Data science

from data driven to deep learning

吳沛燊 Pei-shen Wu, MD

2017-12-11 @ MLDM TW R user group

吳沛燊 peishen.wu

台大醫院復健科總醫師
均一教育平台 Data scientist
深義分析公司 Data scientist

故事 1

當時的大問題：

回答速度趕不上

問題的產生

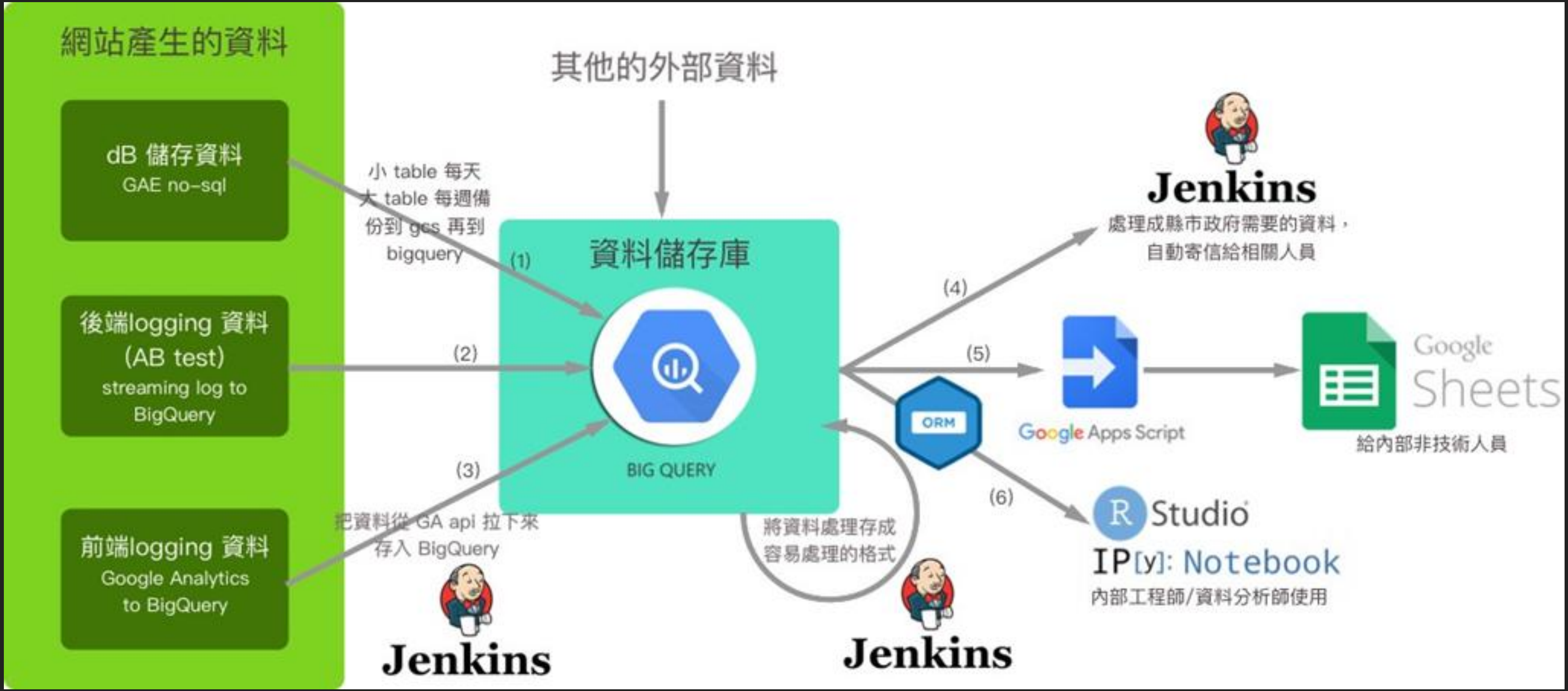
跟複雜度

回答單一問題所需的SQL code

```
SELECT keyID,userID,dot_email,gender,points,user_nickname,username,userEmail,userSendableEmail,userRole,userCity,userSchool,userGrade,joinedTime,userBirthdate,match_score FROM (SELECT pre.keyID AS
keyID,pre.userID AS userID,pre.dot_email AS dot_email,pre.gender AS gender,pre.points AS points,pre.user_nickname AS user_nickname,pre.username AS username,pre.userEmail AS userEmail,pre.userSendableEmail AS
userSendableEmail,pre.userRole AS userRole,pre.userCity AS userCity,pre.userSchool AS userSchool,pre.userGrade AS userGrade,pre.joinedTime AS joinedTime,pre.userBirthdate AS userBirthdate,post.match_score AS
match_score FROM (SELECT * FROM (SELECT pre.keyID AS keyID,pre.userID AS userID,pre.dot_email AS dot_email,pre.gender AS gender,pre.points AS points,pre.user_nickname AS user_nickname,pre.username AS
username,post.userEmail AS userEmail,post.userSendableEmail AS userSendableEmail,post.userRole AS userRole,post.userCity AS userCity,post.userSchool AS userSchool,post.userGrade AS userGrade,post.joinedTime AS
joinedTime,post.userBirthdate AS userBirthdate FROM (SELECT * FROM (SELECT keyID, userID, dot_email, gender, points, user_nickname, username FROM (SELECT __key__.name AS keyID, user_id
AS userID, user_email AS underline_email, current_user.email AS cu_email, user.email AS dot_email, gender, points, user_nickname, username FROM
junyi_20161212.UserData_20161212)) AS pre INNER JOIN EACH (SELECT userEmail, userSendableEmail, userid AS userID, userRole, userCity, userSchool, userGrade,
joinedTime, userBirthdate FROM FinalTable.UserFinalTmpInfo) AS post ON pre.userID = post.userID)) AS pre INNER JOIN EACH (SELECT keyID,match_score FROM (SELECT keyID, match_score FROM
(SELECT keyID,prob1*prob2 AS match_score FROM (SELECT pre.keyID AS keyID,pre.prob1 AS prob1,post.prob2 AS prob2 FROM (SELECT * FROM (SELECT keyID, prob1 FROM (SELECT keyID, match_score AS prob1 FROM
(SELECT keyID, match_score FROM (SELECT keyID, match_score_A*cdf AS match_score FROM (SELECT keyID, MAX(output) AS cdf, match_score_A FROM (SELECT keyID, IF(metric >= input,1,0) AS compare, output,
match_score_A FROM (SELECT pre.keyID AS keyID,pre.comparekey AS comparekey,pre.metric AS metric,pre.match_score_A AS match_score_A,post.input AS input,post.output AS output FROM (SELECT * FROM (SELECT
keyID, 1 AS comparekey, metric, match_score AS match_score_A FROM (SELECT pre.keyID AS keyID,pre.match_score AS match_score,post.dot_email AS dot_email,post.metric AS metric FROM (SELECT * FROM (SELECT
teacher_keyID AS keyID, MAX(match_score) AS match_score FROM(SELECT pre.student_keyID AS student_keyID,pre.classID AS classID,pre.student_underline_email AS student_underline_email,pre.teacher_keyID AS
teacher_keyID,post.match_score AS match_score FROM (SELECT * FROM (SELECT pre.student_keyID AS student_keyID,pre.classID AS classID,pre.student_underline_email AS student_underline_email,post.teacher_keyID AS
teacher_keyID FROM (SELECT * FROM (SELECT student_keyID, classID, student_underline_email FROM FLATTEN((SELECT __key__.name AS student_keyID, user_email AS student_underline_email,
student_lists.path AS classID FROM junyi_20161212.UserData_20161212), classID)) AS pre INNER JOIN EACH (SELECT teacher_keyID, classID FROM(SELECT coaches.name AS teacher_keyID,
__key__.path AS classID, code AS classcode, name AS classname FROM junyi_20161212.StudentList_20161212)) AS post ON pre.classID = post.classID)) AS pre INNER JOIN EACH (SELECT classID,
MAX(match_score) AS match_score FROM (SELECT classID,match_score FROM (SELECT classID, MAX(match_score) AS match_score FROM (SELECT pre.student_keyID AS student_keyID,pre.classID AS
classID,pre.student_underline_email AS student_underline_email,pre.teacher_keyID AS teacher_keyID,post.match_score AS match_score FROM (SELECT * FROM (SELECT pre.student_keyID AS student_keyID,pre.classID AS
classID,pre.student_underline_email AS student_underline_email,post.teacher_keyID AS teacher_keyID FROM (SELECT * FROM (SELECT student_keyID, classID, student_underline_email FROM FLATTEN((SELECT
__key__.name AS student_keyID, user_email AS student_underline_email, student_lists.path AS classID FROM junyi_20161212.UserData_20161212), classID)) AS pre INNER JOIN EACH (SELECT
teacher_keyID, classID FROM(SELECT coaches.name AS teacher_keyID, __key__.path AS classID, code AS classcode, name AS classname FROM
junyi_20161212.StudentList_20161212)) AS post ON pre.classID = post.classID)) AS pre INNER JOIN EACH (SELECT keyID AS student_keyID, match_score FROM (SELECT student_keyID AS keyID, MAX(match_score) AS
match_score FROM(SELECT student_keyID, teacher_keyID, 1 AS match_score FROM (SELECT pre.student_keyID AS student_keyID,pre.classID AS classID,pre.student_underline_email AS
student_[illegible]mail,post.teacher_keyID AS [illegible] FROM [illegible] (SELECT student_keyID, classID, student_underline_email FROM FLATTEN((SELECT __key__.name AS student_keyID, user_email AS
stud[illegible]_email, student_[illegible].path AS classID [illegible] junyi_20161212.UserData_20161212[illegible] classID)) AS [illegible] EACH (SELECT teacher_keyID, classID FROM(SELECT coaches.name
AS teacher_keyID, __key__.path AS classID, [illegible] AS classcode, name AS classname [illegible] junyi_[illegible].StudentList_20161212)) AS post ON pre.[illegible] = post.classID)) GROUP
BY keyID)) AS post ON pre.student_keyID = post.student_keyID)) GROUP BY [illegible] GROUP BY classID)) AS post ON pre.classID = post.classID)) GROUP BY keyID) AS pre INNER JOIN EACH (SELECT pre.dot_email AS
dot_email,pre.metric AS metric,post.keyID AS keyID FROM (SELECT * FROM [illegible] email, SUM(points_earned) [illegible] FROM(SELECT dot_email, points_earned, timestamp, time_consumed FROM
(SELECT dot_email, points_earned, timestamp, time_consumed [illegible]ROM (SELECT dot_email,points_earned,timestamp,time_consumed FROM (SELECT dot_email, points_earned, timestamp,
time_consumed FROM (SELECT user.email AS dot_email, points_earned, seconds_watched AS time_consumed, DATE_ADD(time_watched, 8, 'HOUR') AS timestamp FROM
junyi_20161212.VideoLog_20161212)),(SELECT dot_email, points_earned, timestamp, time_consumed FROM (SELECT user.email AS dot_email, points_earned, time_taken AS time_consumed,
DATE_ADD(time_done, 8, 'HOUR') AS timestamp FROM junyi_20161212.ProblemLog_20161212))) WHERE timestamp >= TIMESTAMP('2016-09-01') AND timestamp <= TIMESTAMP('2016-12-12'))
GROUP BY dot_email) AS pre INNER JOIN EACH (SELECT keyID, dot_email FROM (SELECT __key__.name AS keyID, user.email AS dot_email FROM junyi_20161212.UserData_20161212)) AS post ON
pre.dot_email = post.dot_email)) AS post ON pre.keyID = post.keyID))) AS pre INNER JOIN EACH (SELECT input, AVG(output) AS output, 1 AS comparekey FROM ( SELECT input, output/101 AS output
FROM ( SELECT input, ROW_NUMBER() OVER() AS output FROM ( FLATTEN(( SELECT QUANTILES(metric,101) AS input FROM (SELECT pre.keyID AS keyID,pre.match_score
AS match_score,post.dot_email AS dot_email,post.metric AS metric FROM (SELECT * FROM (SELECT student_keyID AS keyID, MAX(match_score) AS match_score FROM(SELECT student_keyID, teacher_keyID, 1 AS
match_score FROM (SELECT pre.student_keyID AS student_keyID,pre.classID AS classID,pre.student_underline_email AS student_underline_email,post.teacher_keyID AS teacher_keyID FROM (SELECT * FROM
(SELECT student_keyID, classID, student_underline_email FROM FLATTEN((SELECT __key__.name AS student_keyID, user_email AS student_underline_email, student_lists.path AS classID FROM
junyi_20161212.UserData_20161212), classID)) AS pre INNER JOIN EACH (SELECT teacher_keyID, classID FROM(SELECT coaches.name AS teacher_keyID, __key__.path AS classID, code AS
classcode, name AS classname FROM junyi_20161212.StudentList_20161212)) AS post ON pre.classID = post.classID))) GROUP BY keyID) AS pre INNER JOIN EACH (SELECT pre.dot_email AS dot_email,pre.metric
AS metric,post.keyID AS keyID FROM (SELECT * FROM (SELECT dot_email, SUM(points_earned) AS metric FROM(SELECT dot_email, points_earned, timestamp, time_consumed FROM (SELECT dot_email,
points_earned, timestamp, time_consumed FROM (SELECT dot_email,points_earned,timestamp,time_consumed FROM (SELECT dot_email, points_earned, timestamp, time_consumed FROM
(SELECT user.email AS dot_email, points_earned, seconds_watched AS time_consumed, DATE_ADD(time_watched, 8, 'HOUR') AS timestamp FROM
junyi_20161212.VideoLog_20161212)),(SELECT dot_email, points_earned, timestamp, time_consumed FROM (SELECT user.email AS dot_email, points_earned, time_taken AS time_consumed,
DATE_ADD(time_done, 8, 'HOUR') AS timestamp FROM junyi_20161212.ProblemLog_20161212))) WHERE timestamp >= TIMESTAMP('2016-09-01') AND timestamp <= TIMESTAMP('2016-12-12'))
GROUP BY dot_email) AS pre INNER JOIN EACH (SELECT keyID, dot_email FROM (SELECT __key__.name AS keyID, user.email AS dot_email FROM junyi_20161212.UserData_20161212)) AS post ON
pre.dot_email = post.dot_email)) AS post ON pre.keyID = post.keyID))), input)))) GROUP BY input) AS post ON pre.comparekey = post.comparekey))) WHERE compare == 1 GROUP BY keyID,
match_score_A)))) AS pre INNER JOIN EACH (SELECT keyID, prob2 FROM (SELECT keyID, match_score AS prob2 FROM (SELECT keyID, match_score FROM (SELECT keyID, timing_match_denumerator/
timing_match_numerator AS match_score FROM (SELECT keyID, SUM(match_condition) OVER (PARTITION BY keyID) AS timing_match_denumerator, COUNT(match_condition) OVER (PARTITION
BY keyID) AS timing_match_numerator, FROM (SELECT keyID, date, month, weekday, hour, minute, IF((hour == 9 OR hour == 10 OR hour == 11 OR hour [illegible]
match_condition FROM (SELECT keyID, date, month,
```

