Curs superior universitari

CIÈNCIA OBERTA: promoció, suport i avaluació

UNIVERSITAT DE BARCELONA

Unitat 2 – Integritat, transparència i replicabilitat de la ciència

Sessió del dia 15 de Novembre de 2023

Replicabilitat i reproductibilitat de la recerca

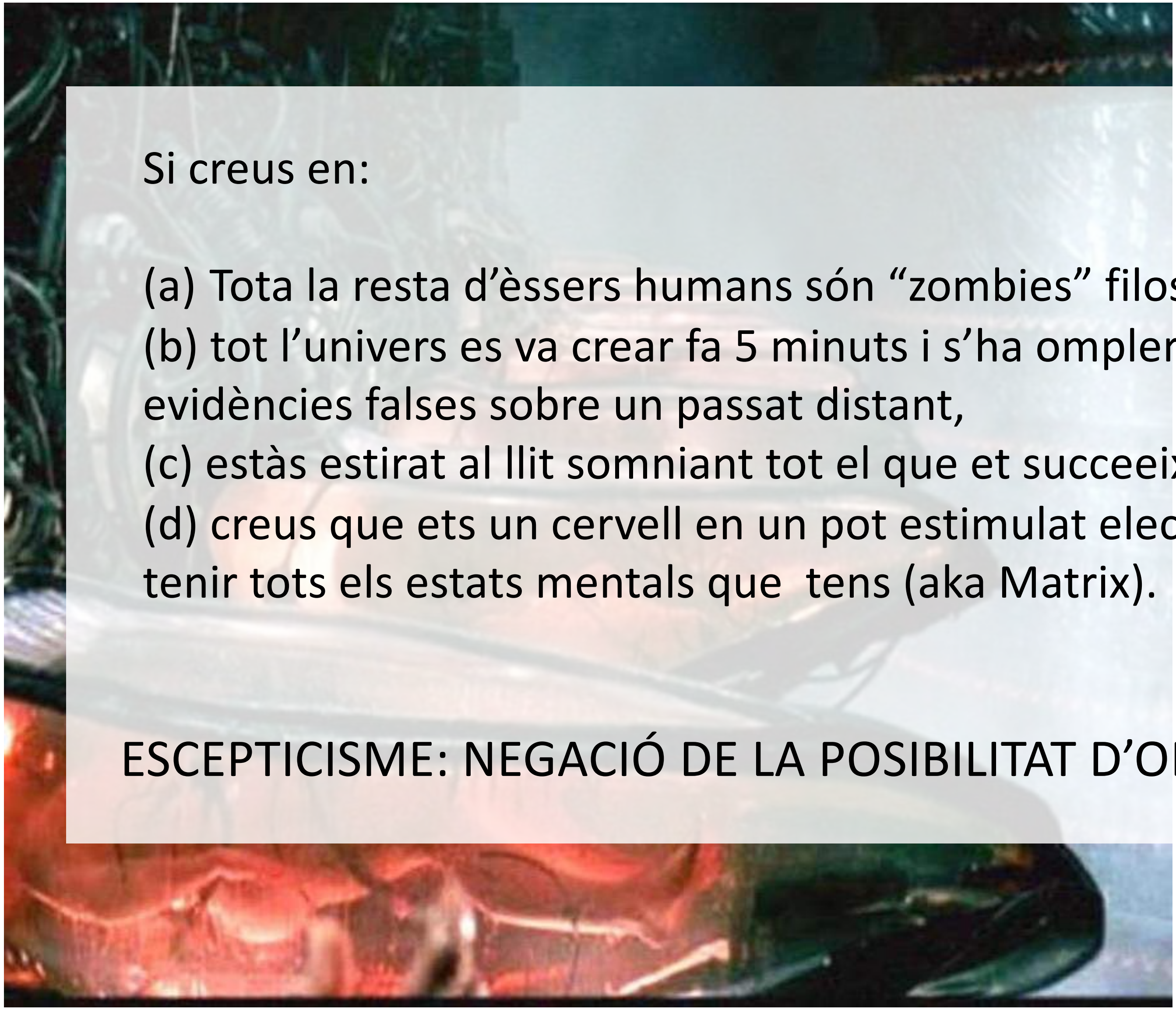
Oriol Pujol Vila
Universitat de Barcelona

Tria una llicència CC. Et recomanem aquestes (per a facilitar la reutilització):

CC BY

CC BY SA

DISCLAIMER



Si creus en:

(a) Tota la resta d’èssers humans són “zombies” filosòfics
(b) tot l’univers es va crear fa 5 minuts i s’ha omplert de memories i evidències falses sobre un passat distant,
(c) estàs estirat al llit somniant tot el que et succeeix ara mateix, o
(d) creus que ets un cervell en un pot estimulat electroquímicament per a tenir tots els estats mentals que tens (aka Matrix).

