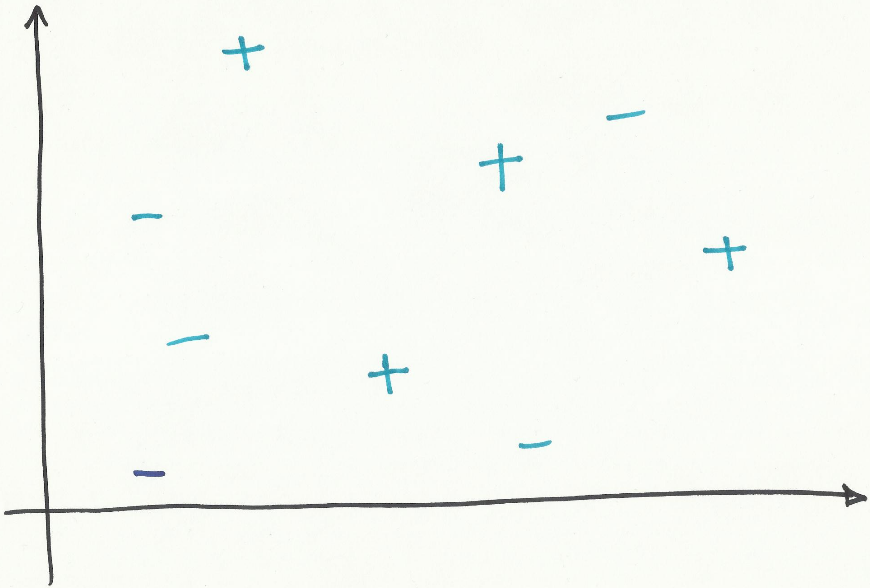
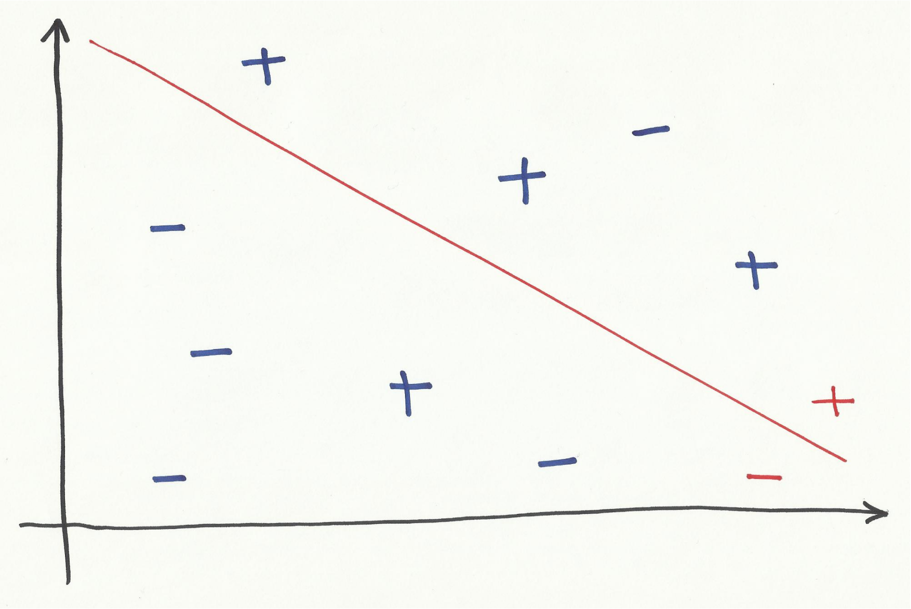
Active Property Testing of Linear Functions Over the Boolean Hypercube