試圖以三週努力
滿足三秒的好奇

均一的 資料pipeline 架構

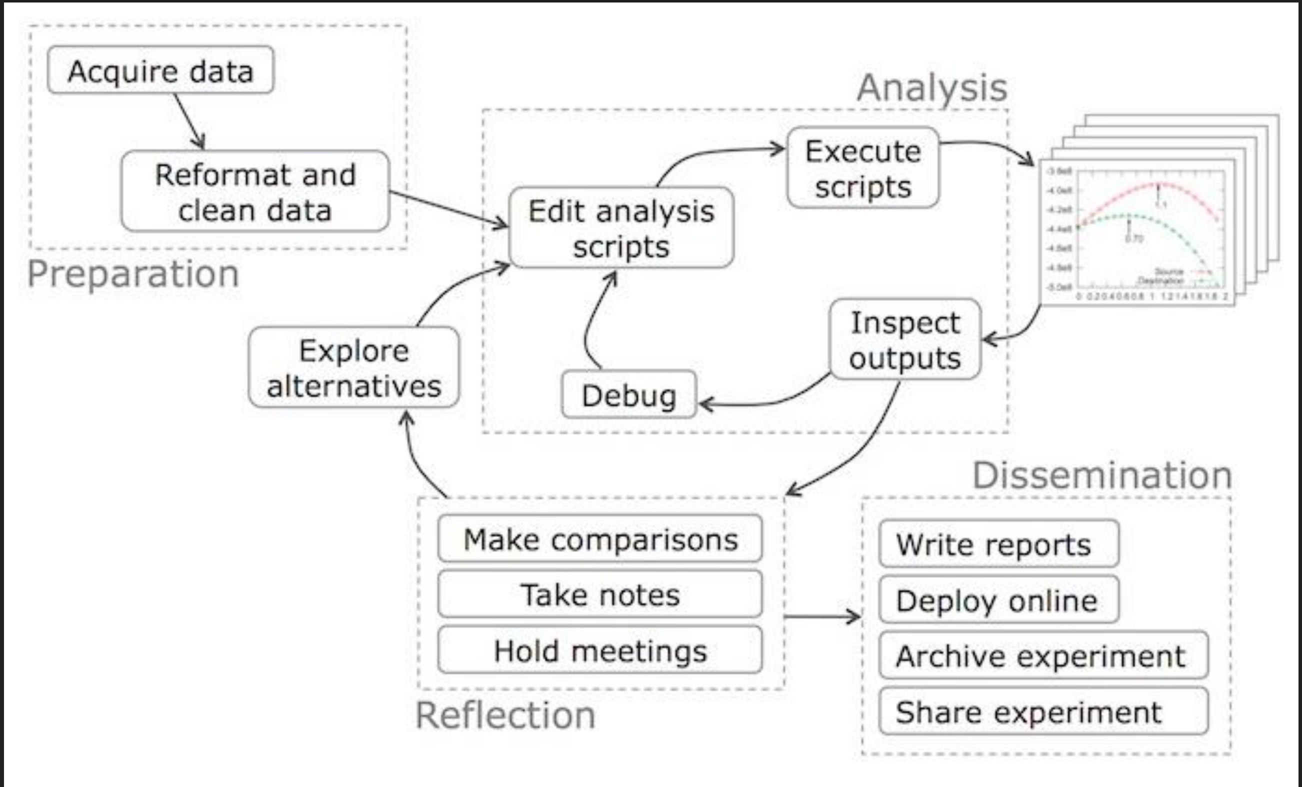


但 資料 =/= 知識

Data 架構 = 資料怎麼被收集、儲存、處理 跟散佈

Information 架構 = 把資料轉換成有用的資訊(知識)，所需要的過程跟practice

Data workflow 才是把資料轉成知識的架構

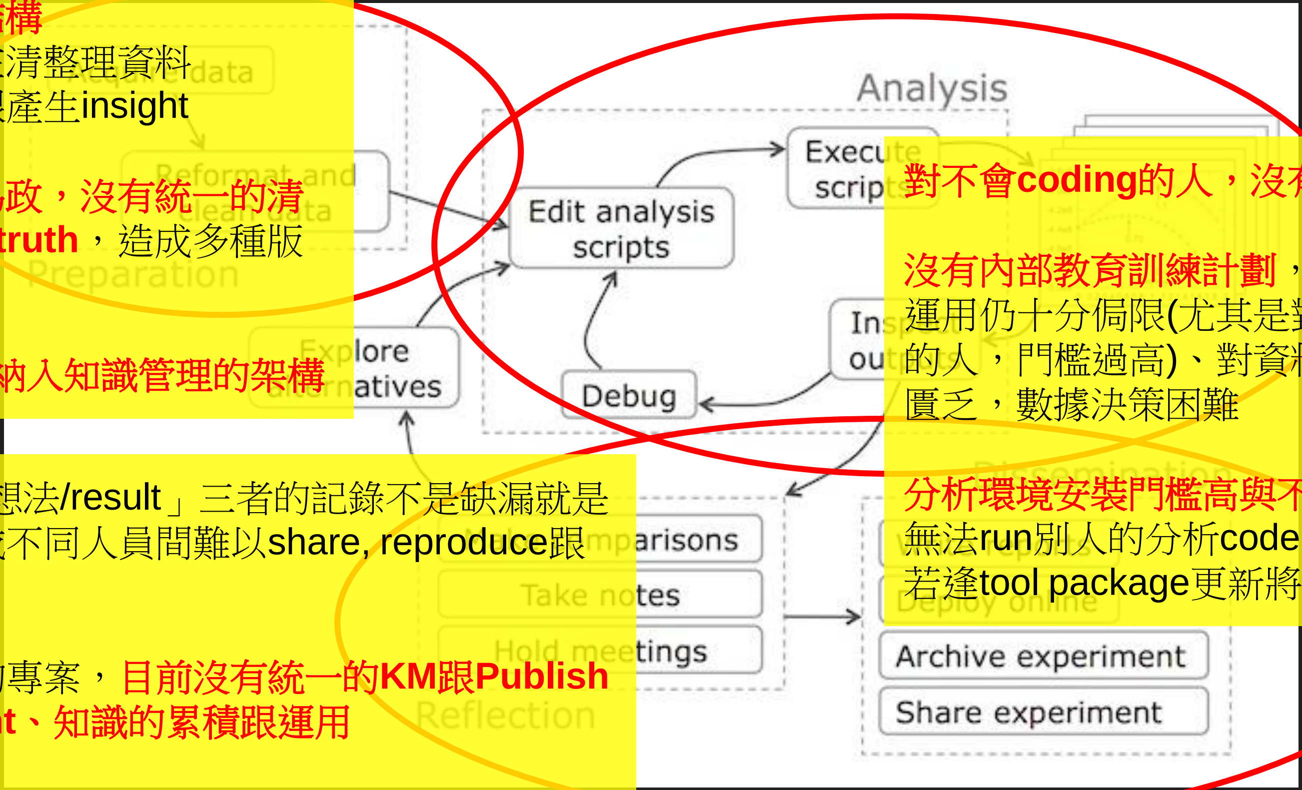


但問題叢生，處處是斷點，阻礙資料發展

不適任的資料庫結構
造成花太多時間在清整理資料
而不是進行分析跟產生insight

資料表創建各自為政，沒有統一的清理流程 與 **single truth**，造成多種版本的數據結果

A/B testing 尚未納入知識管理的架構



對不會**coding**的人，沒有合適工具

沒有內部教育訓練計劃，造成資料的運用仍十分侷限(尤其是對不會coding的人，門檻過高)、對資料表的觀念仍匱乏，數據決策困難

分析環境安裝門檻高與不一致，造成無法run別人的分析code
若逢tool package更新將會是場災難

「code/分析時的想法/result」三者的記錄不是缺漏就是四散在各處，造成不同人員間難以share, reproduce跟verify彼此的結果

此外，對已結案的專案，目前沒有統一的**KM**跟**Publish**機制，不利**Insight**、知識的累積跟運用

目前缺乏「持續有效的擴散數據分析成果」的管道

資料能否產生價值，還是要回歸到架構本身

一個系統的價值能否隨著時間增長的關鍵

= 人員從資料學習的容易度 + 將所得的insight自動化/系統化

(enable to learn from incoming data + rapidly operationalize those learnings)

Programs must be written for people to read, and only incidentally for machines to execute.

— *Hal Abelson*

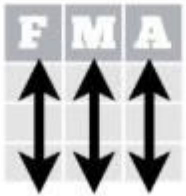
Tidy dataset are all alike; every messy dataset is messy in its own way

— Hadley Wickham

"Data Science with R"
https://www.youtube.com/watch?v=K-ss_ag2k9E

Tidy data definition

In a tidy data set:



Each **variable** is saved in its own **column**

&

Each **observation** is saved in its own **row**

Wickham, H. (2014). Tidy Data. Journal of Statistical Software

關於tidy data principle

1. 三個原則
    a. Each variable forms a column.
    b. Each observation forms a row.
    c. Each type of observational unit forms a table.

1. 這麼做有幾個目的:
    a. 減少data preprocessing/manipulation的次數，降低錯誤的同時增加coding的效率
    b. 直觀的data schema易於溝通
    c. 定義明確data schema，降低interpretation的難度，與保持data consistency
    d. 對於vector-based的tools (eg. R, python pandas) 易於操作

https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html

pipe operator %>% 增加code可讀性

Readable code chunks: The "pipe"-operator

- Readable code chunks can be considered as "grammar" of coding, which follows the similar intuitive logic from language or thinking

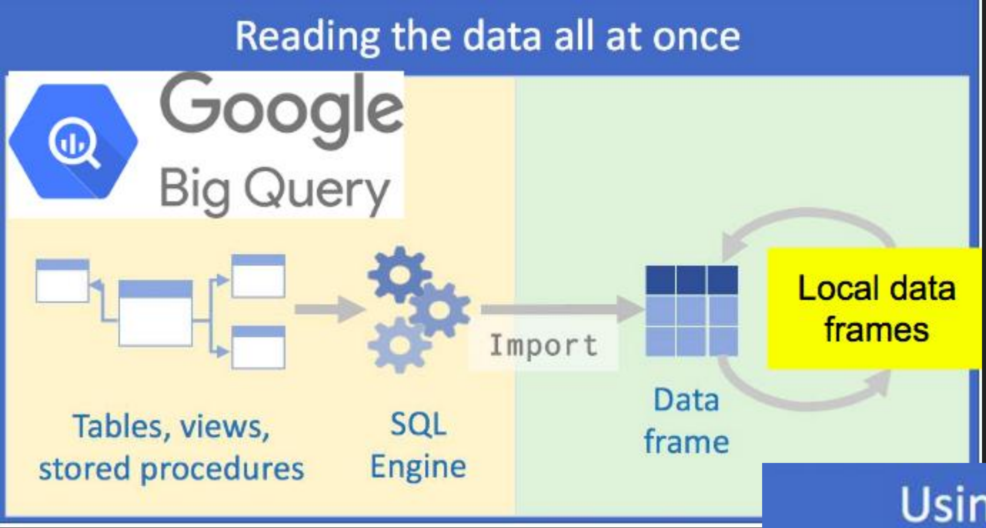
piped code chunk

```
data %>%
  do_first() %>%
  then_second() %>%
  and_then_third() %>%
  finally_last_step()
```