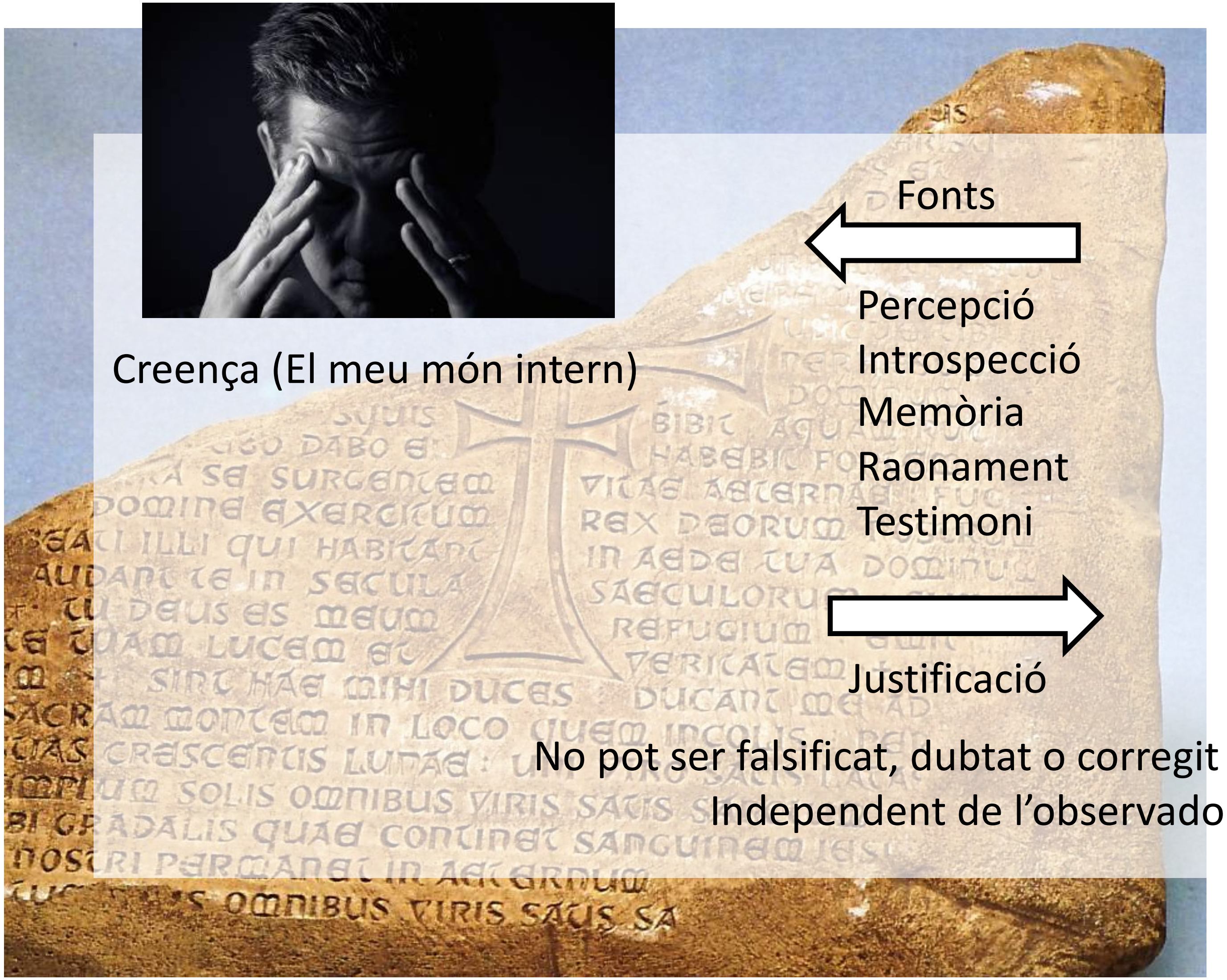
ESCEPTICISME: NEGACIÓ DE LA POSIBILITAT D’OBTENIR CONEIXEMENT

PER QUÈ CREIEU QUE ÉS IMPORTANT LA REPRODUCIBILITAT DE LA RECERCA?

LA RECERCA I LA CERCA DEL CONEIXEMENT

EPISTEMOLOGIA

Coneixement: Creença Veritable Justicada (Edmund Gettier)



Creença (El meu món intern)

