Example: If $S_{0-1} := \{e_1, \ldots, e_n\}$ then the zero-one loss

$$\ell_{0-1} = \hat{\partial} \hat{\sigma}_{S_{0-1}}.$$
Inverse losses needed for Vovk’s aggregating algorithm
- Have multiple predictions, with corresponding losses
- Combine predictors by combining their loss vectors. Seek an actual prediction $q$ corresponding to this pseudo-prediction

$$\ell_1(q)$$

$$\ell_2(q)$$

$$\ell(V)$$

$S_\ell$

Pseudo-prediction

$x = \ell(q)$

$\mathbf{1}_q = \{x : x \cdot q = \ell(q)\}$
What do we mean by “Inverse Loss”?

If $\ell(v) = x$ would like $\ell^{-1}(x) = v$. 
The Technical Tool we Need

**Theorem (Barbara and Crouzeix (1994))**

Suppose $C$ is a convex body of positive type. For all $s, d \in \mathbb{R}^n$,

$$\frac{d}{\hat{\gamma}_C(d)} \in \hat{\partial} \hat{\gamma}_C(s) \iff \frac{s}{\hat{\gamma}_C(s)} \in \hat{\partial} \hat{\gamma}_C(d) \iff \hat{\gamma}_C(d) \hat{\gamma}_C(s) = \langle s, d \rangle.$$
The Consequence . . . the Inverse Loss

Corollary

Given a set $S$ of positive type and hence a proper loss $\ell_S$, the inverse loss

$$\ell_S^{-1} = \ell_{S^\circ}.$$
The inverse is only up to a positive scaling
The inverse is only up to a positive scaling. One is not guaranteed that \( d = \ell_S^{-1}(\ell_S(d)) \) for all \( d \in \mathbb{R}^n_+ \). One is guaranteed that for all \( d \in \mathbb{R}^n_+ \),

\[
\ell_S(\ell_S^{-1}(\ell_S(d))) = \ell_S(d).
\]

(Confer Drazin pseudo-inverse)
The inverse is only up to a positive scaling

One is not guaranteed that \( d = \ell_S^{-1}(\ell_S(d)) \) for all \( d \in \mathbb{R}^n_+ \)

One is guaranteed that for all \( d \in \mathbb{R}^n_+ \),

\[
\ell_S(\ell_S^{-1}(\ell_S(d))) = \ell_S(d).
\]

(Confer Drazin pseudo-inverse)

Not only are losses distorted probabilities, but probabilities are distorted losses
Given a point $d \in \mathbb{R}^n_+$ it is desired to find $s \in \mathbb{R}^n$ such that $\ell_{\log}(s) = x$, where $x$ is the point of intersection of the line segment $[0, d]$ with the curve $\ell_{\log}(\mathbb{R}^n_+)$. Evaluating $\ell_{\log}^{-1}(d)$ gives the point $y$; observe the hyperplane $h_d$ with normal vector $d$ supports $\ell_{\log}^{-1}(\mathbb{R}^n_+)$ at the point $y$ because of properness.

Any positive scaling $s = \alpha y$, for $\alpha > 0$, would also suffice. It can be seen that $h_s$ (with normal vector $s$) supports $\ell_{\log}(\mathbb{R}^n_+)$ at $x$ (again due to properness).
Cf. Shephard’s *Cost and Production Functions* (1953)
Define the parametric family of concave gauges $\hat{\gamma}_p : \mathbb{R}^n_+ \to \mathbb{R}$ for $p \in [-\infty, 0) \cup (0, 1]$:

For $p \in (0, 1]$

$$\hat{\gamma}_p(x) := \begin{cases} \left( \sum_{i=1}^n x_i^p \right)^{\frac{1}{p}}, & x \in \mathbb{R}^n_+, \\ -\infty & \text{otherwise.} \end{cases}$$

For $p \in (-\infty, 0)$,

$$\hat{\gamma}_p(x) := \begin{cases} \left( \sum_{i=1}^n x_i^p \right)^{\frac{1}{p}}, & x \in \text{int} \mathbb{R}^n_+ \\ 0 & x \in \text{bd} \mathbb{R}^n_+ \\ -\infty & \text{otherwise} \end{cases}$$

and

$$\hat{\gamma}_{-\infty}(x) := \begin{cases} \bigwedge_{i=1}^n x_i, & x \in \mathbb{R}^n_+ \\ -\infty & \text{otherwise} \end{cases}$$