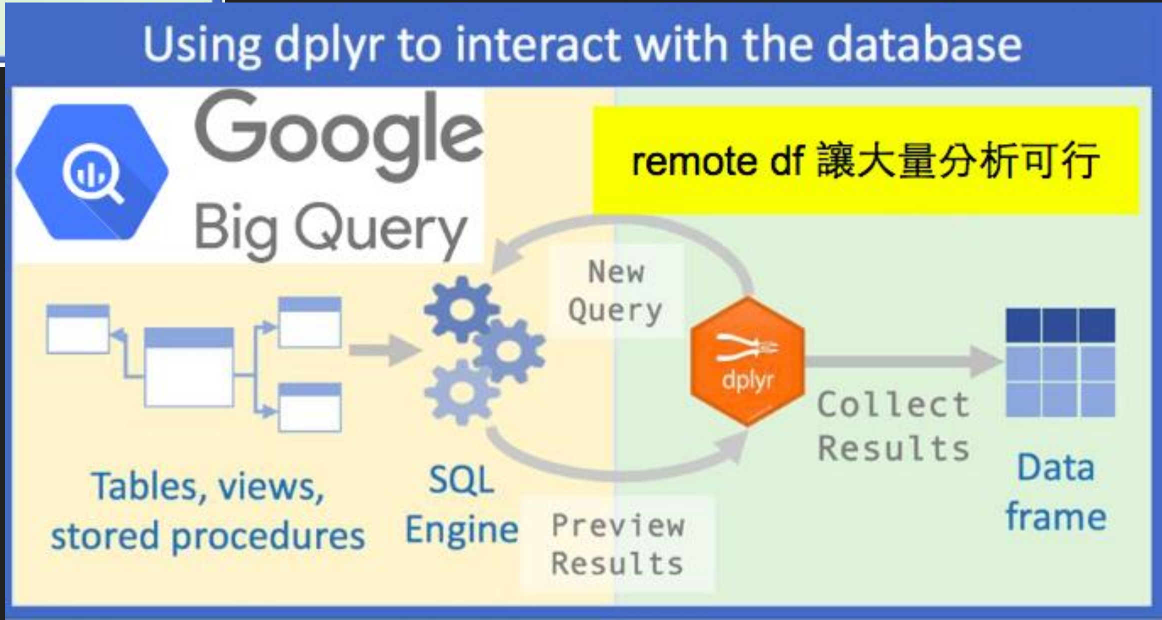
regular code chunk

```
finally_last_step(
  and_then_third(
    then_second(
      do_first(data)
    )
  )
)
```

https://strengejacke.files.wordpress.com/2017/02/data-transformation-and-exploration.pdf



資料處理發生在雲端
只把結果下載到近端電腦內



IP[y]: Notebook spectrogram Last Checkpoint: a few seconds ago (autosaved)

IPython (Python 3)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

Simple spectral analysis

An illustration of the Discrete Fourier Transform using windowing, to reveal the frequency content of a sound signal.

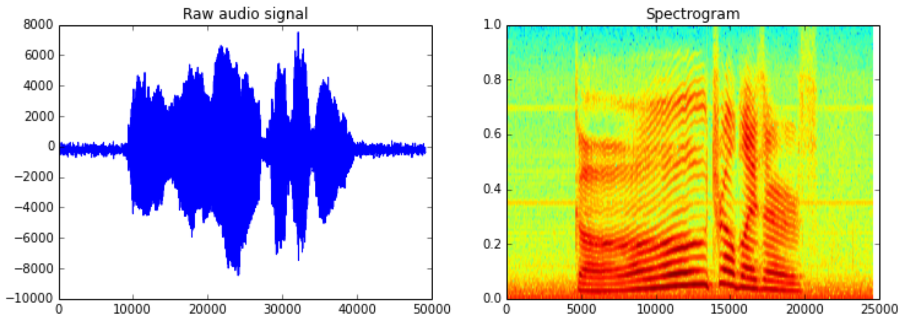
$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \qquad k = 0, \ldots, N-1$$

We begin by loading a datafile using SciPy's audio file support:

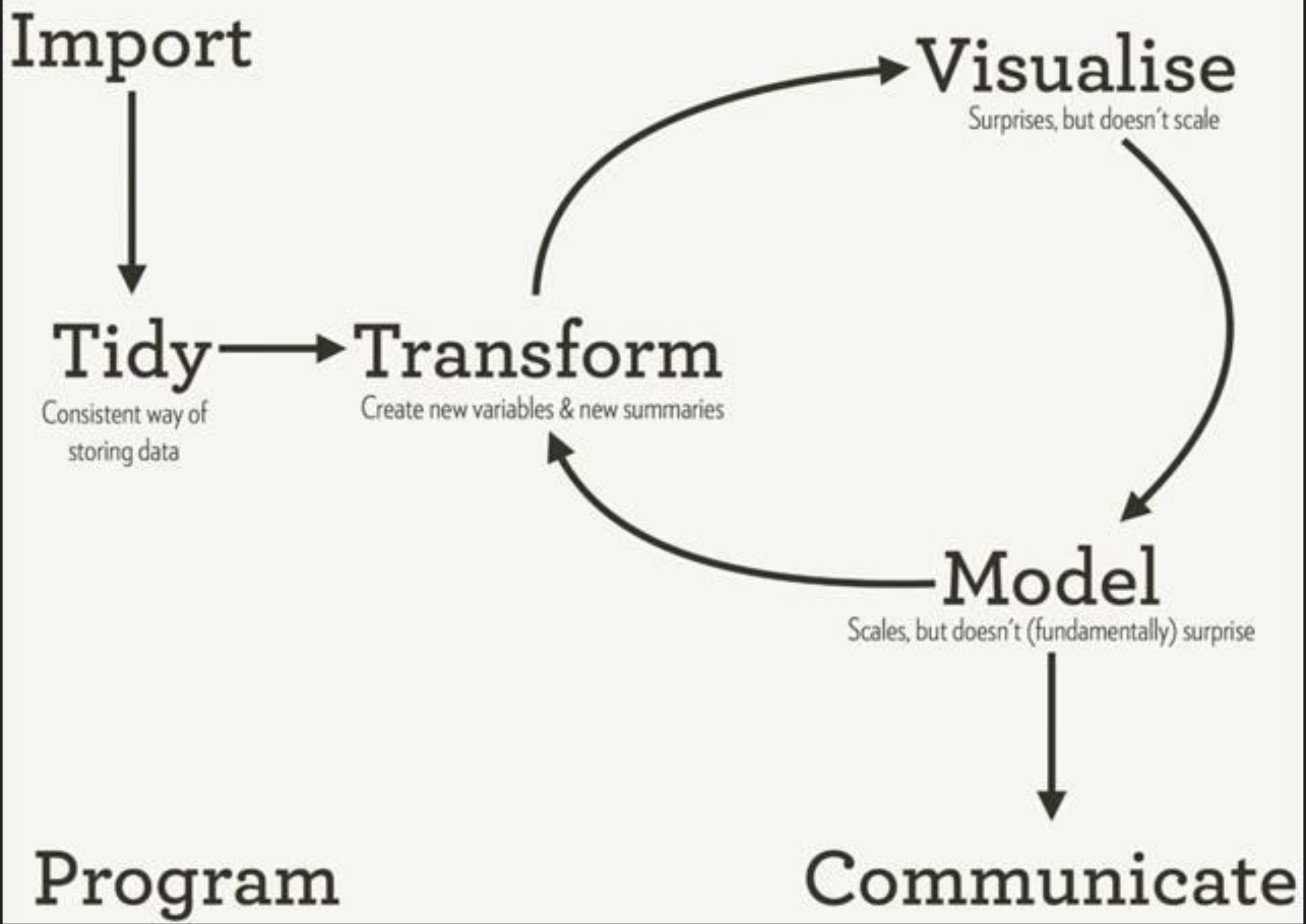
In [1]:
```
from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```

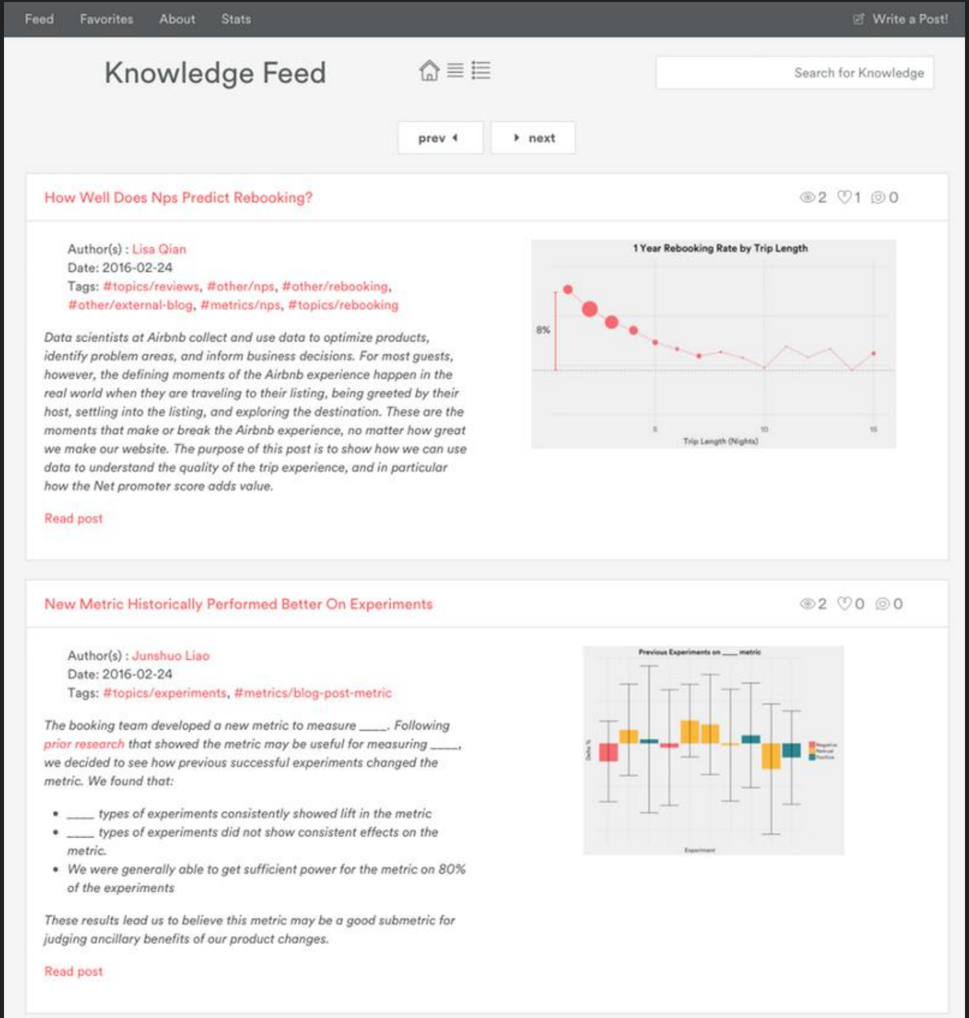
And we can easily view its spectral structure using matplotlib's builtin `specgram` routine:

In [2]:
```
%matplotlib inline
from matplotlib import pyplot as plt
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.specgram(x); ax2.set_title('Spectrogram');
```



以 jupyter nb 作為分析文件統一的交換格式


Import
Tidy
Consistent way of storing data
Transform
Create new variables & new summaries
Visualise
Surprises, but doesn't scale
Model
Scales, but doesn't (fundamentally) surprise
Program
Communicate



Airbnb knowledge repo

https://github.com/airbnb/knowledge-repo

數據成果得以
持續擴散的管道

內部教育訓練

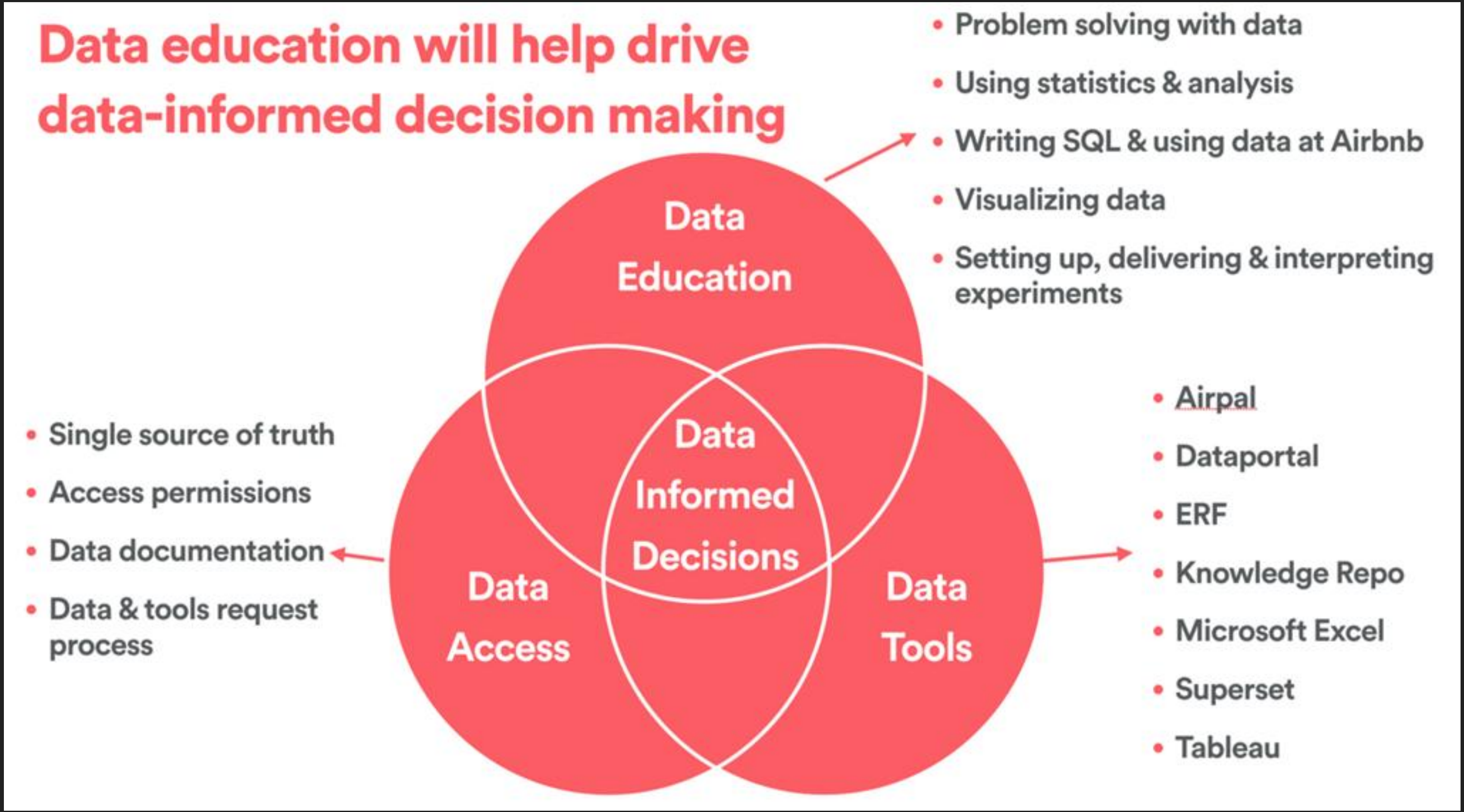
專門例會討論
資料運用議題

唯有 Full-stack solution
才能徹底解決問題

分析工具
與共用的分析環境

Best practice與
Data standards

好的資料庫架構



https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a

故事 2

https://www.junyiacademy.org/

影片　　　　題目

影片

影片 題目

題目

課綱
分類

任務

影片

影片

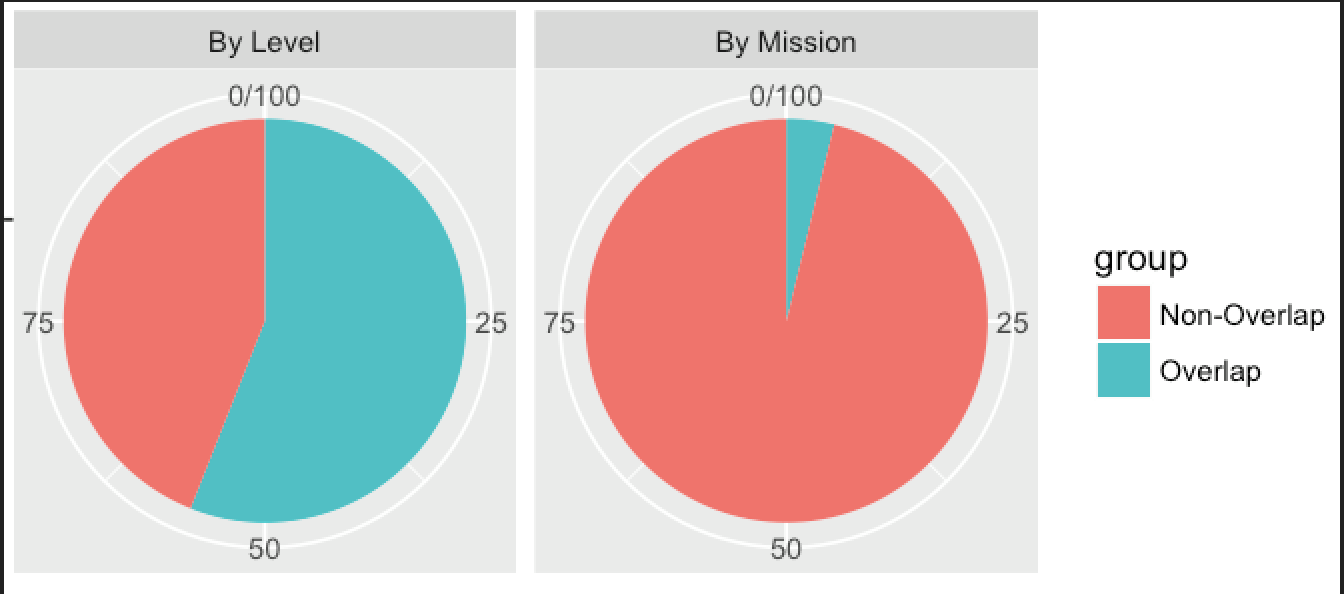
題目

題目

指派作業的UI設計

具有讓使用者自我揭露的作用

揭露：哪些物件具有學習上的關聯性?



由使用者指派的任務裡，許多(題目/影片)組合是在現行課綱找不到的