Abbas Mehrabian
amehrabi@uwaterloo.ca

University of Waterloo

Theoretical Aspects of Computer Science
IPM, Tehran
December 22nd, 2012





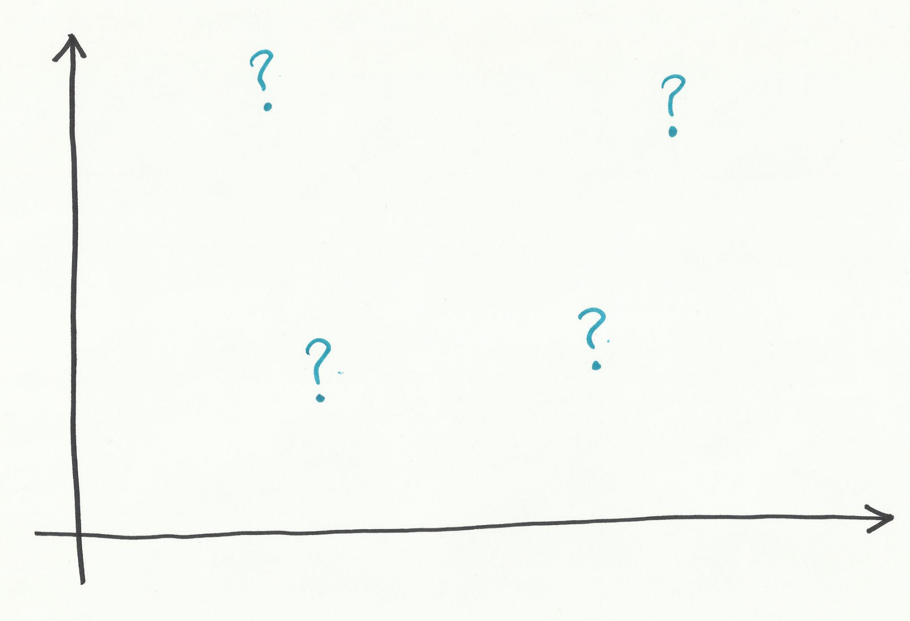
What is *Property Testing*?

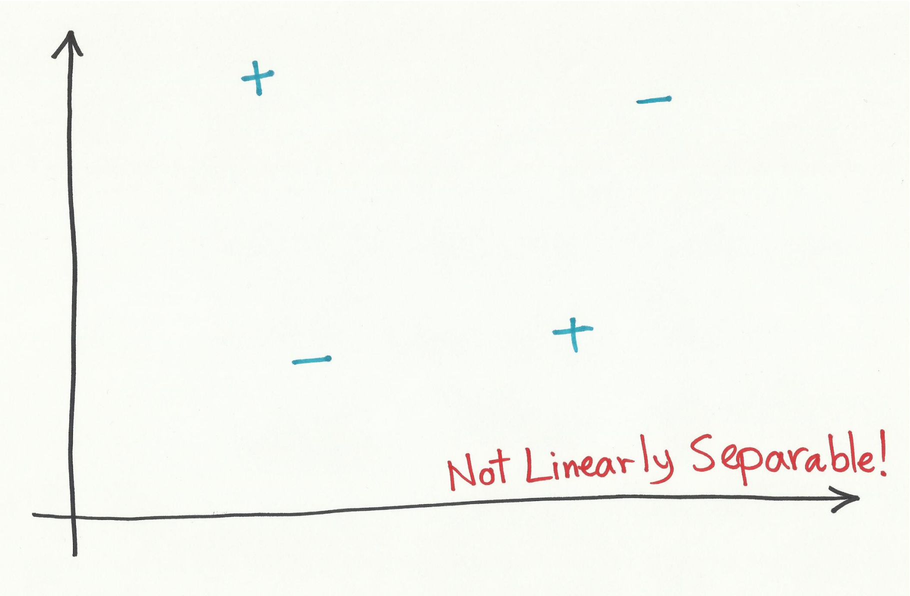
Property Testing is a relaxation of learning.

