Learnability and models of decision making under uncertainty

Pathikrit Basu   Federico Echenique

Caltech

Virginia Tech DT Workshop – April 6, 2018

Pathikrit

To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes, there were nothing but details.

Jorge Luis Borges, “Funes el memorioso”

Motivation

Complex models vs. Occam's razor:

- Use a model of economic behavior to infer welfare
- Make choices for the agent.
- Complex models lead to overfitting.

"Uniform learnability" $\Leftrightarrow$ no overfitting $\Leftrightarrow$ simplicity

(these are applications of old ideas in ML)

Setup

- Ω a finite *state space.*
- $x \in X = \mathbf{R}^{\Omega}$ are *acts*
- $\succsim \subseteq X \times X = Z$ is a *preference*
- $\mathcal{P}$ is a class of preferences.

Learning (informal)

Model: $\mathcal{P}$

Data: choices generated by some $\succeq \in \mathcal{P}$

The choices are among pairs $(x, y) \in Z$ drawn from some *unknown* $\mu \in \Delta(Z)$.

(Uniform) learning: Get arbitrarily close to $\succeq$, with high prob. after a finite sample.

(Uniform) Poly-time learnable: Get arbitrarily close to $\succeq$, with high prob. w/sample size that doesn't explode with $|\Omega|$.

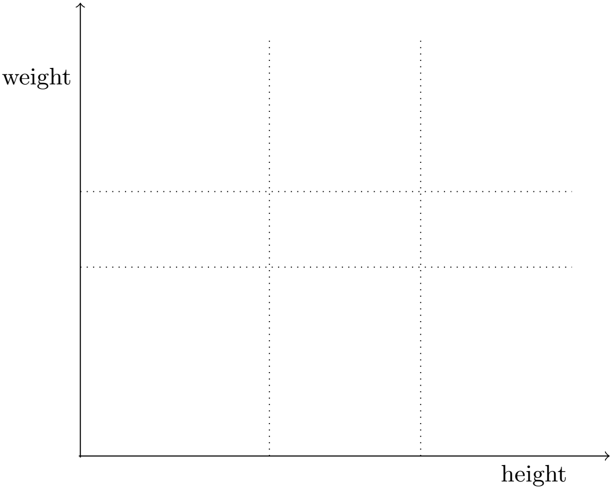
Our results

	Learnable	Sample complexity ($\|\Omega\|$)
Expected utility	✓	Linear
Maxmin (2 states)	✓	NA
Maxmin (states > 2)	X	$+\infty$
Choquet expected utility	✓	Exponential

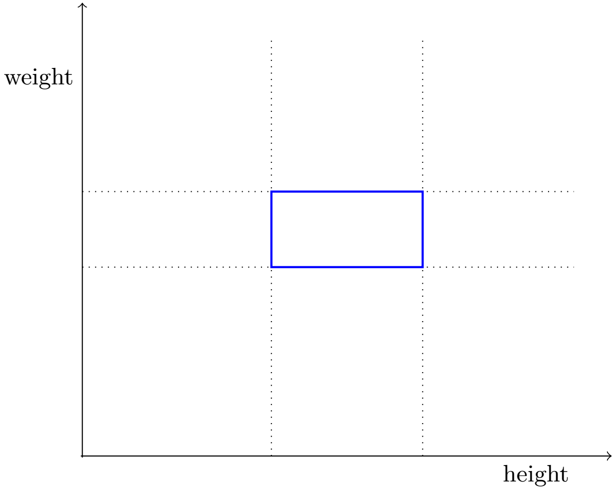
Table: Summary

What is a normal Martian?