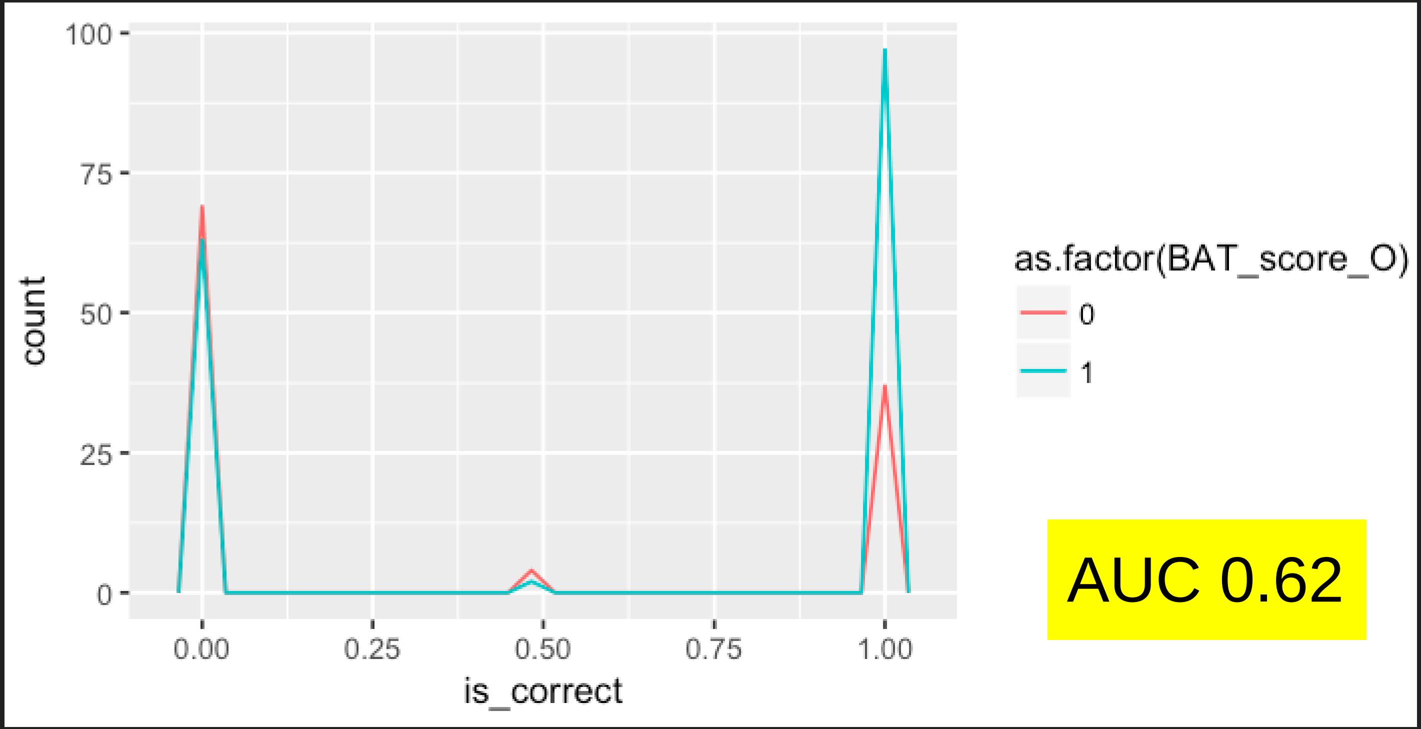
問題：

這個觀察有什麼重要性？

以某知識點的對錯去預測90天後對應的成績

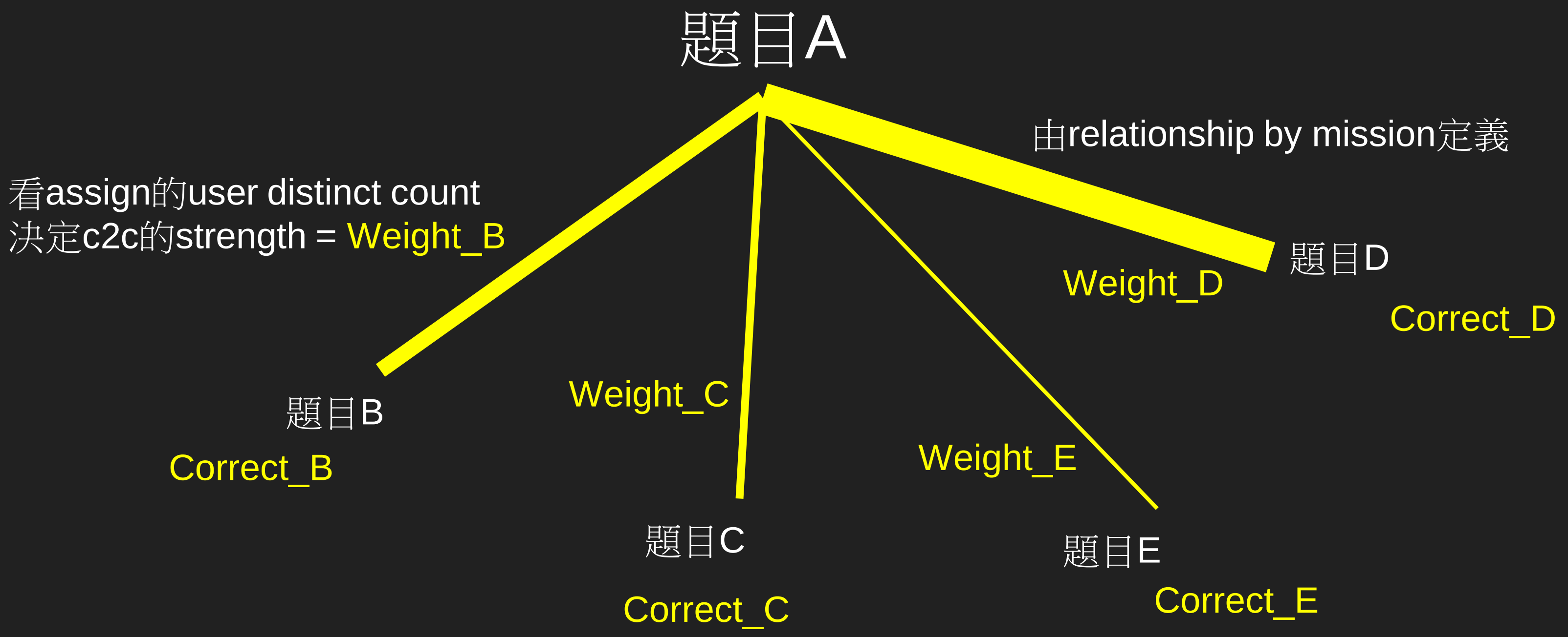
此為「能力」的
True north metric

發現：單點預測效果差

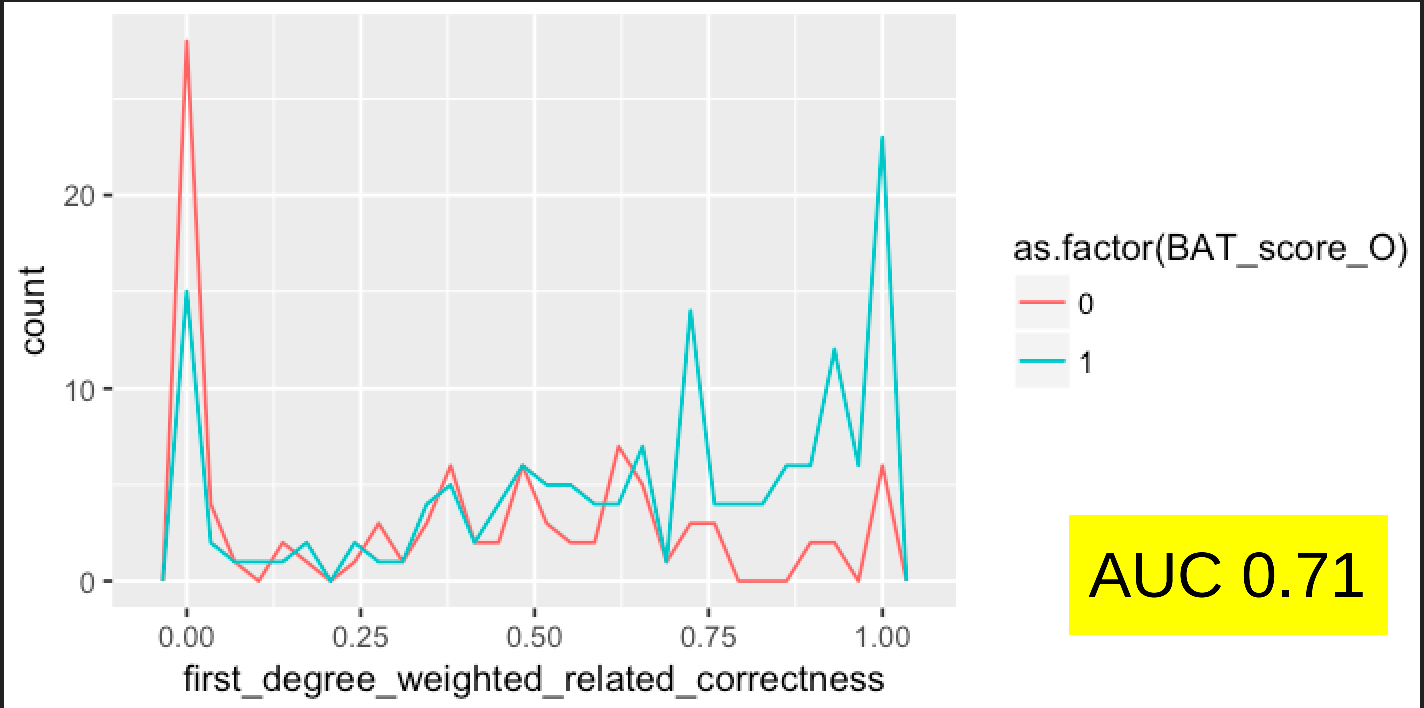


100
75
50
25
0
count
0.00
0.25
0.50
0.75
1.00
is_correct
as.factor(BAT_score_O)
0
1
AUC 0.62

$$\text{Correct\_A} = \frac{\text{SUM(Correct\_B*Weight\_B + … + Correct\_E*Weight\_E)}}{\text{SUM(Weight\_B + … + Weight\_E)}}$$



發現：網絡對單點的預測準確度提高



AUC 0.71

啓發1：

要答對能力測驗不能僅靠單點的能力

(不然不能解釋為何彼此相關的知識點的集體答對狀況，較能預測日後能力測驗成績)

啓發2：

現行課綱的侷限?
存在更好的學習方式?

(可以作為推薦系統的基礎)

啓發3：

網絡/關聯性的資料具有戰略意義

藥物A

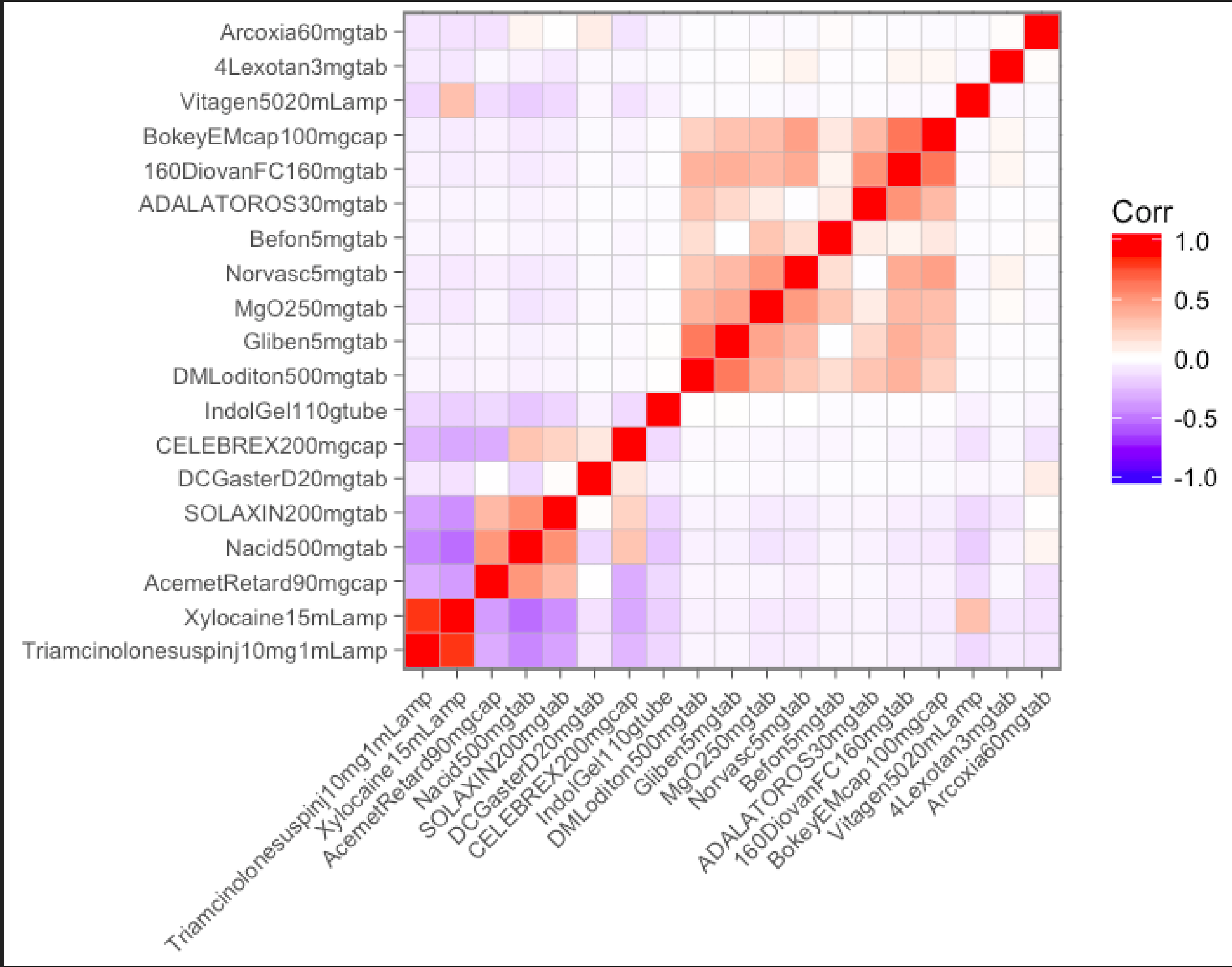
處方B

針劑D

藥物C

某病人的處方

某教授的
用藥習慣


Arcoxia60mgtab
4Lexotan3mgtab
Vitagen5020mLamp
BokeyEMcap100mgcap
160DiovanFC160mgtab
ADALATOROS30mgtab
Befon5mgtab
Norvasc5mgtab
MgO250mgtab
Gliben5mgtab
DMLoditon500mgtab
IndolGel110gtube
CELEBREX200mgcap
DCGasterD20mgtab
SOLAXIN200mgtab
Nacid500mgtab
AcemetRetard90mgcap
Xylocaine15mLamp
Triamcinolonesuspinj10mg1mLamp
Corr
1.0
0.5
0.0
-0.5
-1.0

後續：

寫成SOP

讓服務水平不因人員經驗而有差異

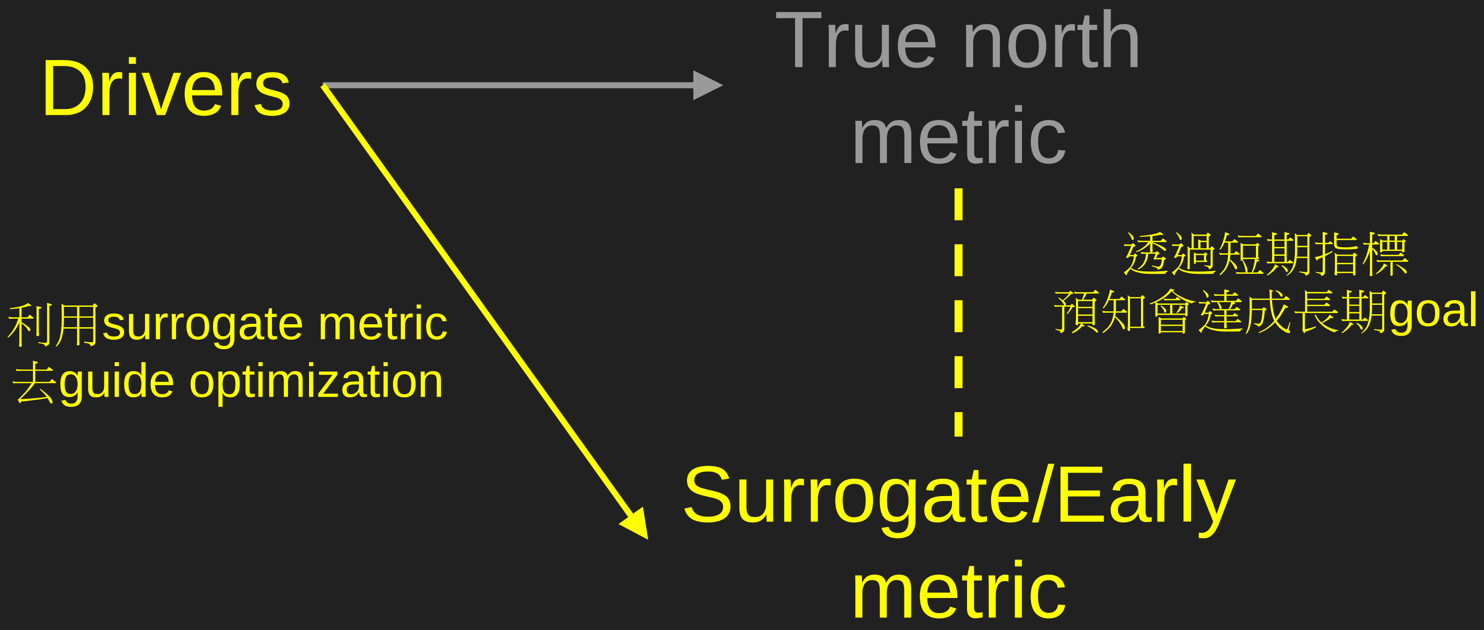
故事 3

改善界面
易用性

6個月後仍
active login

問題：

不能等到6個月後
才來驗證成效(太久了...)


Drivers
True north metric
透過短期指標