Fonts

Percepció
Introspecció
Memòria
Raonament
Testimoni

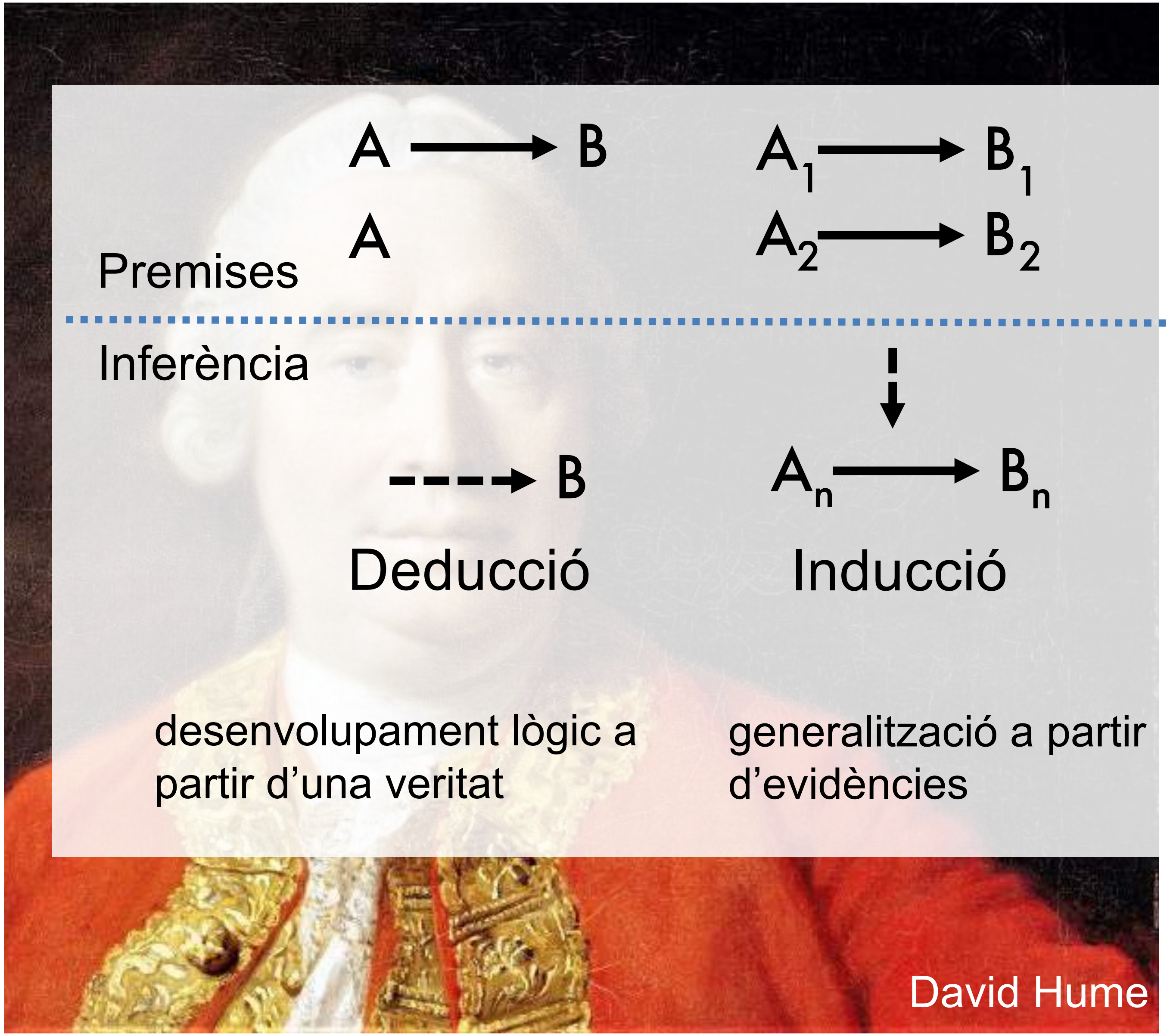
Justificació

Veritable (El Món)

No pot ser falsificat, dubtat o corregit pels altres
Independent de l'observador

EPISTEMOLOGIA

Inferència com a justificació



Premises	$A \longrightarrow B$ A	$A_1 \longrightarrow B_1$ $A_2 \longrightarrow B_2$	$A \longrightarrow B$ B
Inferència	$\dashrightarrow B$	$\downarrow$ $A_n \longrightarrow B_n$	$A \dashleftarrow$
	Deducció	Inducció	Abducció
	desenvolupament lògic a partir d'una veritat	generalització a partir d'evidències	inferencia a la millor explicació

David Hume

RECERCA

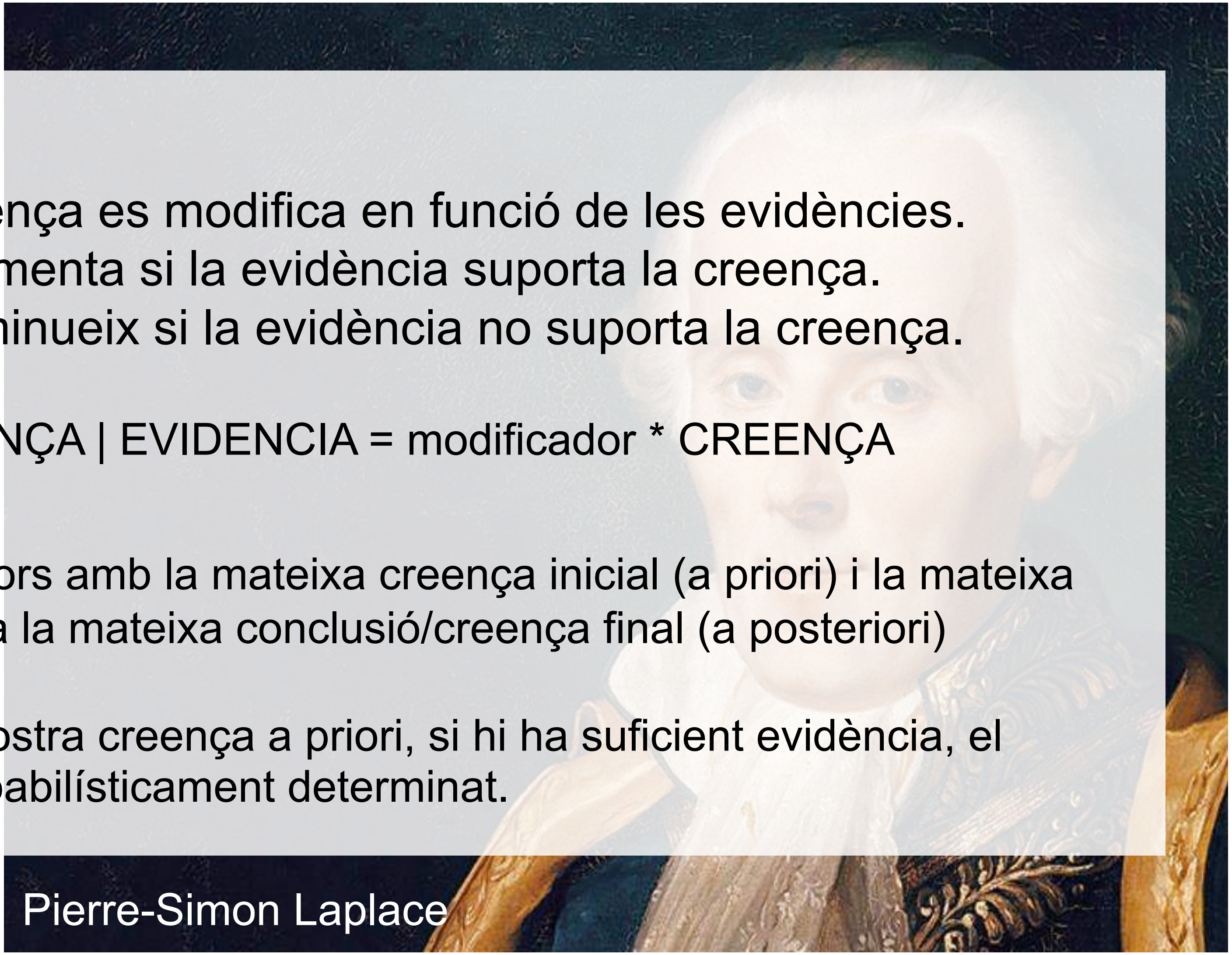
La creació de coneixement es planteja en tres possibilitats

