

# Minería de Datos ¿Revolución Científica y Empresarial?

Alvaro J. Riascos Villegas  
Universidad de los Andes y Quantil

Octubre 31 de 2014

# Contenido

- 1 **Introducción**
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Cuál es la revolución científica?
- 6 La revolución empresarial
- 7 Quantil

## Introducción

- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ( $10^6 GB = 10^3 TB$ ) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.
- Una base de datos como la base de suficiencia que anualmente procesa el Ministerio de Salud tiene aproximadamente 400 millones de registros.

## Introducción

- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ( $10^6 GB = 10^3 TB$ ) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.
- Una base de datos como la base de suficiencia que anualmente procesa el Ministerio de Salud tiene aproximadamente 400 millones de registros.

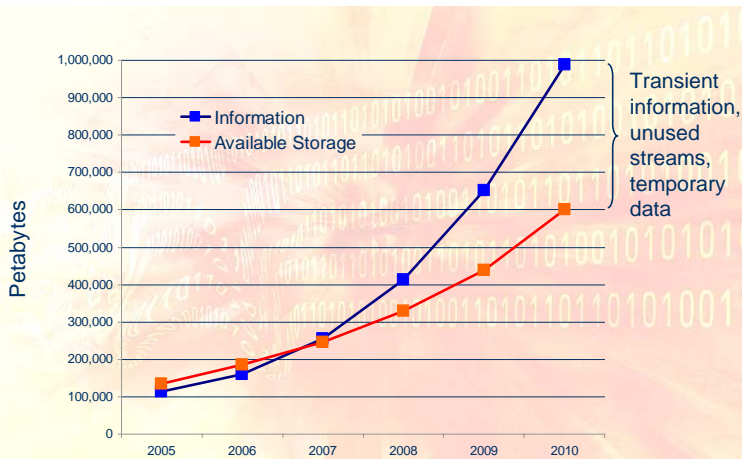
# Introducción

- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ( $10^6 GB = 10^3 TB$ ) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.
- Una base de datos como la base de suficiencia que anualmente procesa el Ministerio de Salud tiene aproximadamente 400 millones de registros.

# Introducción

- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ( $10^6 GB = 10^3 TB$ ) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.
- Una base de datos como la base de suficiencia que anualmente procesa el Ministerio de Salud tiene aproximadamente 400 millones de registros.

# Introducción



# Introducción

**Table 1: Worldwide production of original content, stored digitally using standard compression methods, in terabytes circa 1999.**

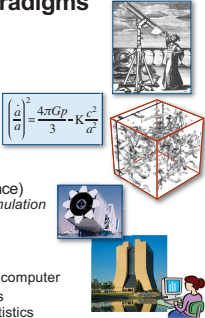
Storage Medium	Type of Content	Terabytes/Year, Upper Estimate	Terabytes/Year, Lower Estimate	Growth Rate, %
<a href="#">Paper</a>	Books	8	1	2
	Newspapers	25	2	-2
	Periodicals	12	1	2
	Office documents	195	19	2
	<b>Subtotal:</b>	<b>240</b>	<b>23</b>	<b>2</b>
<a href="#">Film</a>	Photographs	410,000	41,000	5
	Cinema	16	16	3
	X-Rays	17,200	17,200	2
	<b>Subtotal:</b>	<b>427,216</b>	<b>58,216</b>	<b>4</b>
<a href="#">Optical</a>	Music CDs	58	6	3
	Data CDs	3	3	2
	DVDs	22	22	100
	<b>Subtotal:</b>	<b>83</b>	<b>31</b>	<b>70</b>
<a href="#">Magnetic</a>	Camcorder Tape	300,000	300,000	5
	PC Disk Drives	766,000	7,660	100
	Departmental Servers	460,000	161,000	100
	Enterprise Servers	167,000	108,550	100
	<b>Subtotal:</b>	<b>1,693,000</b>	<b>577,210</b>	<b>55</b>
<b>TOTAL:</b>		<b>2,120,539</b>	<b>635,480</b>	<b>50</b>



- Una nueva forma de hacer ciencia.