預知會達成長期goal
利用surrogate metric
去guide optimization
Surrogate/Early metric

改善界面
易用性

6個月後仍
active login

以predictive modeling
去建構早期指標

新成員在第1週完成個人資料
＋新增人脈的數量

= Quality signup

改善界面
易用性

6個月後仍
active login

利用surrogate metric
去guide optimization

透過短期指標
預知會達成長期goal

新成員在第1週完成個人資料
＋新增人脈的數量
= Quality signup

Linkedin的例子

改變
feedback
方式

12個月後仍有
active session

需要設計A/B testing
去驗證這一段因果關係

未有任何attempt就開Hint (反比)
開了Hint後的attempts > 1 (正比)

我們為均一找到的機會

啓發：

Off-task行為具有反映未來使用狀況的潛力

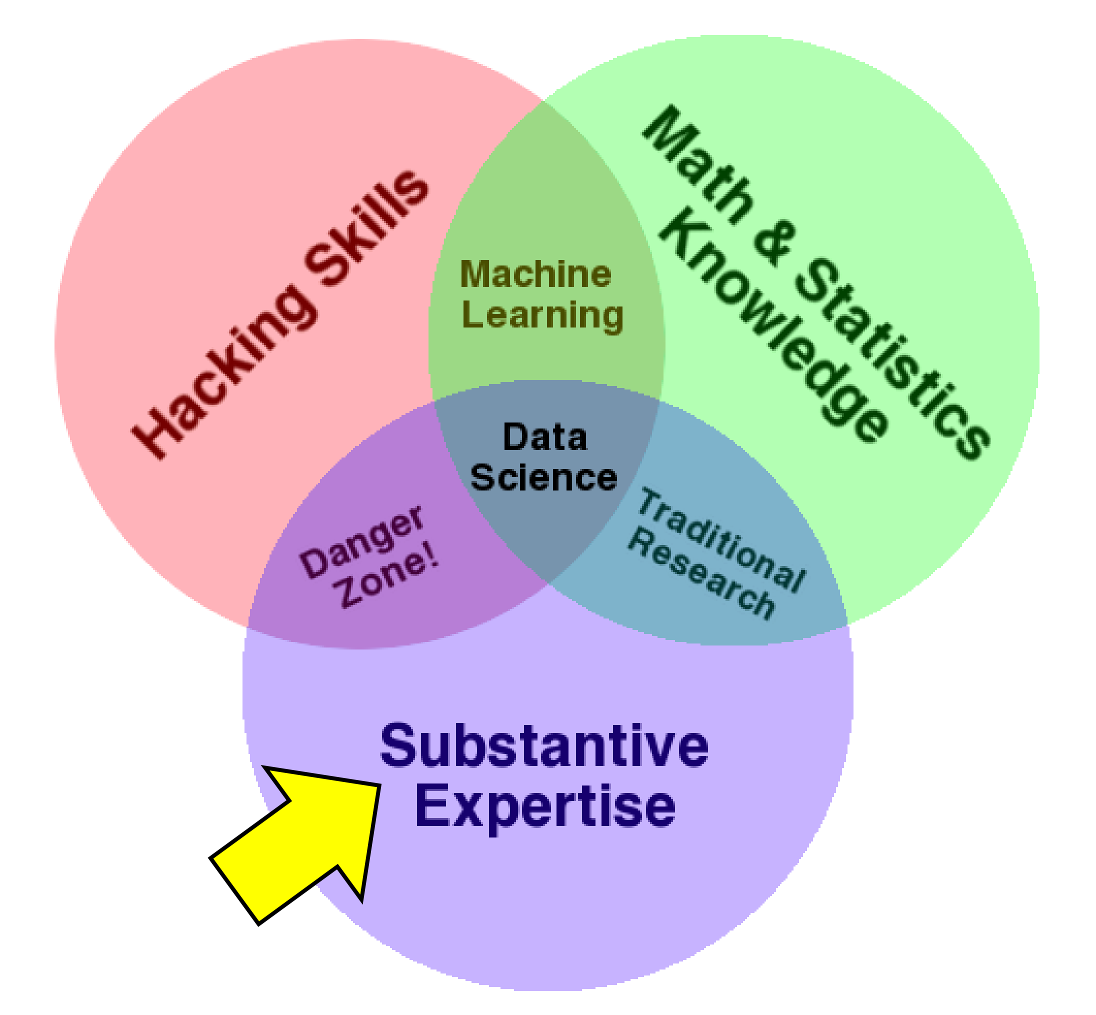
(及早預警/提醒?)

(還有很多gaming the system值得實驗)

DS的重點不在Data

在Science

不光是Data，還有Domain-relevant questions


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

DS的產出是軟體嗎?

models, dashboard, database, pipelines ...

哪些可能的 future products?

潛在的marketing strategies
跟user needs尚未被發現?

哪些資料的收集會帶來的優勢？

DS產出的 Actionable insights 具有策略性質

協助組織運用資料加速成長

才是發展DS最重要的目的

DS產出是 knowledge

而程式/軟體只是工具

Data workflow 才是把資料轉成知識的架構

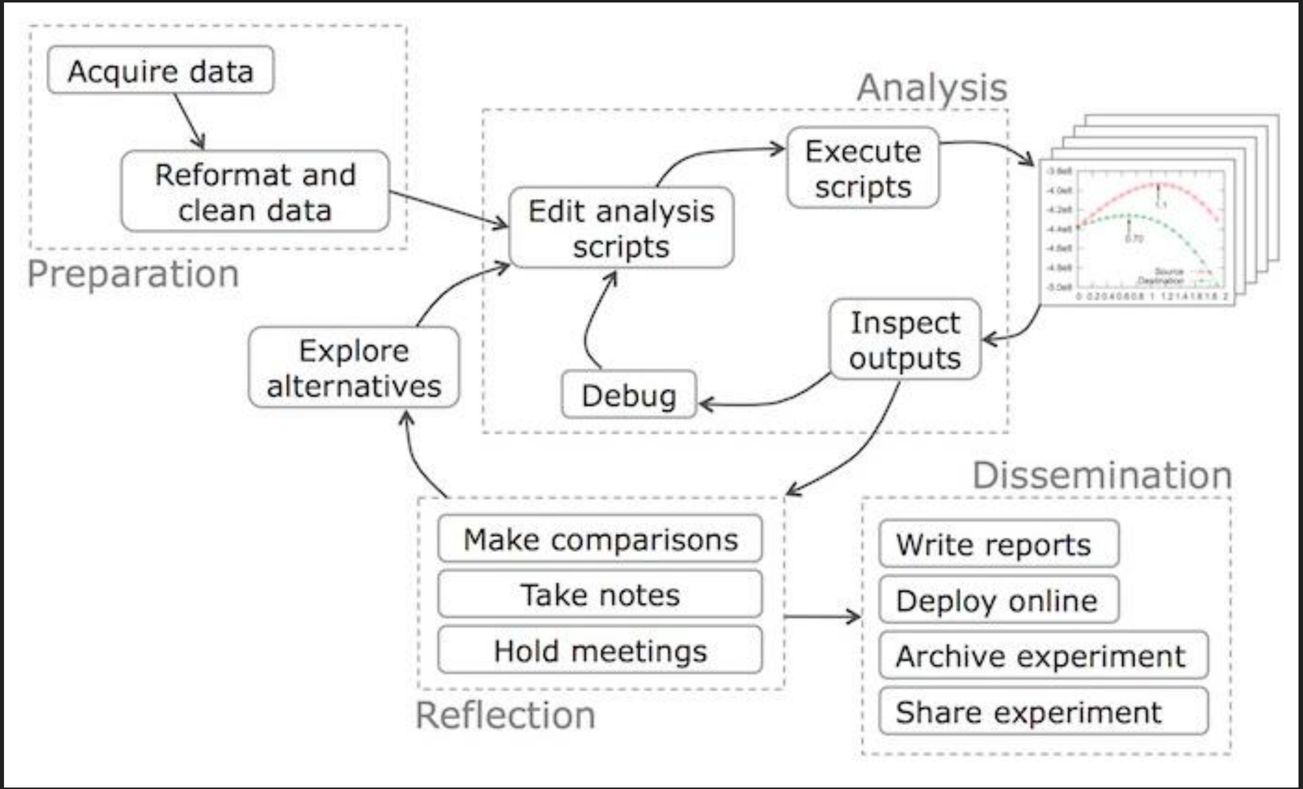
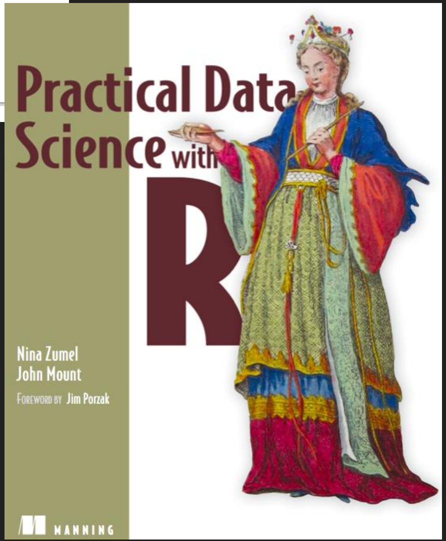


Table 1.1 Data science project roles and responsibilities

Role	Responsibilities
Project sponsor	Represents the business interests; champions the project