Teòric

Experimental

Computacional

INDUCCIÓ I L’INFERÈNCIA BAYESIANA



1. La meva creença es modifica en funció de les evidències.
2. Creença augmenta si la evidència suporta la creença.
3. Creença disminueix si la evidència no suporta la creença.

CREENÇA | EVIDENCIA = modificador * CREENÇA

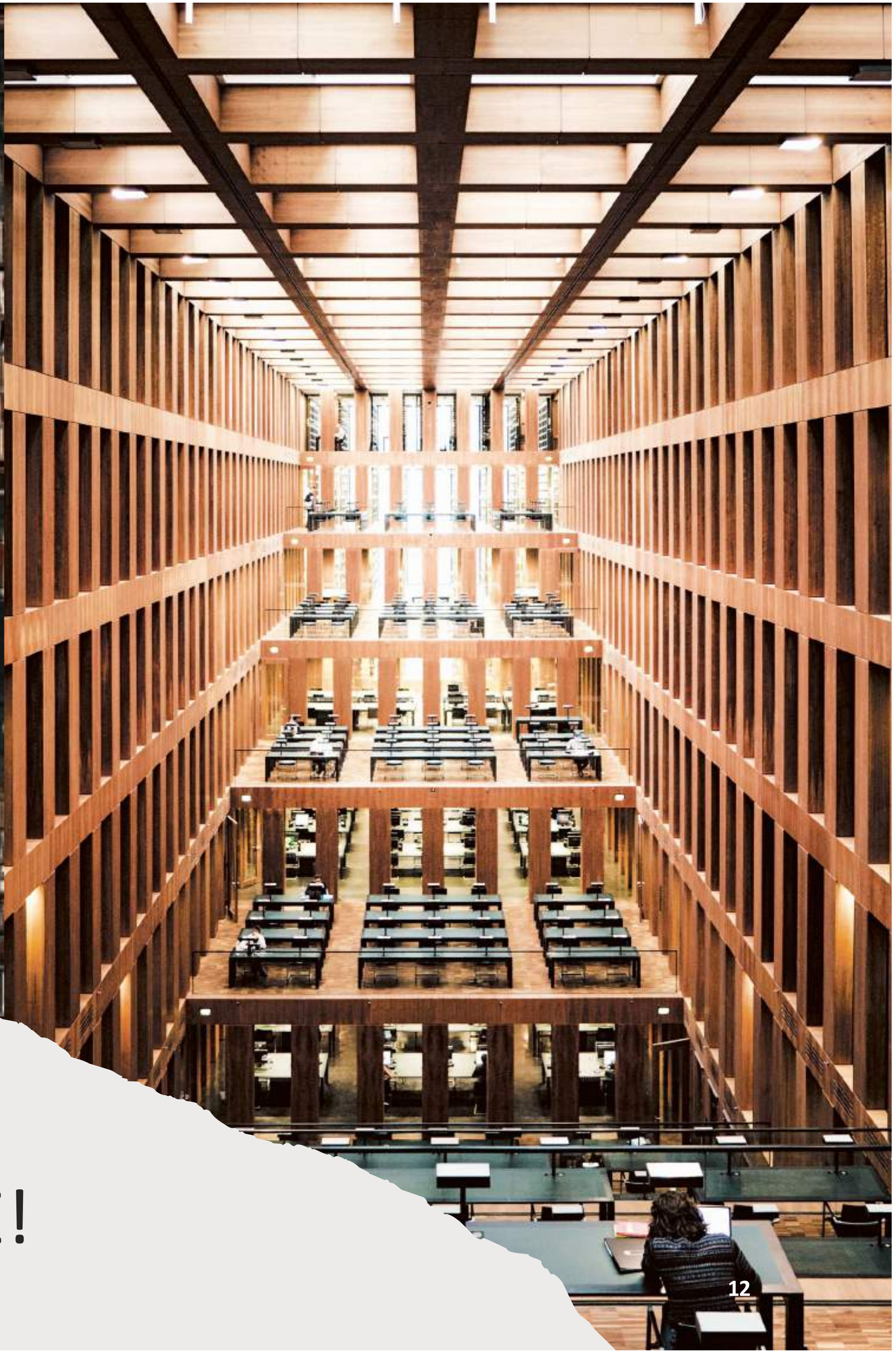
Donat dos observadors amb la mateixa creença inicial (a priori) i la mateixa evidència arribaran a la mateixa conclusió/creença final (a posteriori)

Independent de la nostra creença a priori, si hi ha suficient evidència, el fenomen queda probabilísticament determinat.