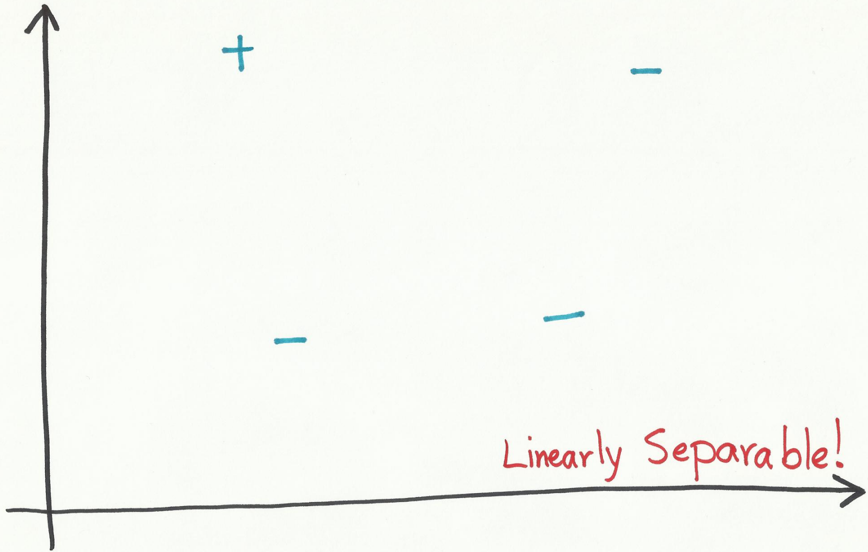
What is *Property Testing*?

Property Testing is a relaxation of learning.

Let's see an example ...




+
−
−
+
Not Linearly Separable!


Linearly Separable!

What is *Property Testing*?

Distance of functions

Definition

For functions $f, g : A \to B$

$$\mathrm{dist}(f, g) := \frac{\#\{a : f(a) \neq g(a)\}}{\#A}.$$

Definition

For function f and a class $\mathcal{C}$ of functions

$$\mathrm{dist}\,(f, \mathcal{C}) := \inf\{\mathrm{dist}(f, g) : g \in \mathcal{C}\}\,.$$

What is *Property Testing*?

Formal definition

Definition (Rubinfeld and Sudan'96)

A (property) tester for class $\mathcal{C}$

- randomized decision algorithm
- is given $\varepsilon > 0$
- asks value of f on q points
- If $f \in \mathcal{C}$ accepts with probability $\geq 2/3$
- If $\mathrm{dist}(f, \mathcal{C}) > \varepsilon$ rejects with probability $\geq 2/3$.

What is *Property Testing*?

Formal definition

Definition (Rubinfeld and Sudan'96)

A (property) tester for class $\mathcal{C}$

- randomized decision algorithm
- is given $\varepsilon > 0$
- asks value of f on q points
- If $f \in \mathcal{C}$ accepts with probability $\geq 2/3$
- If $\mathrm{dist}(f, \mathcal{C}) > \varepsilon$ rejects with probability $\geq 2/3$.

Important parameter is $q :=$ query complexity.

Definition

Function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ is linear if

$$\forall\, x, y \in \mathbb{Z}_2^n \qquad f(x) + f(y) = f(x + y).$$

Definition

Function $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$ is linear if

$$\forall\, x, y \in \mathbb{Z}_2^n \qquad f(x) + f(y) = f(x + y).$$

Theorem (Blum, Luby, Rubinfeld'93)

Query complexity of testing linear functions is $\Theta(1/\epsilon)$.

Property Testing of Linear Functions

Blum et al.'s algorithm

for $i = 0$ to $2/\varepsilon$ **do**
    Uniformly and independently select $X, Y \in \mathbb{Z}_2^n$.
    If $f(X) + f(Y) \neq f(X + Y)$ then reject.
end for
If no iteration caused rejection, then accept.

Property Testing of Linear Functions

Blum et al.'s algorithm

```
for i = 0 to 2/ε do
    Uniformly and independently select X, Y ∈ Z_2^n.
    If f(X) + f(Y) ≠ f(X + Y) then reject.
end for
If no iteration caused rejection, then accept.
```

Key equation: If $\mathcal{L}_n$ is the class of linear functions,

$$\mathbb{P}_{X,Y}\left[f(X) + f(Y) \neq f(X+Y)\right] > \mathrm{dist}(f, \mathcal{L}_n).$$

What is *Active Property Testing*?