Client	Represents end users' interests; domain expert
Data scientist	Sets and executes analytic strategy; communicates with sponsor and client
Data architect	Manages data and data storage; sometimes manages data collection
Operations	Manages infrastructure; deploys final project results



從各端實際需要，去發現資料可以幫忙的地方 = 收集好的question跟痛點

業務端
產品端
客服端
…

Data Product Manager

Data engineer

Data scientist

協助後續追蹤、發展、測試及整合進產品

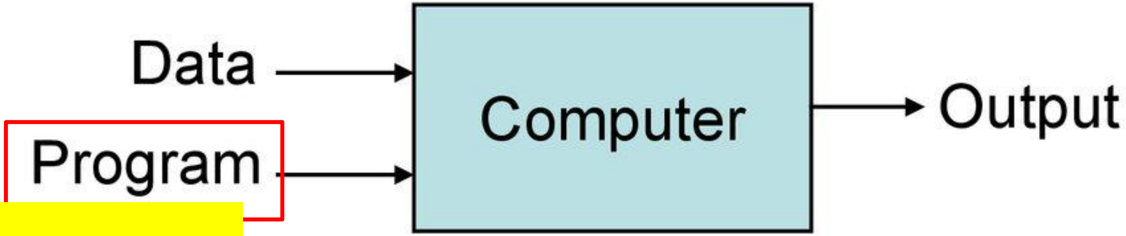
Data science

以data回答問題, 獲取並運用knowledge的手段

Machine learning

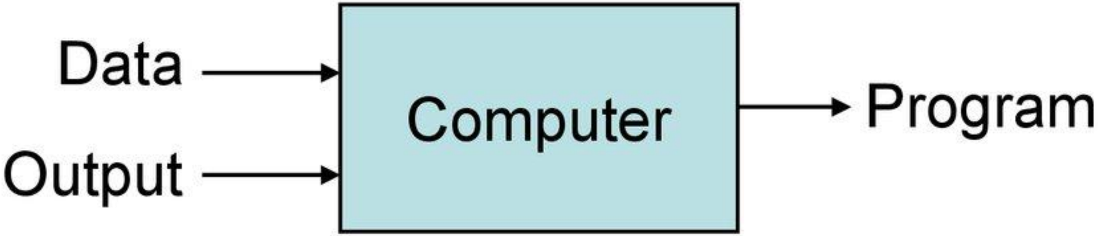
命令電腦做事情的paradigm shift

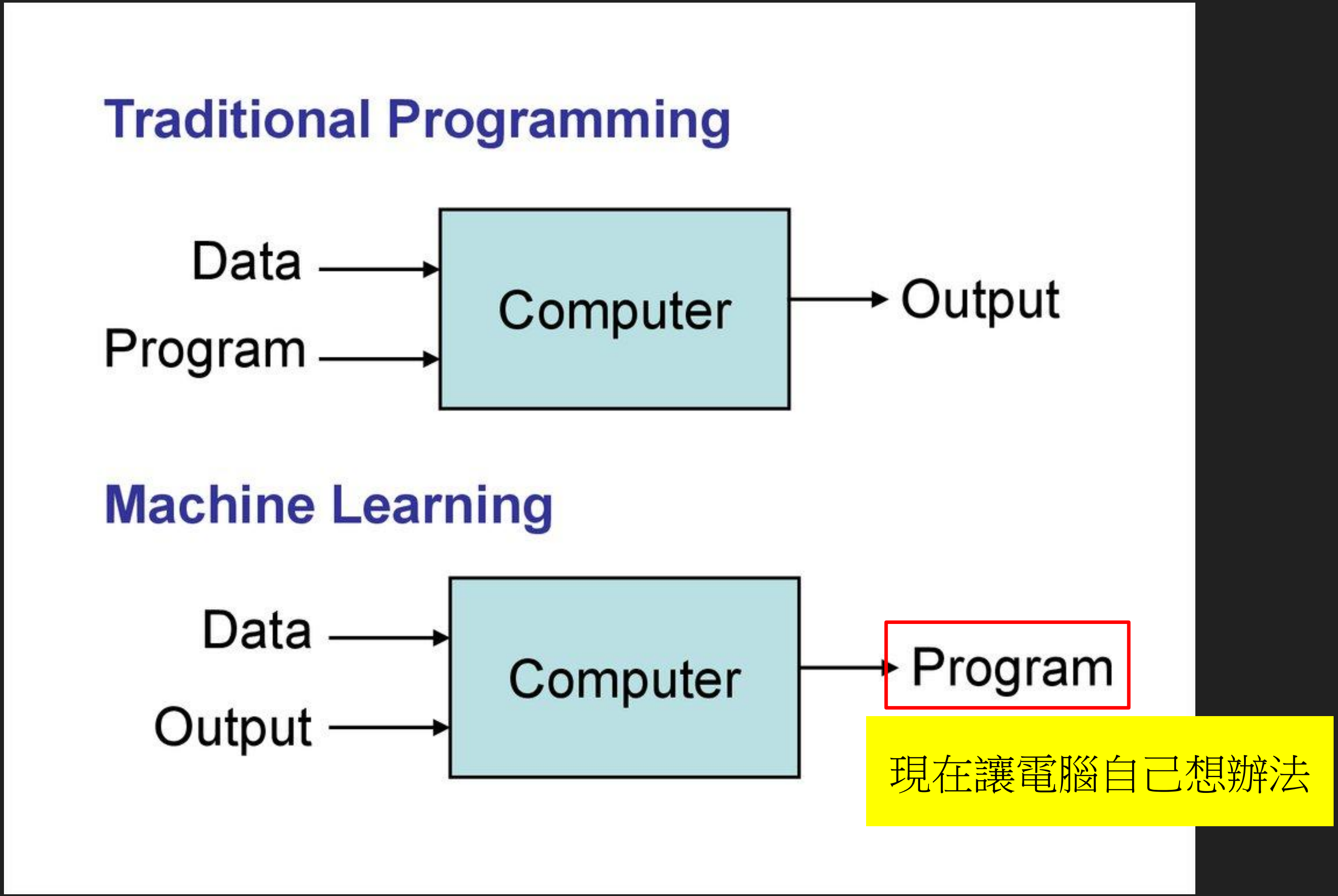
Traditional Programming

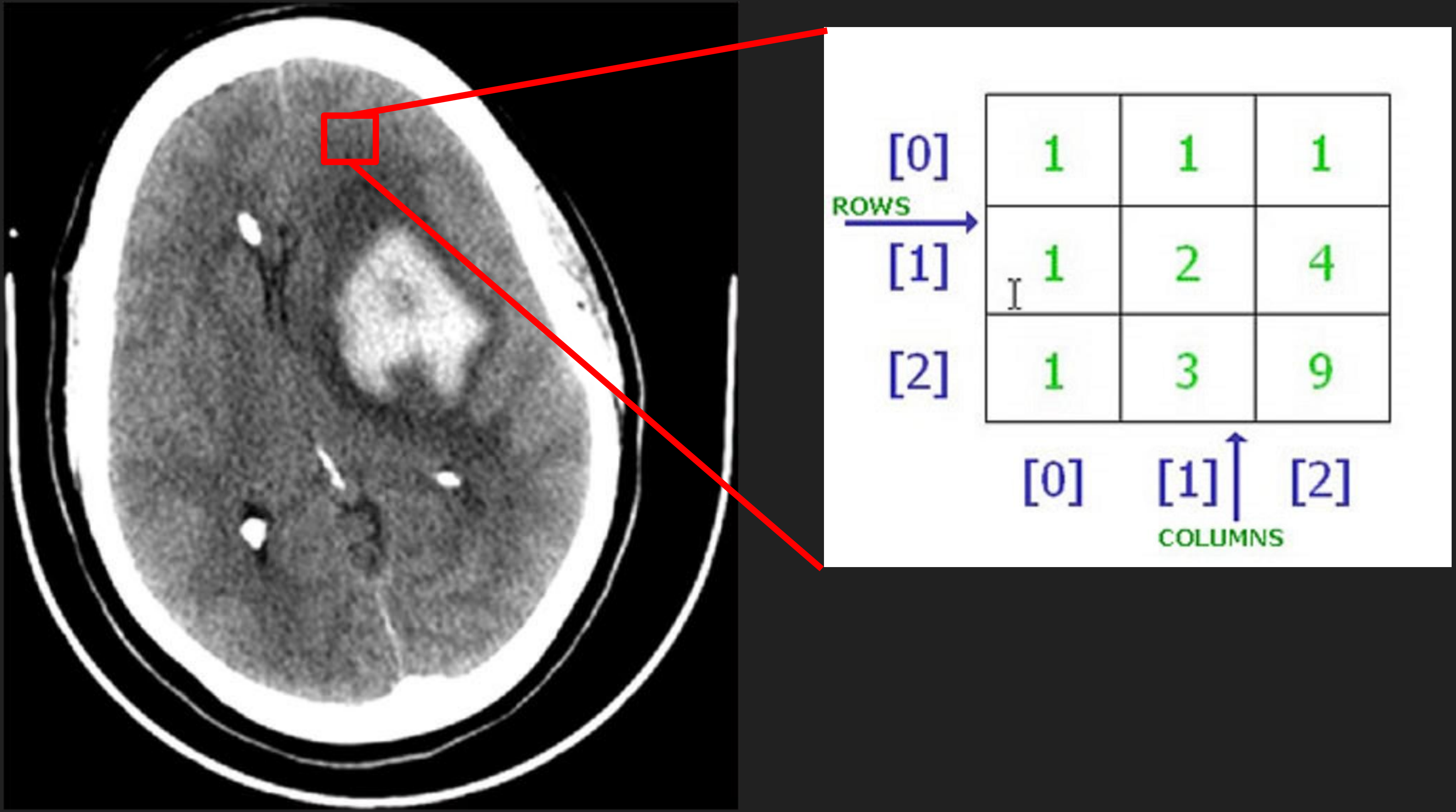


過去這一步最傷腦筋

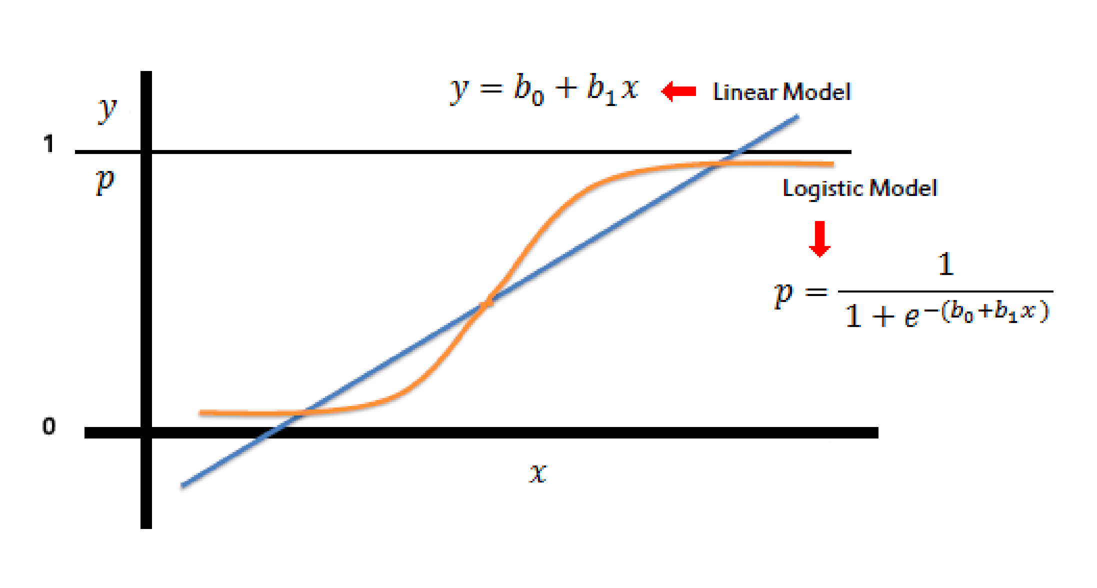
Machine Learning



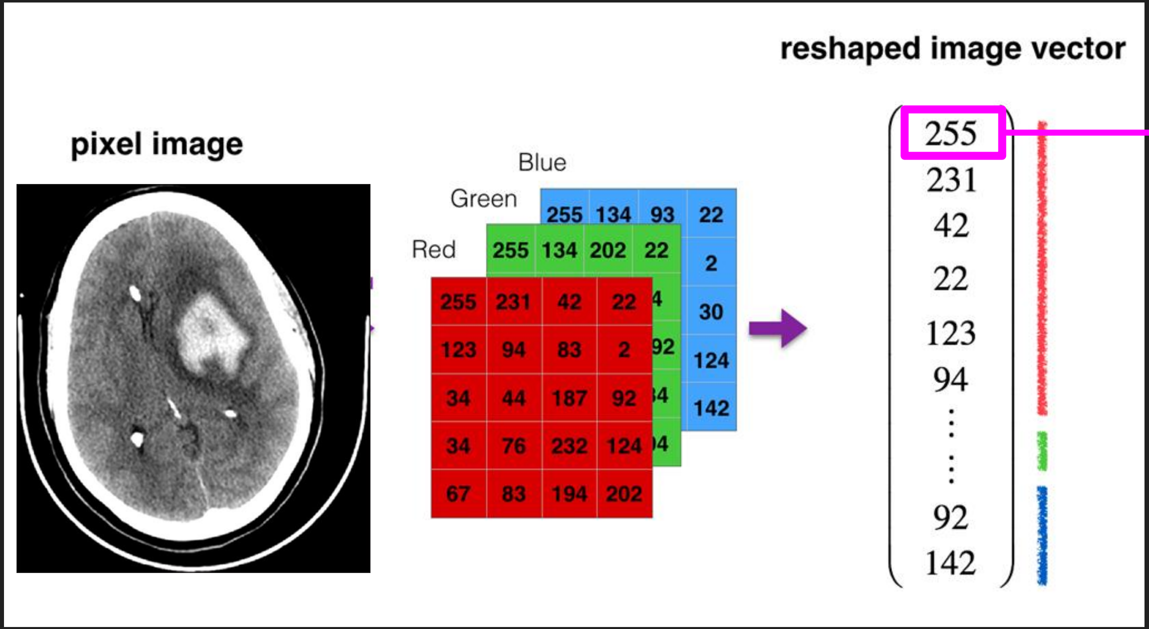

Traditional Programming
Data
Program
Computer
Output
Machine Learning
Data
Output
Computer
Program
現在讓電腦自己想辦法


[0]
ROWS
[1]
[2]
1 1 1
1 2 4
1 3 9
[0] [1] [2]
COLUMNS

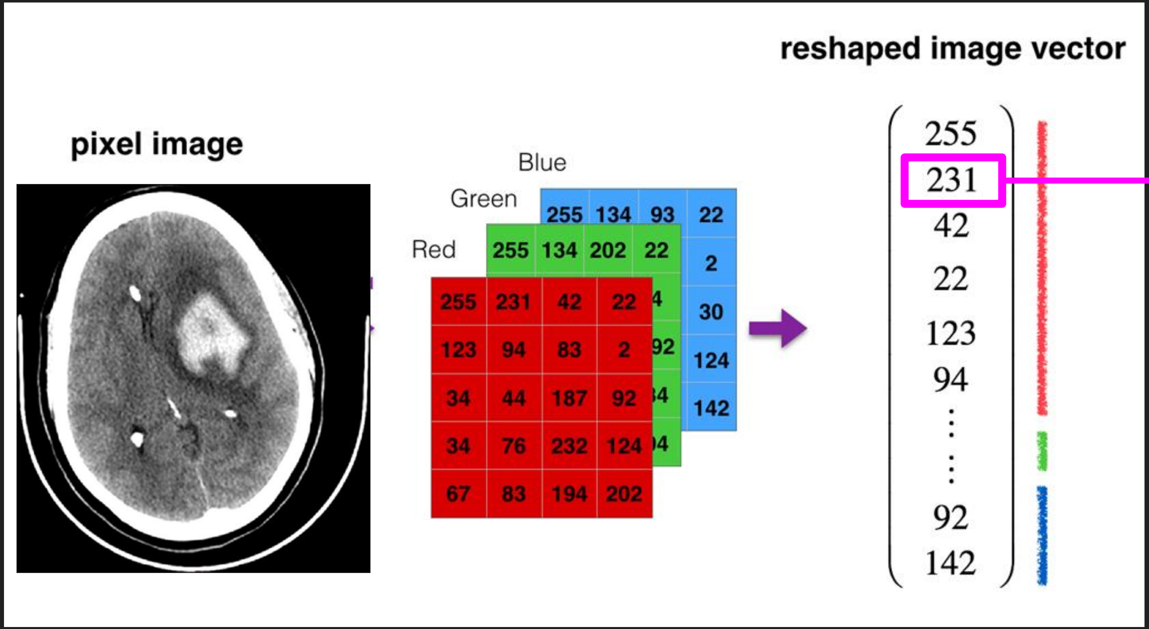
$$Ln\left(\frac{P}{1-P}\right)=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k$$



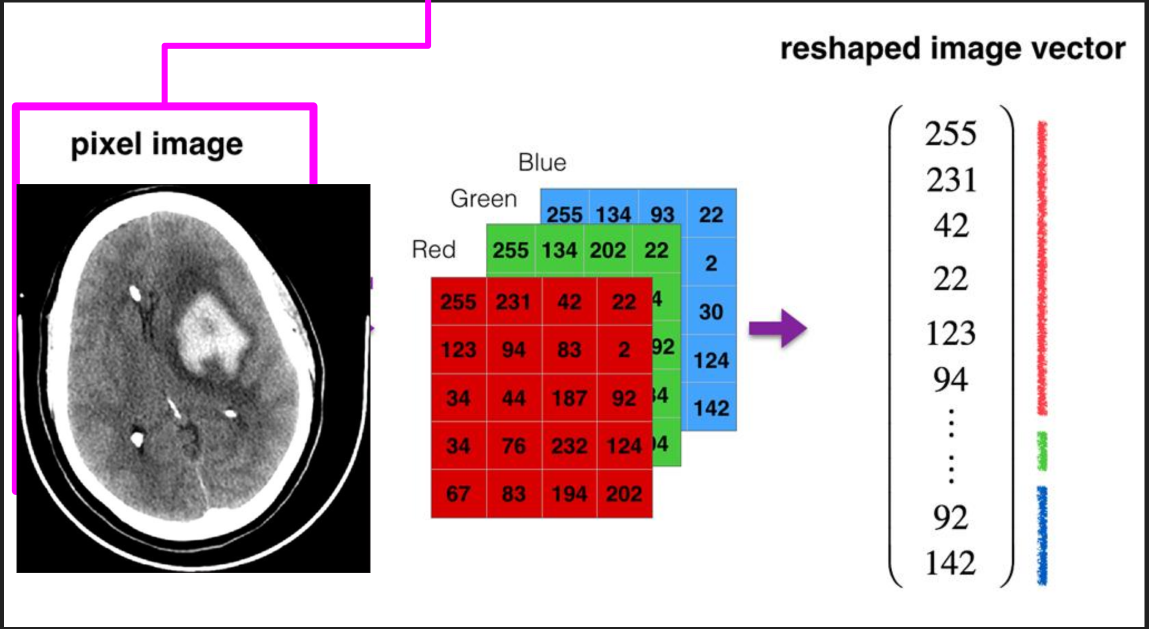