Pierre-Simon Laplace

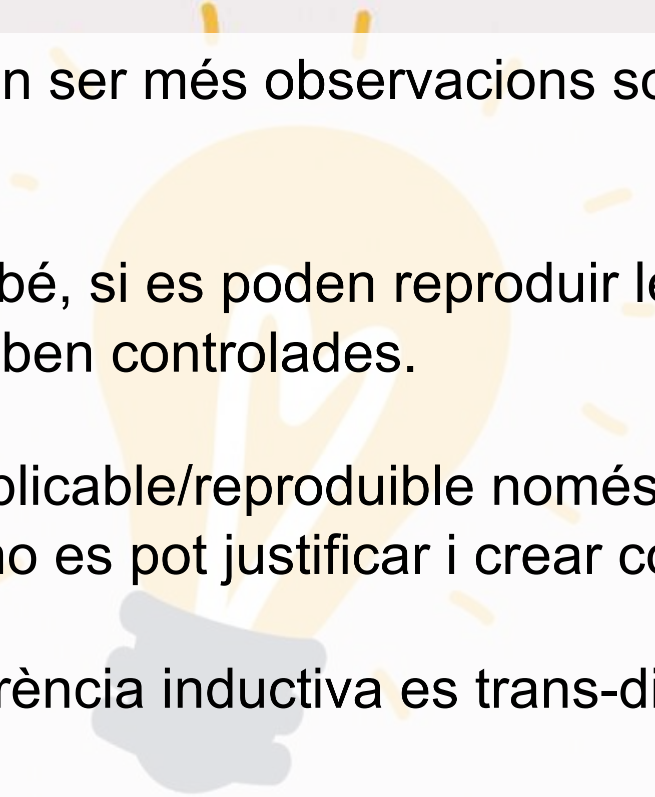
UNA DE ROMANS

https://github.com/oriolpujol/slides/blob/master/Reproducibility_Bayes.ipynb



SI ÉS CONEIXEMENT,
HA DE SER REPRODUIBLE!

IDEES IMPORTANT



1. Calen evidències per crear coneixement.

2. Les evidències poden ser més observacions sota la mateixa hipòtesi

3. Són evidències, també, si es poden reproduir les conclusions amb noves observacions ben controlades.

4. Si quelcom no es replicable/reproduible només hi ha una evidencia i per tant no es pot justificar i crear coneixement.

5. El mecanisme d’inferència inductiva es trans-disciplinar.

6. Si diversos grups de recerca treballen sota la mateixa pregunta es poden juntar evidències i donar suport a una hipòtesi.

7. Permet la reinterpretació de les evidències i la potencial correcció d’errors.

EL “REPLIGATE”

2014 a Psicologia Social es prohibeix el p-value

Més del 70% de recercaires han fallat en reproduir els experiments d’altres.

Més del 50% han fallat a reproduir els seus propis experiments.

STATISTICAL ERRORS

P values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume.

BY REGINA NUZZO

CONSEQÜÈNCIES SOCIALS

Pèrdua de temps,
Pèrdua de diners,
Pèrdua de credibilitat

A QUÈ ES DEU?

QUÈ NECESSITEM PER FER RECERCA REPRODUIBLE?

EL PASTÍS REPRODUIBLE

16:14

TASTEAGLORIAS
Publicaciones

tasteaglorias
Glòcakes

10 Me gusta

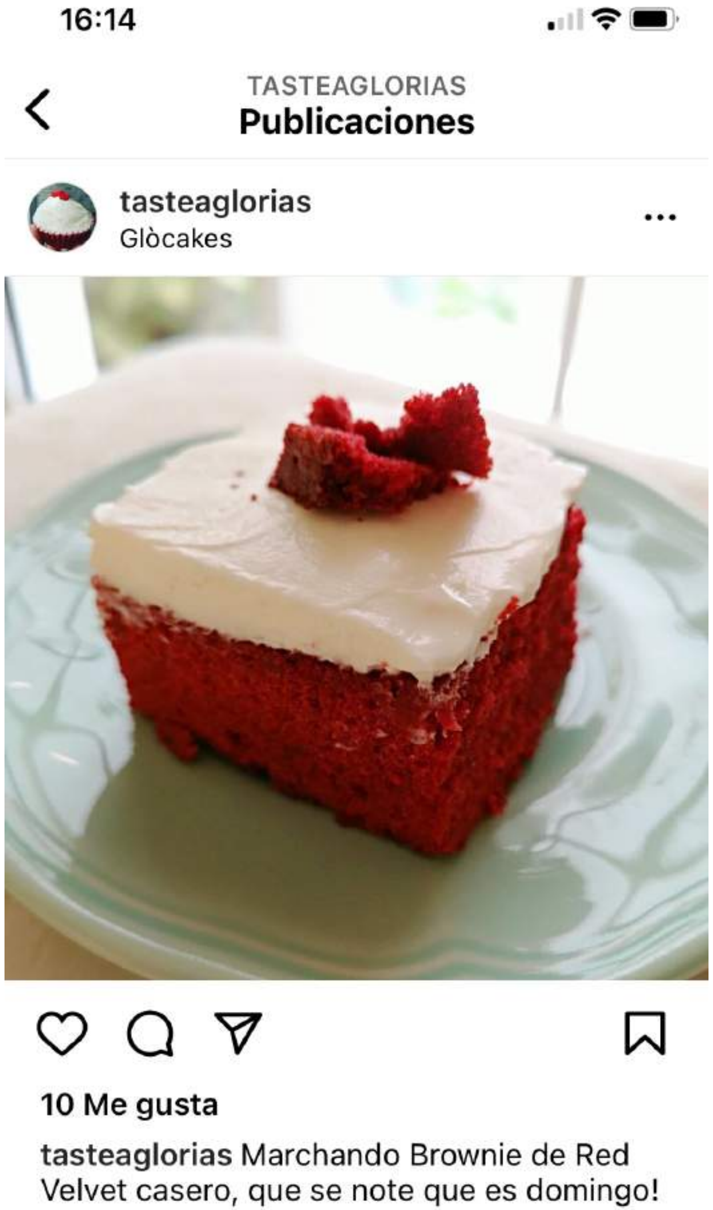
tasteaglorias Marchando Brownie de Red Velvet casero, que se note que es domingo!