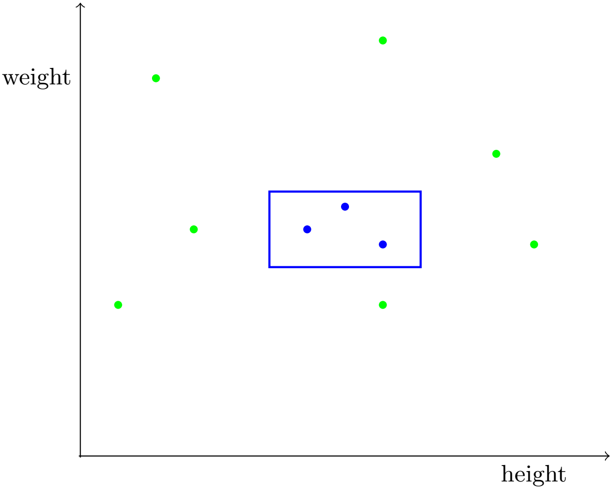
Digression



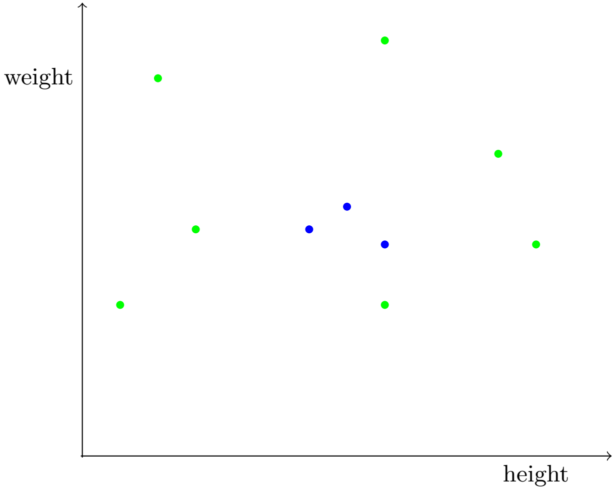
Digression



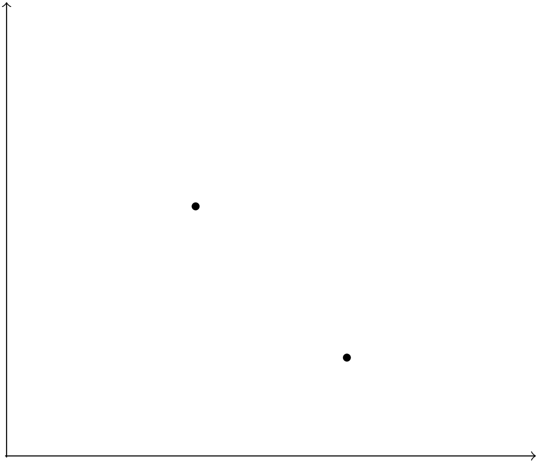
Digression



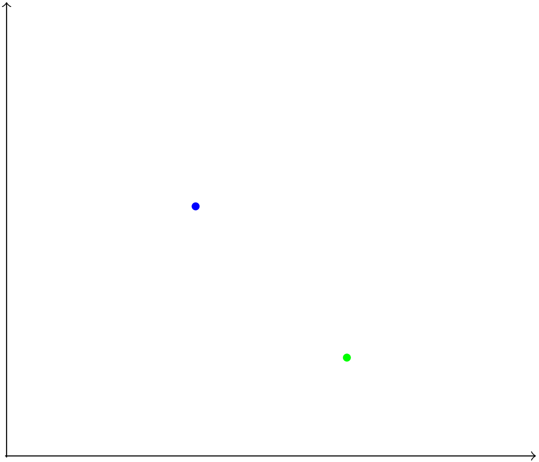
Digression



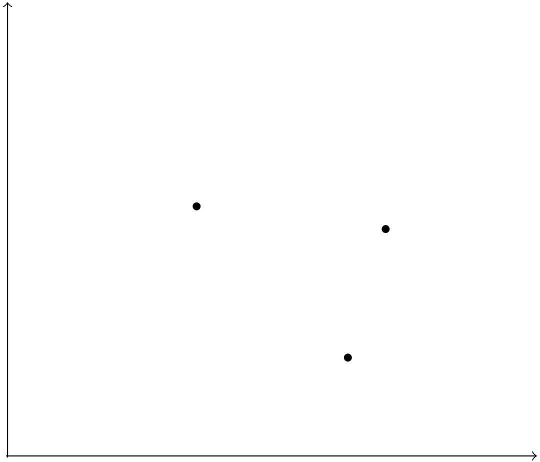
Digression



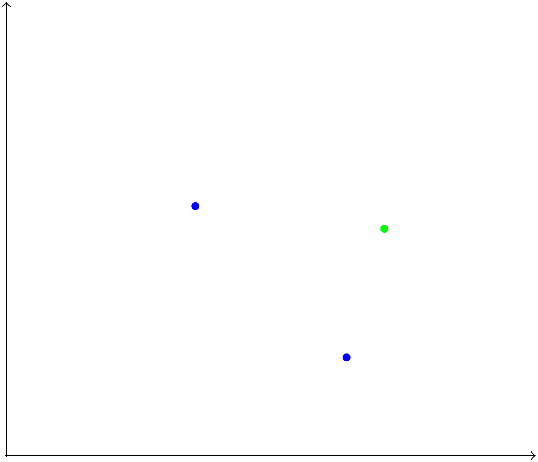
Digression



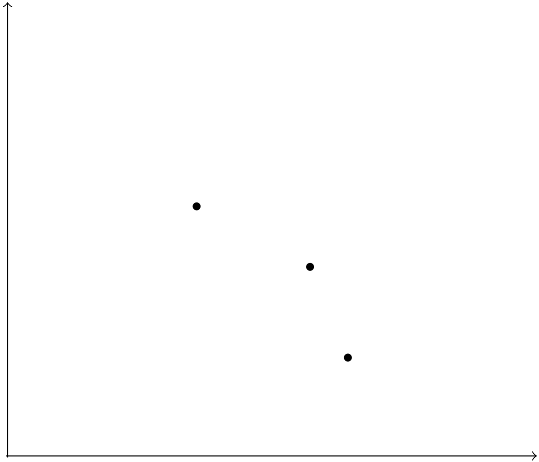
Digression



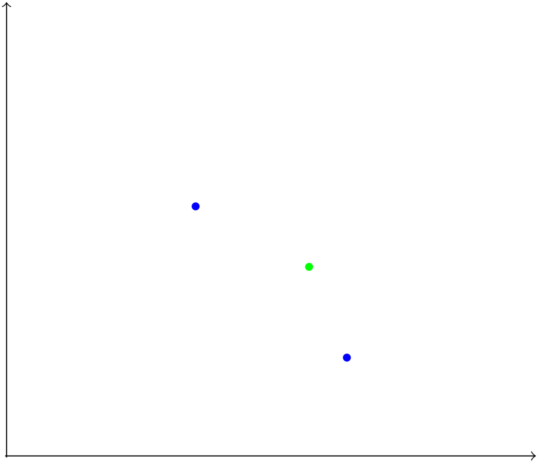
Digression



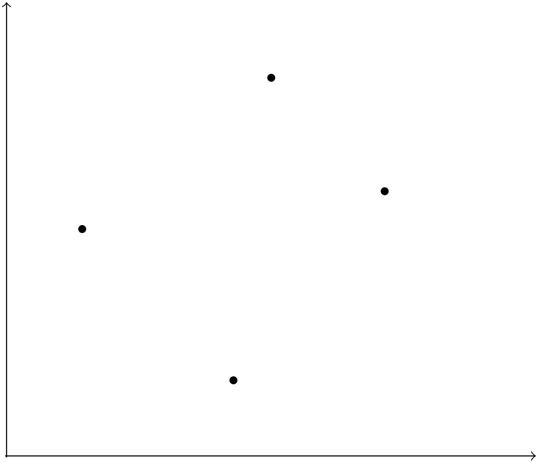
Digression



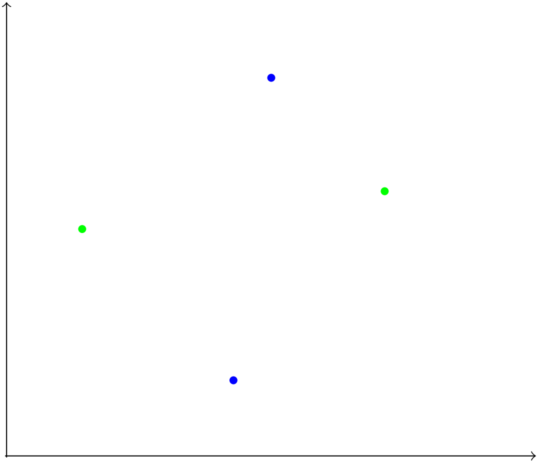
Digression



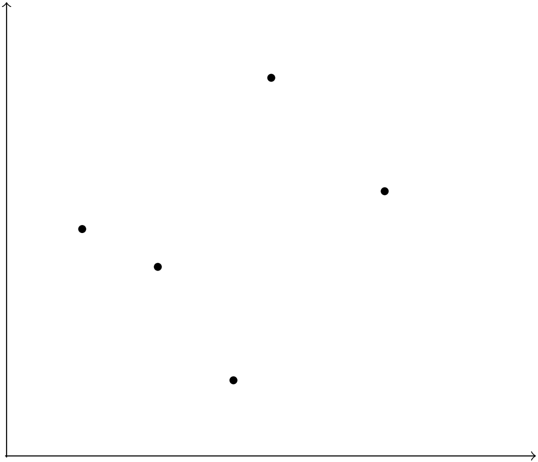
Digression



Digression



Digression



VC dimension

Let $\mathcal{P}$ be a collection of sets.
A finite set A is *always rationalized* ("shattered") by $\mathcal{P}$ if, no matter how A is labeled, $\mathcal{P}$ can rationalize it.

The Vapnik-Chervonenkis (*VC*) *dimension* of a collection of subsets is the *largest cardinality of a set that can always be rationalized.*

VC(rectangles) $= 4$.
VC(all finite sets) $= \infty$

VC dimension