### Science Paradigms

- Thousand years ago:  
science was **empirical**  
*describing natural phenomena*
- Last few hundred years:  
**theoretical** branch  
*using models, generalizations*
- Last few decades:  
a **computational** branch  
*simulating complex phenomena*
- Today: **data exploration** (eScience)  
*unify theory, experiment, and simulation*
  - Data captured by instruments  
or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files  
using data management and statistics



The complex block contains several images: a historical figure with a telescope (top right), a mathematical equation  $\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$  (middle), a 3D data visualization (middle right), a satellite dish (bottom middle), and a modern scientist at a computer (bottom right).

- Tomado de *The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research*

# Introducción

- El proceso típico de minería de datos.

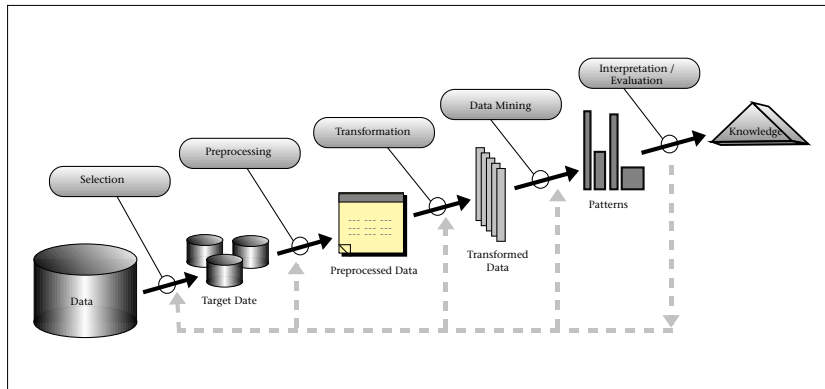
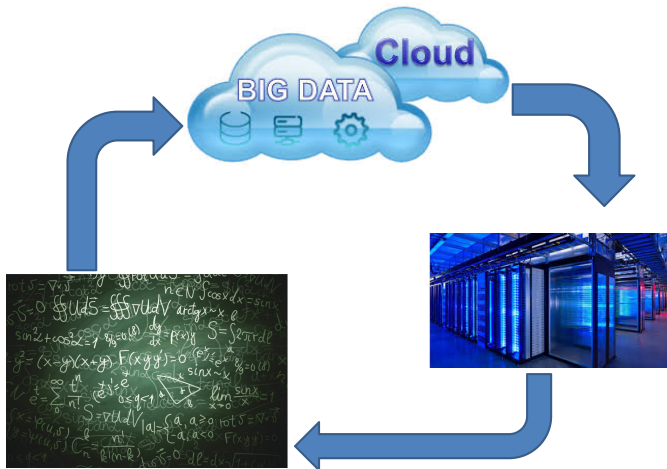


Figure 1. An Overview of the Steps That Compose the KDD Process.

- KDD: *Knowledge Discovery in Datasets*

# Introducción: La verdadera revolución



# Contenido

- 1 Introducción
- 2 Matemáticas en acción**
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Cuál es la revolución científica?
- 6 La revolución empresarial
- 7 Quantil

# Matemáticas en acción

- Un psicólogo de una prisión en California se acercó al Departamento de Estadística de Stanford con esto:

The image shows a piece of lined paper with several lines of handwritten mathematical symbols and numbers. The handwriting is very dense and difficult to decipher, appearing as a series of scribbles and symbols rather than clear equations. Some recognizable symbols include numbers like 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, and various mathematical symbols like  $\frac{1}{2}$ ,  $\frac{1}{3}$ ,  $\frac{1}{4}$ ,  $\frac{1}{5}$ ,  $\frac{1}{6}$ ,  $\frac{1}{7}$ ,  $\frac{1}{8}$ ,  $\frac{1}{9}$ ,  $\frac{1}{10}$ ,  $\frac{1}{11}$ ,  $\frac{1}{12}$ ,  $\frac{1}{13}$ ,  $\frac{1}{14}$ ,  $\frac{1}{15}$ ,  $\frac{1}{16}$ ,  $\frac{1}{17}$ ,  $\frac{1}{18}$ ,  $\frac{1}{19}$ ,  $\frac{1}{20}$ ,  $\frac{1}{21}$ ,  $\frac{1}{22}$ ,  $\frac{1}{23}$ ,  $\frac{1}{24}$ ,  $\frac{1}{25}$ ,  $\frac{1}{26}$ ,  $\frac{1}{27}$ ,  $\frac{1}{28}$ ,  $\frac{1}{29}$ ,  $\frac{1}{30}$ ,  $\frac{1}{31}$ ,  $\frac{1}{32}$ ,  $\frac{1}{33}$ ,  $\frac{1}{34}$ ,  $\frac{1}{35}$ ,  $\frac{1}{36}$ ,  $\frac{1}{37}$ ,  $\frac{1}{38}$ ,  $\frac{1}{39}$ ,  $\frac{1}{40}$ ,  $\frac{1}{41}$ ,  $\frac{1}{42}$ ,  $\frac{1}{43}$ ,  $\frac{1}{44}$ ,  $\frac{1}{45}$ ,  $\frac{1}{46}$ ,  $\frac{1}{47}$ ,  $\frac{1}{48}$ ,  $\frac{1}{49}$ ,  $\frac{1}{50}$ ,  $\frac{1}{51}$ ,  $\frac{1}{52}$ ,  $\frac{1}{53}$ ,  $\frac{1}{54}$ ,  $\frac{1}{55}$ ,  $\frac{1}{56}$ ,  $\frac{1}{57}$ ,  $\frac{1}{58}$ ,  $\frac{1}{59}$ ,  $\frac{1}{60}$ ,  $\frac{1}{61}$ ,  $\frac{1}{62}$ ,  $\frac{1}{63}$ ,  $\frac{1}{64}$ ,  $\frac{1}{65}$ ,  $\frac{1}{66}$ ,  $\frac{1}{67}$ ,  $\frac{1}{68}$ ,  $\frac{1}{69}$ ,  $\frac{1}{70}$ ,  $\frac{1}{71}$ ,  $\frac{1}{72}$ ,  $\frac{1}{73}$ ,  $\frac{1}{74}$ ,  $\frac{1}{75}$ ,  $\frac{1}{76}$ ,  $\frac{1}{77}$ ,  $\frac{1}{78}$ ,  $\frac{1}{79}$ ,  $\frac{1}{80}$ ,  $\frac{1}{81}$ ,  $\frac{1}{82}$ ,  $\frac{1}{83}$ ,  $\frac{1}{84}$ ,  $\frac{1}{85}$ ,  $\frac{1}{86}$ ,  $\frac{1}{87}$ ,  $\frac{1}{88}$ ,  $\frac{1}{89}$ ,  $\frac{1}{90}$ ,  $\frac{1}{91}$ ,  $\frac{1}{92}$ ,  $\frac{1}{93}$ ,  $\frac{1}{94}$ ,  $\frac{1}{95}$ ,  $\frac{1}{96}$ ,  $\frac{1}{97}$ ,  $\frac{1}{98}$ ,  $\frac{1}{99}$ ,  $\frac{1}{100}$ .

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para  $f$  es muy grande.
- Para juzgar la plausibilidad de que una  $f$  en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo  $x$  seguía el símbolo  $y$ .
- Interpretamos esto como la probabilidad de transición de  $x$  a  $y$  en el idioma Inglés (i.e. una cadena de Markow). Denotamos por  $M(x, y)$  esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para  $f$  es muy grande.
- Para juzgar la plausibilidad de que una  $f$  en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo  $x$  seguía el símbolo  $y$ .
- Interpretamos esto como la probabilidad de transición de  $x$  a  $y$  en el idioma Inglés (i.e. una cadena de Markow). Denotamos por  $M(x, y)$  esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para  $f$  es muy grande.
- Para juzgar la plausibilidad de que una  $f$  en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo  $x$  seguía el símbolo  $y$ .
- Interpretamos esto como la probabilidad de transición de  $x$  a  $y$  en el idioma Inglés (i.e. una cadena de Markow). Denotamos por  $M(x, y)$  esta frecuencia.



- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para  $f$  es muy grande.
- Para juzgar la plausibilidad de que una  $f$  en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo  $x$  seguía el símbolo  $y$ .
- Interpretamos esto como la probabilidad de transición de  $x$  a  $y$  en el idioma Inglés (i.e. una cadena de Markov). Denotamos por  $M(x, y)$  esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para  $f$  es muy grande.
- Para juzgar la plausibilidad de que una  $f$  en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo  $x$  seguía el símbolo  $y$ .
- Interpretamos esto como la probabilidad de transición de  $x$  a  $y$  en el idioma Inglés (i.e. una cadena de Markow). Denotamos por  $M(x, y)$  esta frecuencia.

- Ahora resolvemos este problema:

$$\max_{f \in \mathfrak{F}} \prod M(f(s_i), f(s_{i+1})) \quad (1)$$

- Este problema se resuelve usando al algoritmo de Metropolis - Hasting (un truco muy creativo!)
- Ahora, por qué esto funcionaría? Para ganar un poco de confianza en el modelo veamos el resultado en un caso conocido.

- Ahora resolvemos este problema:

$$\max_{f \in \mathfrak{F}} \prod M(f(s_i), f(s_{i+1})) \quad (1)$$

- Este problema se resuelve usando al algoritmo de Metropolis - Hasting (un truco muy creativo!)
- Ahora, por qué esto funcionaría? Para ganar un poco de confianza en el modelo veamos el resultado en un caso conocido.

- Ahora resolvemos este problema:

$$\max_{f \in \mathfrak{F}} \prod M(f(s_i), f(s_{i+1})) \quad (1)$$

- Este problema se resuelve usando al algoritmo de Metropolis - Hasting (un truco muy creativo!)
- Ahora, por qué esto funcionaría? Para ganar un poco de confianza en el modelo veamos el resultado en un caso conocido.

- Revolver todos los símbolos de este párrafo de Hamlet:

ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS  
NOBLER IN THE MIND TO SUFFER THE SLINGS AND ARROWS OF OUTRAGEOUS  
FORTUNE OR TO TAKE ARMS AGAINST A SEA OF TROUBLES AND BY OPPOSING END

- Metropolis Hasting despues de  $n$  iteraciones:

```
100 ER ENOHDLAE OHDLO UOZEOUNORU O UOZEO HD OITO HEOQSET IUROFHE HENO ITORUZAEN
200 ES ELOHRNDE OHRNO UOVEOULOSU O UOVEO HR OITO HEOQAET IUSOPHE HELO ITOSUVDEL
300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOPHE HELO ITOSUVDEL
400 ES ELOHINME OHINO UOVEOULOSU O UOVEO HI OATO HEOQRET AUSOWHE HELO ATOSUVMEL
500 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
600 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
900 ES ELOHANME OHANO UODEOULOSU O UODEO HA OITO HEOQRET IUSOWHE HELO ITOSUDMEL
1000 IS ILOHANMI OHANO RODIORLOS R O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODILO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODILO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODILO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFILO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHEHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHEHEL TIN SOBREL
1800 ESTER HAMLET HAM TO BE OR SOT TO BE THAT IN THE QUENTIOS WHETHER TIN SOBLER
1900 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
2000 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
```

- Resultado decodificación mensaje presos:

to bat-rb. con todo mi respeto. i was sitting down playing chess with danny de emf and boxer de el centro was sitting next to us. boxer was making loud and loud voices so i tell him por favor can you kick back homie cause im playing chess a minute later the vato starts back up again so this time i tell him con respecto homie can you kick back. the vato stop for a minute and he starts up again so i tell him check this out shut the f\*\*k up cause im tired of your voice and if you got a problem with it we can go to celda and handle it. i really felt disrespected thats why i told him. anyways after i tell him that the next thing I know that vato slashes me and leaves. dy the time i figure im hit i try to get away but the c.o. is walking in my direction and he gets me right dy a celda. so i go to the hole. when im in the hole my home boys hit doxer so now "b" is also in the hole. while im in the hole im getting schoold wrong and



# Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas**
- 4 Computación y sus límites
- 5 ¿Cuál es la revolución científica?
- 6 La revolución empresarial
- 7 Quantil

## Límites al aprendizaje de máquinas

### Theorem (Arbitrariamente cerca al clasificador aleatorio)

*Sea  $f$  un clasificador entrenado con una muestra finita. Entonces existe una distribución de probabilidad  $P$  (generador de datos) tal que:*

- 1 *En promedio  $P$  genera la muestra finita.*
- 2 *El error de clasificación de  $f$  de acuerdo a  $P$  es arbitrariamente cerca a 0,5.*

- No hay forma de aprender sin hipótesis sobre la estructura del fenómeno que se quiere estudiar (una teoría!).

## Límites al aprendizaje de máquinas

### Theorem (Arbitrariamente cerca al clasificador aleatorio)

*Sea  $f$  un clasificador entrenado con una muestra finita. Entonces existe una distribución de probabilidad  $P$  (generador de datos) tal que:*

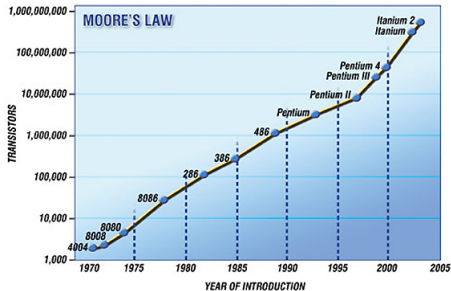
- 1 *En promedio  $P$  genera la muestra finita.*
  - 2 *El error de clasificación de  $f$  de acuerdo a  $P$  es arbitrariamente cerca a 0,5.*
- No hay forma de aprender sin hipótesis sobre la estructura del fenómeno que se quiere estudiar (una teorá!).

# Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites**
- 5 ¿Cuál es la revolución científica?
- 6 La revolución empresarial
- 7 Quantil

# Computación y sus límites

- Pocas dudas sobre la capacidad de calcular de los computadores.
- Moore's law



- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
  - 1 La capacidad de almacenamiento del Universo visible es:  $10^{92}$  bits.
  - 2 Número de operaciones realizadas desde el Big Bang:  $10^{122}$ .
  - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits.

- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
  - 1 La capacidad de almacenamiento del Universo visible es:  $10^{92}$  bits.
  - 2 Número de operaciones realizadas desde el Big Bang:  $10^{122}$ .
  - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.



# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

# Computación y sus límites

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

- Juego de Rabin (1957). Dos jugadores  $A$  y  $B$ , y un conjunto  $W \subset \mathbb{N}^3$ .
- Reglas: En tres rondas,  $A$  y  $B$  escogen tres números naturales.
- El resultado final es una tripla  $(a, b, c)$ .
- Si  $(a, b, c) \in W$  gana  $B$ .
- ¿Qué sabemos sobre este juego?
  - 1 Es determinado (inducción hacia atrás).
  - 2 Existe  $W$  decidible tal que el juego es determinado en favor de  $B$ .
  - 3  $B$  no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.



# Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Cuál es la revolución científica?**
- 6 La revolución empresarial
- 7 Quantil

## ¿Cuál es la revolución científica?

- En mi opinión no es el fin de la teoría como se profesa:
  - Anderson: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete
  - Napoletani, Panza, Struppa: Agnostic Science. Towards a Philosophy of Data Analysis.
  - Big data: A revolution that will transform how we live work and think.