Replicable – es pot repetir?
(es poden tornar a recollir els ingredients i replicar les condicions per realitzar el pastís? Es poden obtenir resultats consistents per respondre la mateixa qüestió?)

Reproduible – parla de la validació del procés. S’arriben a les conclusions similars amb les mateixes dades i mètode?

Què necessitem per fer el pastís reproduible?

EL PASTÍS REPRODUIBLE



10 Me gusta

tasteaglorias Marchando Brownie de Red Velvet casero, que se note que es domingo!

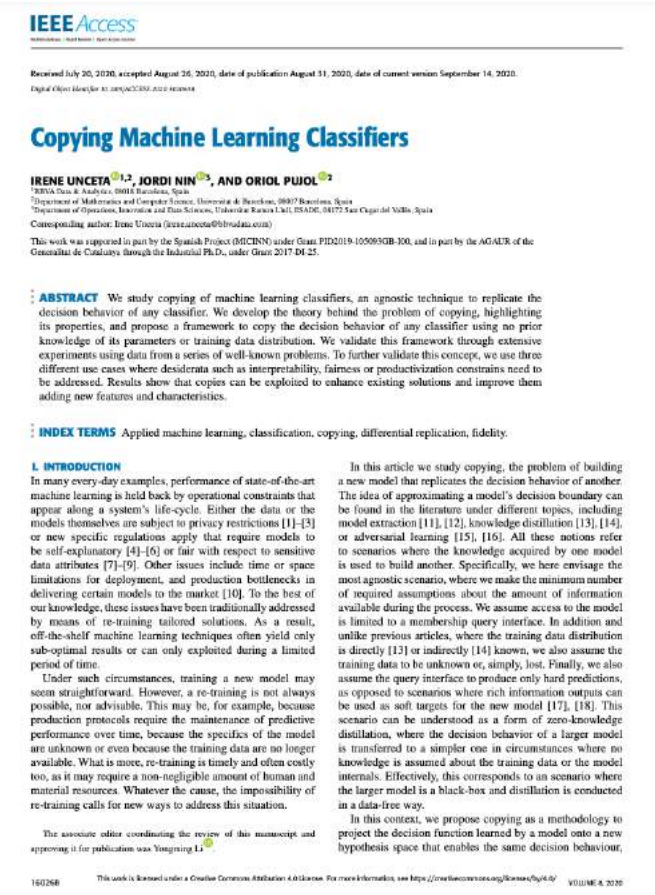
Ingredients: Els podem trovar?

Quantitats: Quina quantitat s'ha fet servir?

Procés: Tenim la recepta?

Entorn: Tenim les condicions per reproduir la cocció?

RECERCA REPRODUIBLE



IEEE Access

Received July 20, 2020, accepted August 26, 2020, date of publication August 31, 2020, date of current version September 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.[illegible]

Copying Machine Learning Classifiers

IRENE UNCETA[1,2], JORDI NIN[3], AND ORIOL PUJOL[2]

[1] BBVA Data & Analytics, 08018 Barcelona, Spain
[2] Department of Mathematics and Computer Science, Universitat de Barcelona, 08007 Barcelona, Spain
[3] Department of Operations, Innovation and Data Sciences, Universitat Ramon Llull, ESADE, 08172 Sant Cugat del Vallès, Spain

Corresponding author: Irene Unceta (irene.unceta@bbvadata.com)

This work was supported in part by the Spanish Project (MICINN) under Grant PID2019-105093GB-I00, and in part by the AGAUR of the Generalitat de Catalunya through the Industrial Ph.D., under Grant 2017-DI-25.

ABSTRACT We study copying of machine learning classifiers, an agnostic technique to replicate the decision behavior of any classifier. We develop the theory behind the problem of copying, highlighting its properties, and propose a framework to copy the decision behavior of any classifier using no prior knowledge of its parameters or training data distribution. We validate this framework through extensive experiments using data from a series of well-known problems. To further validate this concept, we use three different use cases where desiderata such as interpretability, fairness or productivization constrains need to be addressed. Results show that copies can be exploited to enhance existing solutions and improve them adding new features and characteristics.

INDEX TERMS Applied machine learning, classification, copying, differential replication, fidelity.

I. INTRODUCTION

In many every-day examples, performance of state-of-the-art machine learning is held back by operational constraints that appear along a system's life-cycle. Either the data or the models themselves are subject to privacy restrictions [1]–[3] or new specific regulations apply that require models to be self-explanatory [4]–[6] or fair with respect to sensitive data attributes [7]–[9]. Other issues include time or space limitations for deployment, and production bottlenecks in delivering certain models to the market [10]. To the best of our knowledge, these issues have been traditionally addressed by means of re-training tailored solutions. As a result, off-the-shelf machine learning techniques often yield only sub-optimal results or can only exploited during a limited period of time.