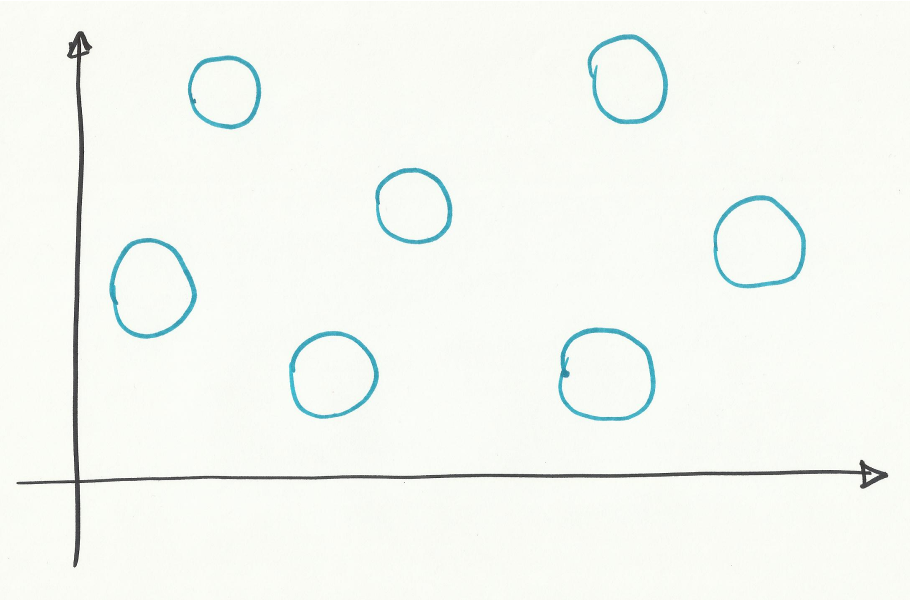
A Motivating Example

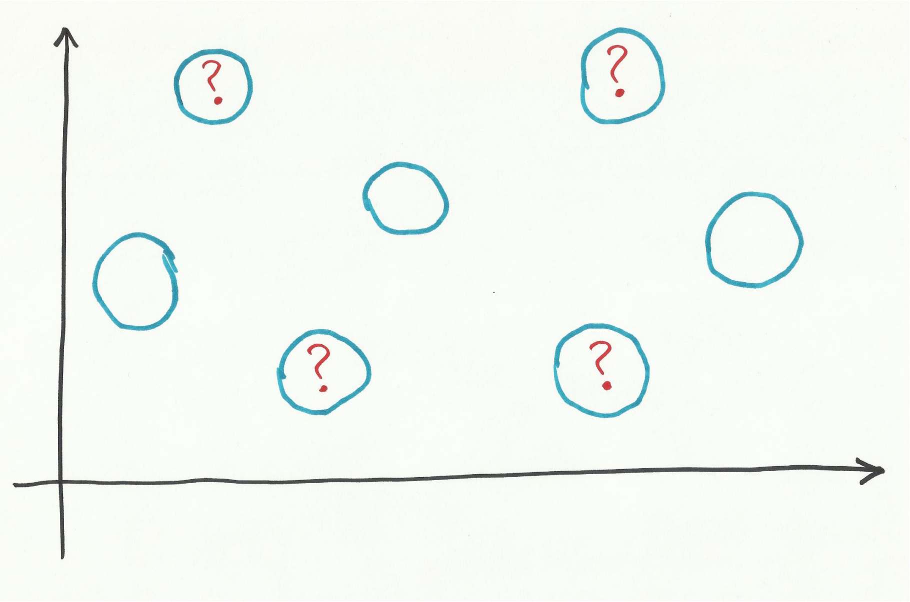
A patient is described as a vector of various features, e.g.:
(weight, age, blood pressure, blood sugar) $\in \mathbb{R}^4$

What is *Active Property Testing*?

A Motivating Example

A patient is described as a vector of various features, e.g.:
(weight, age, blood pressure, blood sugar) $\in \mathbb{R}^4$
Want to learn a function $\mathtt{diabetes} : \mathbb{R}^4 \to \{0, 1\}$.





What is *Active Property Testing*?

Formal definition

Definition (Balcan, Blais, Blum, Yang'12)

An active (property) tester for class $\mathcal{C}$

- randomized decision algorithm
- is given $\varepsilon > 0$
- is given a random sample
- asks value of f on q sample points
- If $f \in \mathcal{C}$ accepts with probability $\geq 2/3$
- If $\mathrm{dist}(f, \mathcal{C}) > \varepsilon$ rejects with probability $\geq 2/3$.