$\Pi_{\mathcal{P}}(k)$ = the largest number of labelings that can be rationalized for a data of cardinality S.

A measure of how "rich" or "complex" $\mathcal{P}$ is. How prone to overfitting.

VC dimension

$\Pi_{\mathcal{P}}(k)$ = the largest number of labelings that can be rationalized for a data of cardinality S.

A measure of how "rich" or "complex" $\mathcal{P}$ is. How prone to overfitting.

Observe: if $k \leq VC(\mathcal{P})$ then $\Pi_{\mathcal{P}}(k) = 2^k$.

Thm (Sauer's lemma): If $VC(\mathcal{P}) = d$ then

$$\Pi_{\mathcal{P}}(k) \leq \left(\frac{ke}{d}\right)^d$$

for $k > d$.

Data

A dataset consists of a finite set of pairs $(x_i, y_i) \in Z$:

$$\begin{array}{cc} (x_1, y_1) & a_1 \\ (x_2, y_2) & a_2 \\ \vdots & \vdots \\ (x_n, y_2) & a_n, \end{array}$$

with a *labeling* $a_i \in \{0, 1\}$; where $a_i = 1$ iff x_i is chosen over y_i.

Data

A *dataset* is a finite sequence

$$D \in \bigcup_{n \geq 1} (Z \times \{0,1\})^n.$$

The set of all datasets is denoted by $\mathcal{D}$

Learning

A *learning rule* is a map $\sigma : \mathcal{D} \to \mathcal{P}$.