Under such circumstances, training a new model may seem straightforward. However, a re-training is not always possible, nor advisable. This may be, for example, because production protocols require the maintenance of predictive performance over time, because the specifics of the model are unknown or even because the training data are no longer available. What is more, re-training is timely and often costly too, as it may require a non-negligible amount of human and material resources. Whatever the cause, the impossibility of re-training calls for new ways to address this situation.

In this article we study copying, the problem of building a new model that replicates the decision behavior of another. The idea of approximating a model's decision boundary can be found in the literature under different topics, including model extraction [11], [12], knowledge distillation [13], [14], or adversarial learning [15], [16]. All these notions refer to scenarios where the knowledge acquired by one model is used to build another. Specifically, we here envisage the most agnostic scenario, where we make the minimum number of required assumptions about the amount of information available during the process. We assume access to the model is limited to a membership query interface. In addition and unlike previous articles, where the training data distribution is directly [13] or indirectly [14] known, we also assume the training data to be unknown or, simply, lost. Finally, we also assume the query interface to produce only hard predictions, as opposed to scenarios where rich information outputs can be used as soft targets for the new model [17], [18]. This scenario can be understood as a form of zero-knowledge distillation, where the decision behavior of a larger model is transferred to a simpler one in circumstances where no knowledge is assumed about the training data or the model internals. Effectively, this corresponds to an scenario where the larger model is a black-box and distillation is conducted in a data-free way.

In this context, we propose copying as a methodology to project the decision function learned by a model onto a new hypothesis space that enables the same decision behaviour,

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Ingredients: Dades (detalls d'aquisició, disseny experimental, metadades, etc.)

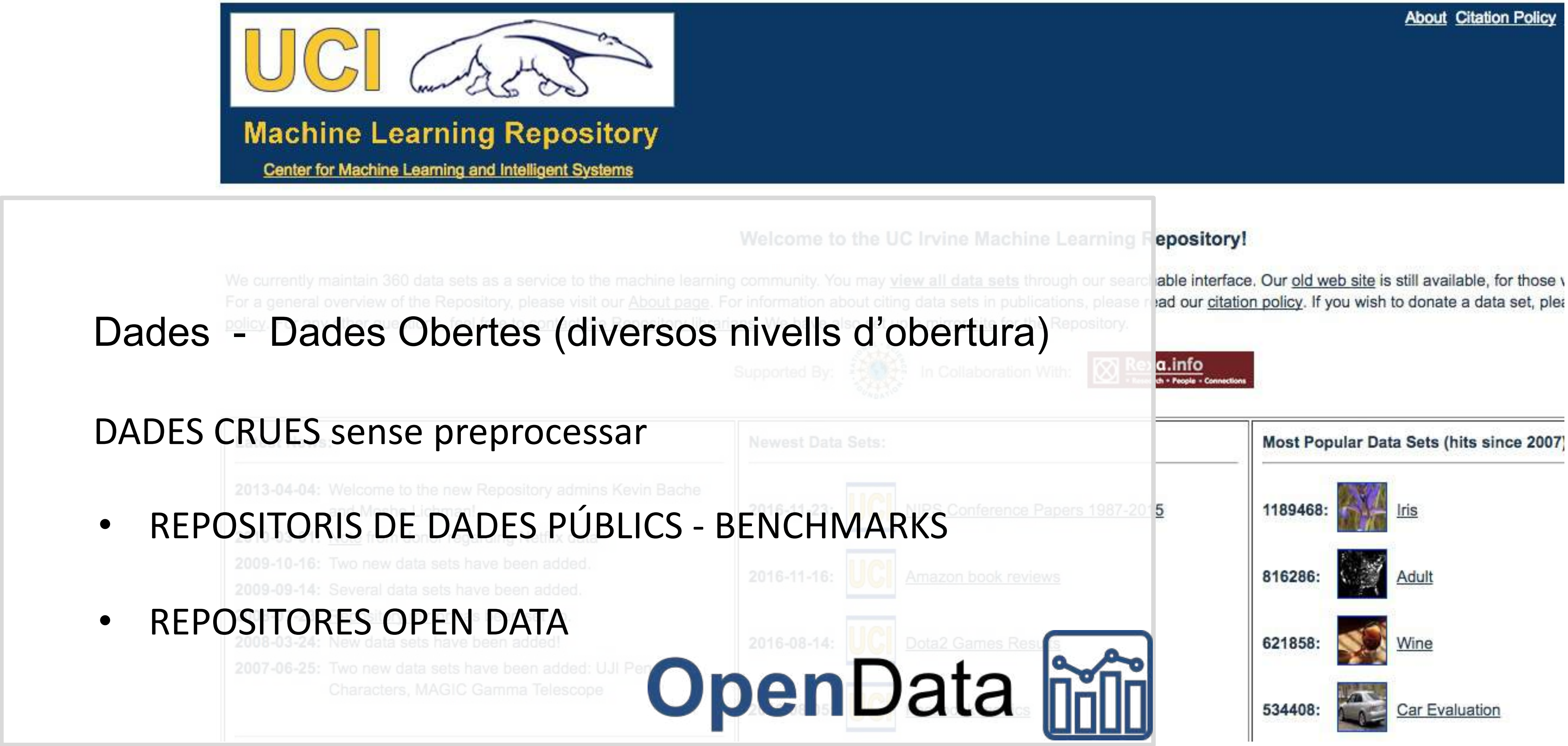
Recepta: Decisions fetes (netejant, processant, recodificant, etc)

Resultats (clars i tant objectius com sigui posible)

Entorn: Eines i entorns (materials, programari, paquets, etc.)