What is *Active Property Testing*?

Formal definition

Definition (Balcan, Blais, Blum, Yang'12)

An active (property) tester for class $\mathcal{C}$

- randomized decision algorithm
- is given $\varepsilon > 0$
- is given a random sample
- asks value of f on q sample points
- If $f \in \mathcal{C}$ accepts with probability $\geq 2/3$
- If $\mathrm{dist}(f, \mathcal{C}) > \varepsilon$ rejects with probability $\geq 2/3$.

Important parameter is $q :=$ query complexity.

What is *Active Property Testing*?

Formal definition

Definition (Balcan, Blais, Blum, Yang'12)

An active (property) tester for class $\mathcal{C}$

- randomized decision algorithm
- is given $\varepsilon > 0$
- is given a random sample
- asks value of f on q sample points
- If $f \in \mathcal{C}$ accepts with probability $\geq 2/3$
- If $\text{dist}(f, \mathcal{C}) > \varepsilon$ rejects with probability $\geq 2/3$.

Important parameter is $q :=$ query complexity.

Sample size $<<$ Domain size

Active property testing of linear functions $\mathbb{Z}_2^n \to \mathbb{Z}_2$.

Theorem (M'12+)

$$c_1 \frac{n}{\log n} \leq \text{Query complexity for active property testing of linear functions} \leq c_2 \frac{n}{\varepsilon \log n}.$$

Active Property Testing of Linear Functions

The algorithm

```
m ← n/ log n
for i = 0 to 16/ε do
    Sample a set S of size n^2 from Z_2^n
    choose a random m-tuple (X_1, X_2, ..., X_m) ∈ S^m with Σ_{i=1}^m X_i = 0
    choose a random (m−1)-tuple (Y_1, Y_2, ..., Y_{m−1}) ∈ S^{m−1}
    with Σ_{i=1}^{m−1} Y_i = 0
    if Σ_{i=1}^m f(X_i) ≠ 0 then reject.
    if Σ_{i=1}^{m−1} f(Y_i) ≠ 0 then reject.
end for
If no iteration caused rejection, then accept.
```

$m \leftarrow n/\log n$
for $i = 0$ to $16/\varepsilon$ **do**
    Sample a set S of size n^2 from $\mathbb{Z}_2^n$
    choose a random m-tuple $(X_1, X_2, \ldots, X_m) \in S^m$ with $\sum_{i=1}^{m} X_i = 0$
    choose a random $(m-1)$-tuple $(Y_1, Y_2, \ldots, Y_{m-1}) \in S^{m-1}$ with $\sum_{i=1}^{m-1} Y_i = 0$
    if $\sum_{i=1}^{m} f(X_i) \neq 0$ then reject.
    if $\sum_{i=1}^{m-1} f(Y_i) \neq 0$ then reject.
end for
If no iteration caused rejection, then accept.

1. Existence of $m = n/\log n$ elements adding up to zero (tight!)
2. If function is far from linear, probability of failure is large

Active property testing of linear functions $\mathbb{Z}_2^n \to \mathbb{Z}_2$.

Theorem (M'12+)

$$c_1 \frac{n}{\log n} \leq \text{Query complexity} \leq c_2 \frac{n}{\varepsilon \log n}.$$

Active property testing of linear functions $\mathbb{Z}_2^n \to \mathbb{Z}_2$.

Theorem (M'12+)

$$c_1 \frac{n}{\log n} \leq \text{Query complexity} \leq c_2 \frac{n}{\varepsilon \log n}.$$

1. Close the gap
2. Generalize distribution
3. Generalize domain and range

Active property testing of linear functions $\mathbb{Z}_2^n \to \mathbb{Z}_2$.

Theorem (M'12+)

$$c_1 \frac{n}{\log n} \leq \text{Query complexity} \leq c_2 \frac{n}{\varepsilon \log n}.$$

1. Close the gap
2. Generalize distribution
3. Generalize domain and range

Thanks for your attention :-)