Data generating process

Given $\succsim \in \mathcal{P}$.

- $\mu \in \Delta(Z)$ (full support)
- (x, y) drawn iid $\sim \mu$
- (x, y) labeled according to $\succsim$.

Distance between $\succsim, \succsim' \in \mathcal{P}$:

$$d_\mu(\succsim, \succsim') = \mu(\succsim \triangle \succsim'),$$

where

$$\begin{aligned} \succsim \triangle \succsim' = & \{(x,y) \in Z : x \succsim y \text{ and } x \not\succsim' y\} \cup \\ & \{(x,y) \in Z : x \not\succsim y \text{ and } x \succsim' y\}. \end{aligned}$$

Learning

$\mathcal{P}' \subseteq \mathcal{P}$ is *learnable*, if $\exists$ a learning rule σ s.t.

$$\forall \varepsilon, \delta > 0 \quad \exists s(\varepsilon, \delta) \in \mathbf{N}$$

s.t. $\forall n \geq s(\varepsilon, \delta)$,

$$(\forall \succsim \in \mathcal{P}')(\forall \mu \in \Delta^f(Z))(\mu^n(d_\mu(\sigma_n, \succsim) > \varepsilon) < \delta)$$

Decisions under uncertainty

- Ω a finite *state space.*
- $x \in X = \mathbf{R}^{\Omega}$ are *acts*
- $\succsim \subseteq = X \times X = Z$ is a *preference*
- $\mathcal{P}$ is a class of preferences.