Barbara and Crouzeix (1994) show that for all $p \in [-\infty, 0) \cup (0, 1]$, $\hat{\gamma}_p$ is a concave gauge and if $\frac{1}{p} + \frac{1}{q} = 1$ then

$$\hat{\gamma}_p = \hat{\gamma}_q$$
$\ell_{2/3}^{-1} = \ell_{-2}$
Algebra of Convex Bodies and Hence Losses

\[
\begin{align*}
A \oplus_1 B &= A + B & \text{Minkowski sum} \\
A \oplus_\infty B &= \text{co}(A \cup B) & \text{convex hull of union} \\
A \odot_1 B &= A \cap B & \text{intersection} \\
A \odot_\infty B &= A \# B & \text{inverse sum} \\
&= \bigcup \{\lambda_1 A + \lambda_2 B : \lambda_1 + \lambda_2 = 1, \lambda_1 \geq 0, \lambda_2 \geq 0\}
\end{align*}
\]
Algebra of Convex Bodies and Hence Losses

\[
\begin{align*}
A \hat{\oplus}_1 B &= A + B \quad \text{Minkowski sum} \\
A \hat{\oplus}_\infty B &= \text{co}(A \cup B) \quad \text{convex hull of union} \\
A \hat{\sqcap}_1 B &= A \cap B \quad \text{intersection} \\
A \hat{\sqcap}_\infty B &= A \# B \quad \text{inverse sum} \\
&= \bigcup \{\lambda_1 A + \lambda_2 B : \lambda_1 + \lambda_2 = 1, \lambda_1 \geq 0, \lambda_2 \geq 0\}
\end{align*}
\]

- Seeger (1990) studied general operation on convex bodies in \( \mathbb{R}^n \) parametrised by a third convex body \( C \subset \mathbb{R}^2 \).
- He considered \( C = C_p = B_p \cap \mathbb{R}^2_+ \), where \( B_p = \{x : \|x\|_p \leq 1\} \) and \( p \in [1, \infty] \).
\[ A \hat{\cdot} B = A + B \quad \text{Minkowski sum} \]
\[ A \hat{\cdot} \infty B = \text{co}(A \cup B) \quad \text{convex hull of union} \]
\[ A \hat{\cdot} \boxgreater B = A \cap B \quad \text{intersection} \]
\[ A \hat{\cdot} \infty B = A \# B \quad \text{inverse sum} \]
\[ = \bigcup \{ \lambda_1 A + \lambda_2 B : \lambda_1 + \lambda_2 = 1, \lambda_1 \geq 0, \lambda_2 \geq 0 \} \]

Seeger (1990) studied general operation on convex bodies in \( \mathbb{R}^n \) parametrised by a third convex body \( C \subset \mathbb{R}^2 \).

He considered \( C = C_p = B_p \cap \mathbb{R}^2_+ \), where \( B_p = \{ x : \|x\|_p \leq 1 \} \) and \( p \in [1, \infty] \).

We extend to more general \( C \) — provides insight (and beauty!)

Blue expressions above have \( \hat{\cdot} C_p \) as \( \hat{\cdot} p \) and \( \hat{\cdot} C_p \) as \( \hat{\cdot} p \).
Let $A, B \subset \mathbb{R}^n$ be sets of negative type (closed, convex, and containing the origin). Let $C \subset \mathbb{R}_+^2$ also be of negative type. For a set $S$ and $\alpha \geq 0$ define

$$\alpha \star S := \begin{cases} 
\alpha S, & \alpha > 0 \\
0^+S, & \alpha = 0 
\end{cases}.$$  

$$A \mathbin{\hat{\ast}} C B := \bigcup_{\lambda \in C \otimes} \lambda_1 A + \lambda_2 B$$

$$A \mathbin{\hat{\otimes}} C B := \bigcup_{\lambda \in C \otimes} \lambda_1 \star A + \lambda_2 \star B$$

$$A \mathbin{\hat{\odot}} C B := \bigcup_{\lambda \in C \odot} \lambda_1 A \cap \lambda_2 B$$

$$A \mathbin{\hat{\boxdot}} C B := \bigcup_{\lambda \in C \boxdot} \lambda_1 \star A \cap \lambda_2 \star B$$
Closedness

Theorem

If \( A, B \subset \mathbb{R}^n \) are sets of negative type and \( C \subset \mathbb{R}^2_+ \) is of negative type then \( A \check{\bowtie}_C B \) and \( A \check{\ll} C B \) are also of negative type.