# ¿Cuál es la revolución científica?

- En mi opinión la revolución está por llegar y es en esta dirección:
- Humphreys (computational science and scientific method, extending ourselves, the philosophical novelty of computer simulation methods).
  - *...our own intellectual and computational capabilities as human beings is no more a benchmark of scientific thought...*
  - *...the situation within which humans deal with science that is carried out at least in part by machines [is] the hybrid scenario and the more extreme situation of a completely automated science [is] the automated scenario*
- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steve Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*

# ¿Cuál es la revolución científica?

- En mi opinión la revolución está por llegar y es en esta dirección:
- Humphreys (computational science and scientific method, extending ourselves, the philosophical novelty of computer simulation methods).
  - *...our own intellectual and computational capabilities as human beings is no more a benchmark of scientific thought...*
  - *...the situation within which humans deal with science that is carried out at least in part by machines [is] the hybrid scenario and the more extreme situation of a completely automated science [is] the automated scenario*
- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steve Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*

# ¿Cuál es la revolución científica?

- En mi opinión la revolución está por llegar y es en esta dirección:
- Humphreys (computational science and scientific method, extending ourselves, the philosophical novelty of computer simulation methods).
  - *...our own intellectual and computational capabilities as human beings is no more a benchmark of scientific thought...*
  - *...the situation within which humans deal with science that is carried out at least in part by machines [is] the hybrid scenario and the more extreme situation of a completely automated science [is] the automated scenario*
- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steve Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*

# ¿Cuál es la revolución científica?

- En mi opinión la revolución está por llegar y es en esta dirección:
- Humphreys (computational science and scientific method, extending ourselves, the philosophical novelty of computer simulation methods).
  - *...our own intellectual and computational capabilities as human beings is no more a benchmark of scientific thought...*
  - *...the situation within which humans deal with science that is carried out at least in part by machines [is] the hybrid scenario and the more extreme situation of a completely automated science [is] the automated scenario*
- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steve Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*

# ¿Cuál es la revolución científica?

- En mi opinión la revolución está por llegar y es en esta dirección:
- Humphreys (computational science and scientific method, extending ourselves, the philosophical novelty of computer simulation methods).
  - *...our own intellectual and computational capabilities as human beings is no more a benchmark of scientific thought...*
  - *...the situation within which humans deal with science that is carried out at least in part by machines [is] the hybrid scenario and the more extreme situation of a completely automated science [is] the automated scenario*
- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steve Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*

# ¿Cuál es la revolución científica?

- En mi opinión la revolución está por llegar y es en esta dirección:
- Humphreys (computational science and scientific method, extending ourselves, the philosophical novelty of computer simulation methods).
  - *...our own intellectual and computational capabilities as human beings is no more a benchmark of scientific thought...*
  - *...the situation within which humans deal with science that is carried out at least in part by machines [is] the hybrid scenario and the more extreme situation of a completely automated science [is] the automated scenario*
- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steve Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*



# Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Cuál es la revolución científica?
- 6 La revolución empresarial**
- 7 Quantil

## La revolución empresarial

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX

## La revolución empresarial

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX

## La revolución empresarial

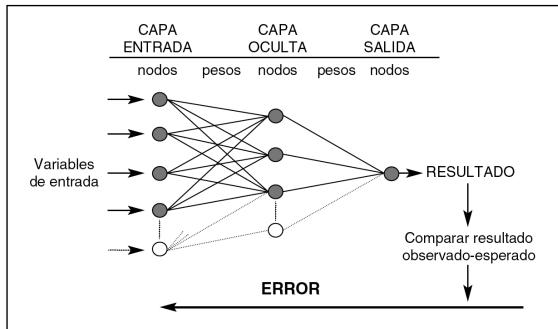
- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX

# La revolución empresarial

- Lo que hace particularmente interesante esta revolución empresarial es que está surgiendo de la mano de la academia y los empresarios.
- Considere el problema de reconocimiento de voz. Los grandes competidores son Google, IBM y Microsoft.

# La revolución empresarial: *deep learning*

- Big data, capacidad de computo y redes neuronales.



- Una red típica tiene entre 5 y 7 capas ocultas.
- 1000 neuronas por capa.
- Es necesario estimar miles de parámetros.
- Se necesitan 1000 horas de grabación.
- Un super computador se demora dos semanas estimando.

# La revolución empresarial: *deep learning*

- [www.clarifai.com](http://www.clarifai.com)

# Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Cuál es la revolución científica?
- 6 La revolución empresarial
- 7 Quantil**



# Quantil

quantil | matemáticas aplicadas