ACCÉS a tots aquests!!!!!

INGREDIENTS



About Citation Policy

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 360 data sets as a service to the machine learning community. You may view all data sets through our searchable interface. Our old web site is still available, for those
For a general overview of the Repository, please visit our About page. For information about citing data sets in publications, please read our citation policy. If you wish to donate a data set, plea

Dades - Dades Obertes (diversos nivells d'obertura)

DADES CRUES sense preprocessar

- REPOSITORIS DE DADES PÚBLICS - BENCHMARKS
- REPOSITORES OPEN DATA

OpenData

Newest Data Sets:

2016-11-20: NIPS Conference Papers 1987-2015

2016-11-16: Amazon book reviews

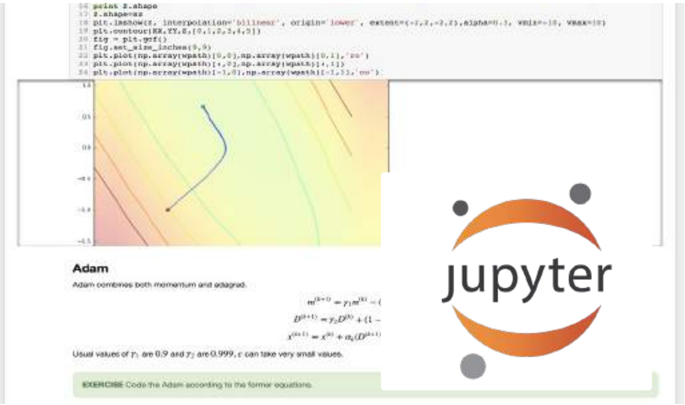
2016-08-14: Dota2 Games Results

Most Popular Data Sets (hits since 2007)

1189468:	Iris
816286:	Adult
621858:	Wine
534408:	Car Evaluation

PROCÉS I RESULTATS

Decisions fetes (publicacions, codi, github, jupyter notebooks)
Resultats (publicacions, metadades)



```
print Z.shape
Z.shape=ss
plt.imshow(Z, interpolation='bilinear', origin='lower', extent=(-2,2,-2,2),alpha=0.3, vmin=-10, vmax=10)
plt.contour(XX,YY,Z,[0,1,2,3,4,5])
fig = plt.gcf()
fig.set_size_inches(9,9)
plt.plot(np.array(wpath)[0,0],np.array(wpath)[0,1],'ro')
plt.plot(np.array(wpath)[:,0],np.array(wpath)[:,1])
plt.plot(np.array(wpath)[-1,0],np.array(wpath)[-1,1],'ro')
```

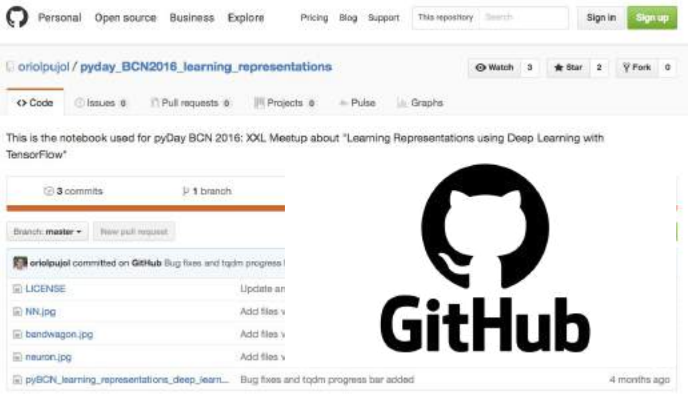
Adam

Adam combines both momentum and adagrad.

$$m^{(k+1)} = \gamma_1 m^{(k)} - ($$
$$D^{(k+1)} = \gamma_2 D^{(k)} + (1 -$$
$$x^{(k+1)} = x^{(k)} + \alpha_k (D^{(k+1)}$$

Usual values of γ_1 are 0.9 and γ_2 are 0.999, ϵ can take very small values.

EXERCISE Code the Adam according to the former equations.



oriolpujol / pyday_BCN2016_learning_representations

This is the notebook used for pyDay BCN 2016: XXL Meetup about "Learning Representations using Deep Learning with TensorFlow"

3 commits　1 branch

Branch: master ▾　New pull request

oriolpujol committed on GitHub Bug fixes and tqdm progress

LICENSE	Update an	
NN.jpg	Add files v	
bandwagon.jpg	Add files v	
neuron.jpg	Add files v	
pyBCN_learning_representations_deep_learn...	Bug fixes and tqdm progress bar added	4 months ago

arXiv is a free distribution service and an open-access archive for 2,022,298 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. Materials on this site are not peer-reviewed by arXiv.

OPEN ACCESS

ENTORN

Eines i entorns (entorn de treball – Docker)

Contenidor: Encapsulament del programari, totes les llibreries i dependències en una imatge que funciona sobre un Sistema Operatiu virtualitzat.

- Encapsulament de l'experiment (freeze)
- Transportabilitat (mínim espai)
- Interoperabilitat i reusabilitat (independent del sistema operatiu hoste)

CONCLUSIONS

1. Recerca és més que una publicació
2. Al cor de la recerca es troba la reproducibilitat… i això sempre involucra dades.
3. Però dades no són suficients… el codi, metadades, i descripció del procés.
4. Requeriment: **Accessibilitat**
5. Alfabetització sobre les dades i el seu tractament (programació?) són necessaries.

GRÀCIES!