The operations \( A \check{\ll} C B \) and \( A \check{\lll} C B \) are identical when \( A \) and \( B \) are bounded since in that case \( 0 \bowtie A = 0 \bowtie B = \{0\} \). However for unbounded sets they differ. (Needed for pesky closure issues.)
Binary Operations on Functions

Definition

Suppose $C \subseteq \mathbb{R}_+^2$ is of negative type and suppose $f, g : \mathbb{R}^n \to [0, \infty]$. The direct and inverse sum of type $C$ of $f$ and $g$ are respectively

$$(f \odot_C g)(x^*) := \gamma_C((f(x^*), g(x^*))')$$

$$(f \oslash_C g)(x^*) := \inf_{x_1^* + x_2^* = x^*} \gamma_C((f(x_1^*), g(x_2^*))').$$

As with the set operations, for $p \in [1, \infty]$ we abbreviate $\odot_{C_p}$ by $\oplus_p$ and $\oslash_{C_p}$ by $\ominus_p$. Special cases of these operations are

$$(f \oplus_1 g)(x^*) = f(x^*) + g(x^*) \quad \text{sum}$$

$$(f \oplus_\infty g)(x^*) = f(x^*) \lor g(x^*) \quad \text{maximum}$$

$$(f \ominus_1 g)(x^*) = \inf_{x_1^* + x_2^* = x^*} (f(x_1^*) + g(x_2^*)) \quad \text{infimal convolution}$$

$$(f \ominus_\infty g)(x^*) = \inf_{x_1^* + x_2^* = x^*} (f(x_1^*) \lor g(x_2^*)) \quad \text{inf-max convolution}$$
Theorem

Suppose $C \subset \mathbb{R}^2_+$ is of negative type and $A, B \subset \mathbb{R}^n$ are of negative type. Then

$$\check{\sigma}_{A \check{\oplus} C B} = \check{\sigma}_A \check{\oplus} C \check{\sigma}_B$$

$$\check{\sigma}_{A \check{\boxdot} C B} = \check{\sigma}_A \check{\boxdot} C \check{\sigma}_B.$$ 

Theorem

Suppose $C \subset \mathbb{R}^2_+$ is of positive type and $A, B \subset \mathbb{R}^n$ are of positive type. Then

$$(A \check{\oplus} C B) \check{\circ} = A \check{\circ} \check{\boxdot} C \check{\circ} B \check{\circ}.$$
Can parametrize proper losses via convex bodies
Can parametrize proper losses via convex bodies
Loss function is subgradient of support function of the body
Conclusion

- Can parametrize proper losses via convex bodies
- Loss function is subgradient of support function of the body
- Losses so induced are 0-homogeneous
Can parametrize proper losses via convex bodies
Loss function is subgradient of support function of the body
Losses so induced are 0-homogeneous
Inverse losses are subgradient of support function of polar losses “≡” probabilities
Can calculate inverse loss explicitly for $l_p$ losses
(also Cobb-Douglas — see paper)
Can parametrize proper losses via convex bodies
Loss function is subgradient of support function of the body
Losses so induced are 0-homogeneous
Inverse losses are subgradient of support function of polar losses “≡” probabilities
Can calculate inverse loss explicitly for $l_p$ losses (also Cobb-Douglas — see paper)
Algebra on sets induces algebra of gains / losses (New general binary operation on convex sets)
Further Work

- Define $f$-divergences via convex bodies
- Bridge between divergences and losses becomes very simple
- Surrogate regret bounds via geometrical arguments
- Kernels between distributions

One might also be able to:

- Characterise convexity of loss function in terms of properties of $S$
- Study mixability and its generalisations in terms of $S$ — how does mixability constant behave under the binary operations?
- Use connection with kernels to extend kernels to multiple arguments