$x, y \in X$ are *comonotonic* if there are no ω, ω' s.t

$$x(\omega) > x(\omega') \text{ but } y(\omega) < y(\omega').$$

Axioms

- (*Weak order*) $\succsim$ is complete and transitive.
- (*Independence*) $\forall x, y, z \in X$ $\lambda \in (0,1)$,

$$x \succsim y \text{ iff } \lambda x + (1-\lambda)z \succsim \lambda y + (1-\lambda)z$$

- (*Continuity*) $\forall x \in X$,

$$U_x = \{y \in X \mid y \succsim x\} \text{ and } L_x = \{y \in X \mid x \succsim y\}$$

  are closed.
- (*Convex*) $\forall x \in X$, the upper contour set

$$U_x = \{y \in X \mid y \succsim x\}$$

  is a convex set.

Axioms

- (*Comonotic Independence*) $\forall x, y, z \in X$ that are comonotonic and $\lambda \in (0,1)$,

$$x \succsim y \text{ iff } \lambda x + (1-\lambda)z \succsim \lambda y + (1-\lambda)z$$

- (*C-Independence*) $\forall x, y \in X$, constant act $c \in X$ and $\lambda \in (0,1)$,

$$x \succsim y \text{ iff } \lambda x + (1-\lambda)c \succsim \lambda y + (1-\lambda)c$$

Decisions under uncertainty

- $\mathcal{P}_{\mathcal{EU}}$: set of preferences satisfying weak order and independence
- $\mathcal{P}_{\mathcal{MEU}}$: set of preferences satisfying weak order, monotonicity, c-independence, continuity, convexity and homotheticity.
- $\mathcal{P}_{\mathcal{CEU}}$: set of preferences satisfying comonotonic independence, continuity and monotonicity.

Theorem

- $VC(\mathcal{P}_{\mathcal{EU}}) = |\Omega| + 1.$
- *If* $|\Omega| \geq 3$*, then* $VC(\mathcal{P}_{\mathcal{MEU}}) = +\infty$ *and* $\mathcal{P}_{\mathcal{MEU}}$ *is not learnable*
- *If* $|\Omega| = 2$*, then* $VC(\mathcal{P}_{\mathcal{MEU}}) \leq 8$ *and* $\mathcal{P}_{\mathcal{MEU}}$ *is learnable.*
- $\binom{|\Omega|}{|\Omega|/2} \leq VC(\mathcal{P}_{\mathcal{CEU}}) \leq (|\Omega|!)^2(2|\Omega| + 1) + 1$

Corollary

- *$\mathcal{P}_{\mathcal{EU}}$, $\mathcal{P}_{\mathcal{CEU}}$ and, when $|\Omega| = 2$, $\mathcal{P}_{\mathcal{MEU}}$ are learnable.*
- *$\mathcal{P}_{\mathcal{EU}}$ requires a minimum sample size that grows linearly with $|\Omega|$,*
- *$\mathcal{P}_{\mathcal{CEU}}$ requires a minimum sample size that grows exponentially with $|\Omega|$.*
- *$\mathcal{P}_{\mathcal{MEU}}$ is not learnable when $|\Omega| \geq 3$.*

Ideas in the proof

For EU:

If $A \subseteq \mathbf{R}^n$ and $|A| \geq n+2$, then $A = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$ and $cvh(A_1) \cap cvh(A_2) \neq \emptyset$.

For max-min. $|\Omega| \geq 3$.

Model can be characterized by a single upper contour set $\{x : x \succeq 0\}$. This upper contour set is a closed convex cone. Consider a circle C in $\{x \in \mathbf{R}^\Omega : \sum_i x_i = 1\}$ distance 1 to $(1/2, \ldots, 1/2)$.
For any n, choose n points $x^1, \ldots, x^n$ on C: label any subset. The closed conic hull of the labeled points will exclude all the non-labeled points.

Ideas in the proof

For CEU:

For a large enough sample, a large enough number of acts must be comonotonic. Apply similar ideas to those used for EU to comonotonic acts, (via comonotonic independence).
This shows that VC is finite (and exact upper bound can be calculated).

Ideas in the proof

For the exponential-sized lower bound: choose exponentially many unordered events in Ω and consider a dataset of bets on each event. Since events are unordered one can construct a CEU that explains any labeling of the data.