$$Ln\left(\frac{P}{1-P}\right)=\beta_0+\beta_1\boxed{X_1}+\beta_2X_2+\ldots+\beta_kX_k$$



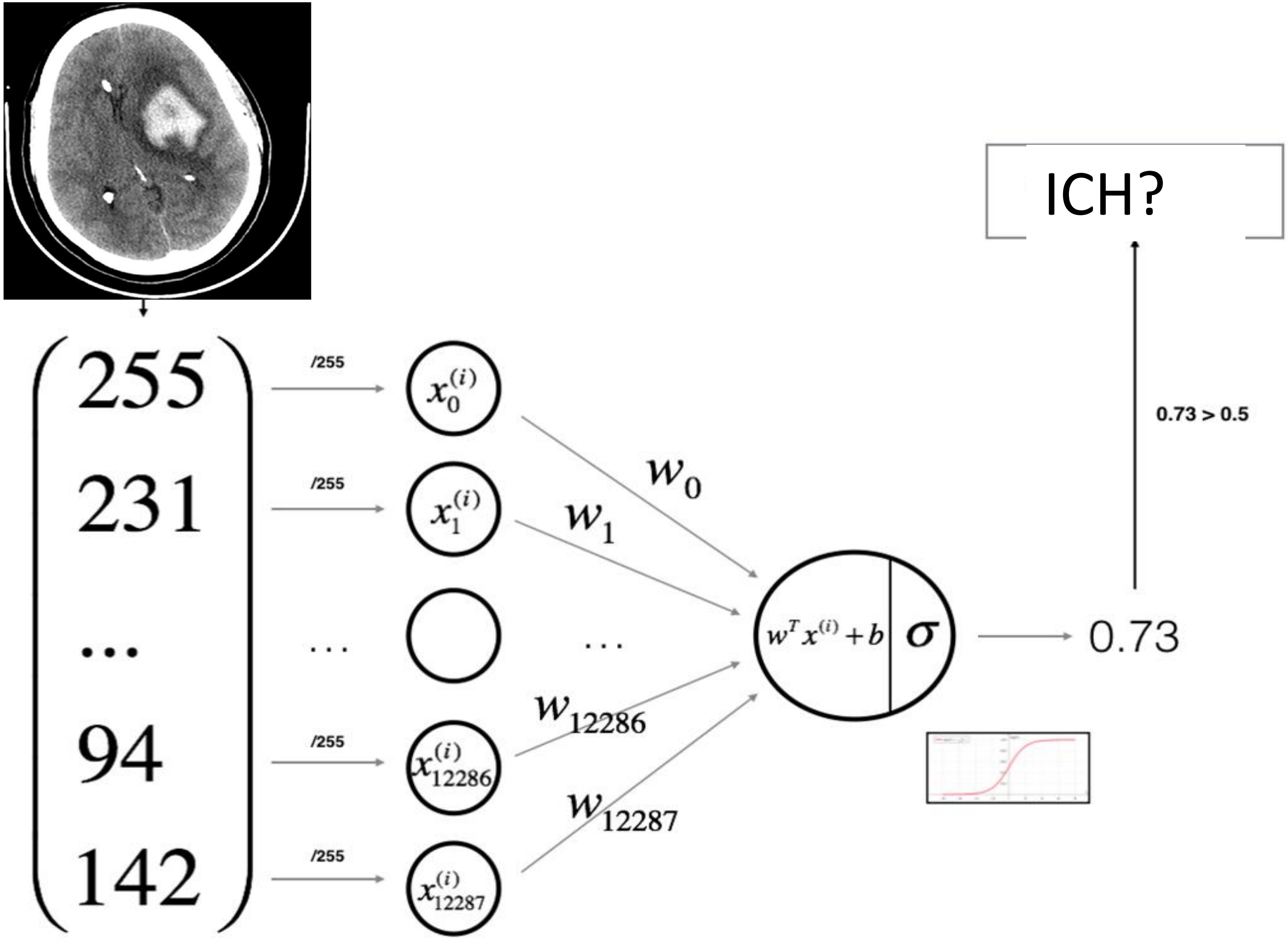
$$Ln\left(\frac{P}{1-P}\right)=\beta_0+\beta_1 X_1+\beta_2 \boxed{X_2}+\dots+\beta_k X_k$$

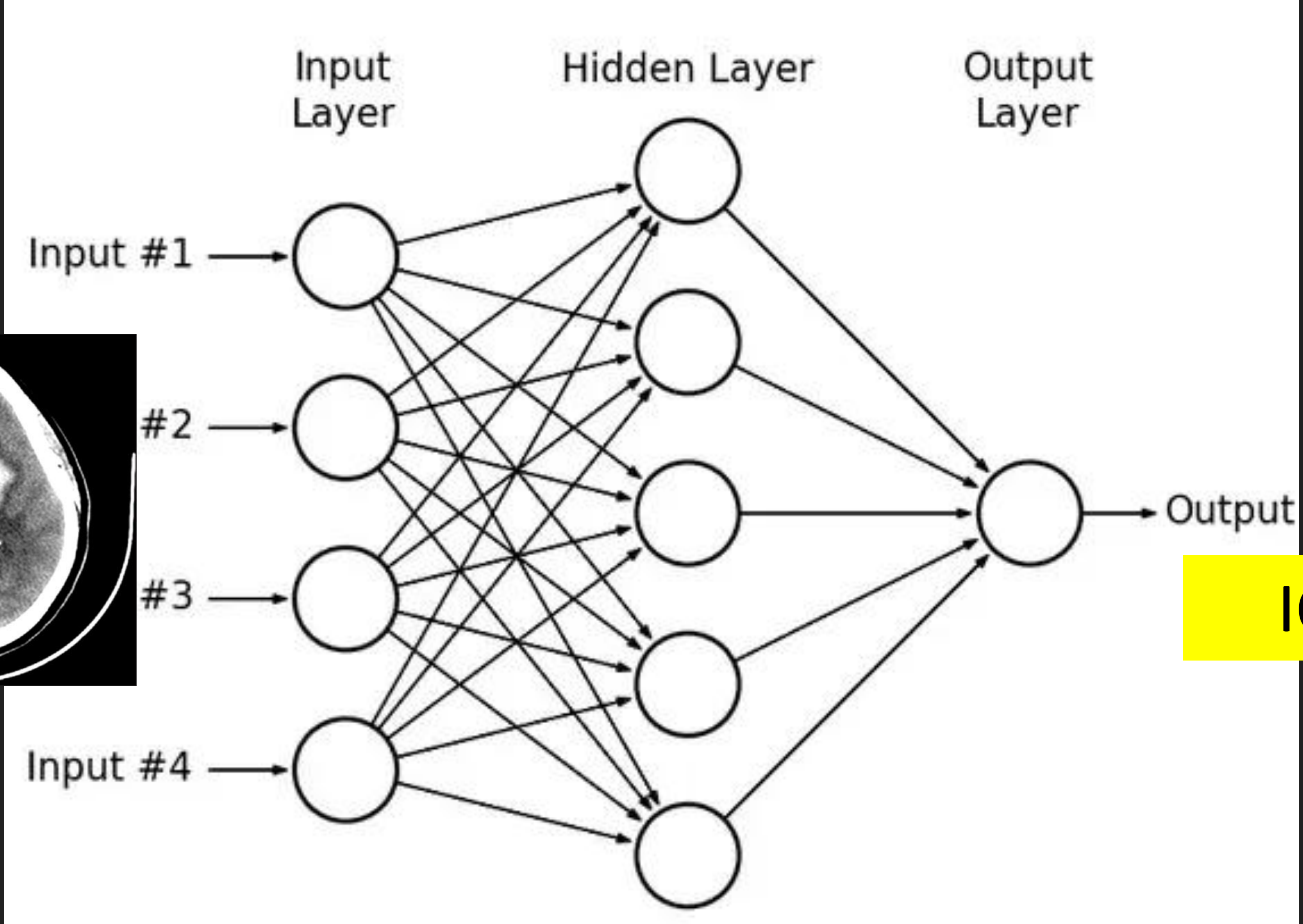


$$Ln\left(\boxed{\frac{P}{1-P}}\right)=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k$$

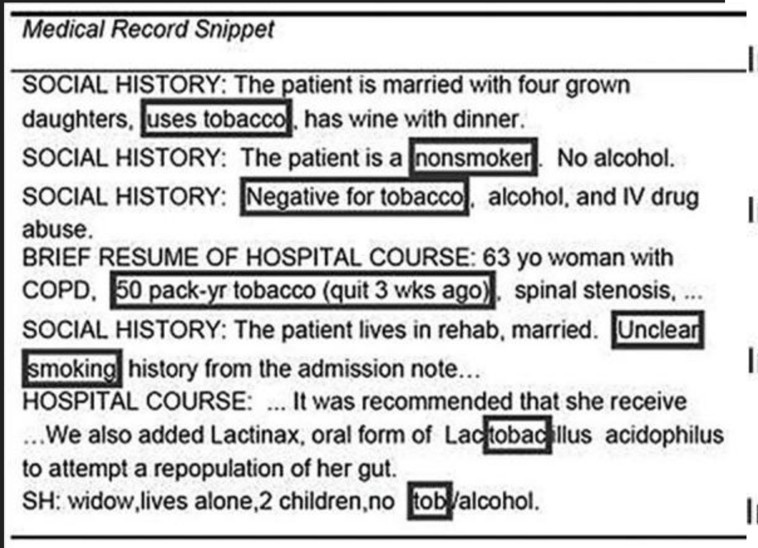


$$Ln\left(\frac{P}{1-P}\right)=\underbrace{\boxed{\beta_0}+\boxed{\beta_1}X_1+\boxed{\beta_2}X_2+\ldots+\boxed{\beta_k}X_k}_{\text{權重電腦自己會去”學”}}$$


ICH?
0.73 > 0.5
255
231
...
94
142
/255
/255
/255
/255
$x_0^{(i)}$
$x_1^{(i)}$
$x_{12286}^{(i)}$
$x_{12287}^{(i)}$
w_0
w_1
w_{12286}
w_{12287}
$w^T x^{(i)} + b$
σ
0.73


Input Layer
Hidden Layer
Output Layer
Input #1
#2
#3
Input #4
Output

ICH?



Medical Record Snippet

SOCIAL HISTORY: The patient is married with four grown daughters, uses tobacco, has wine with dinner.

SOCIAL HISTORY: The patient is a nonsmoker. No alcohol.

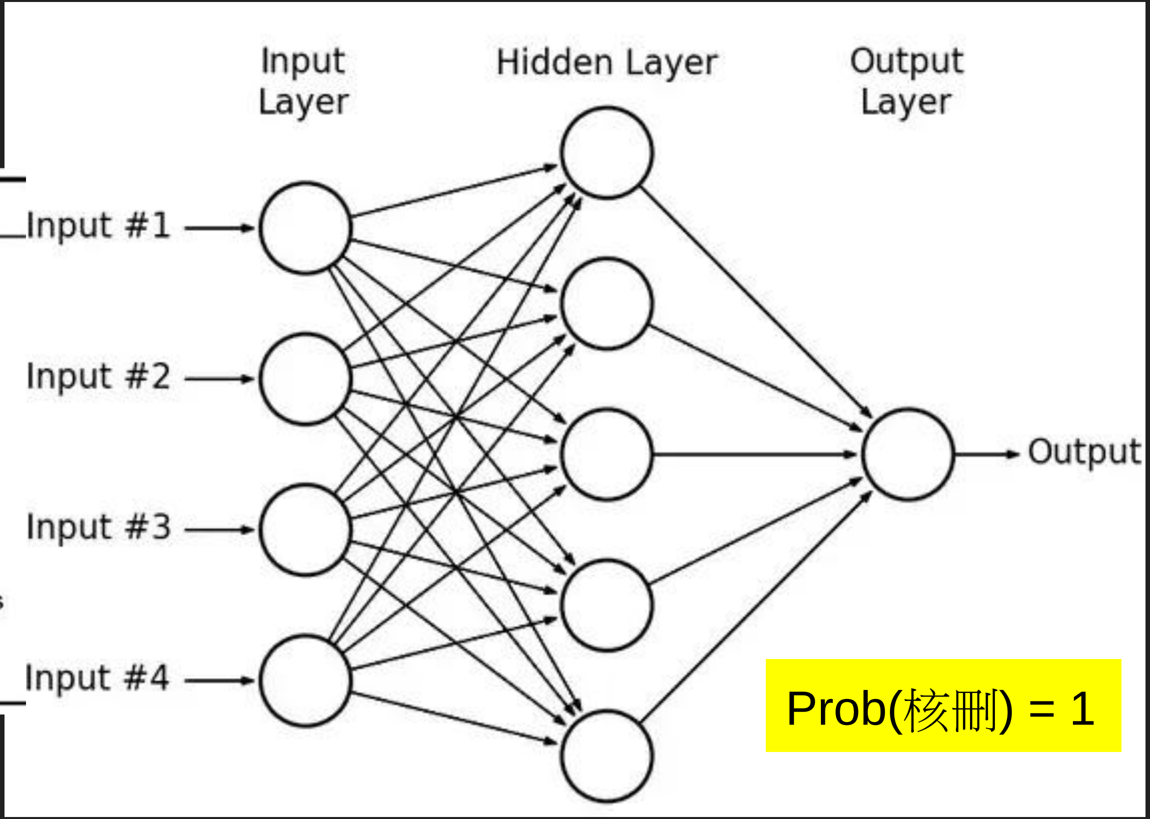
SOCIAL HISTORY: Negative for tobacco, alcohol, and IV drug abuse.

BRIEF RESUME OF HOSPITAL COURSE: 63 yo woman with COPD, 50 pack-yr tobacco (quit 3 wks ago), spinal stenosis, ...

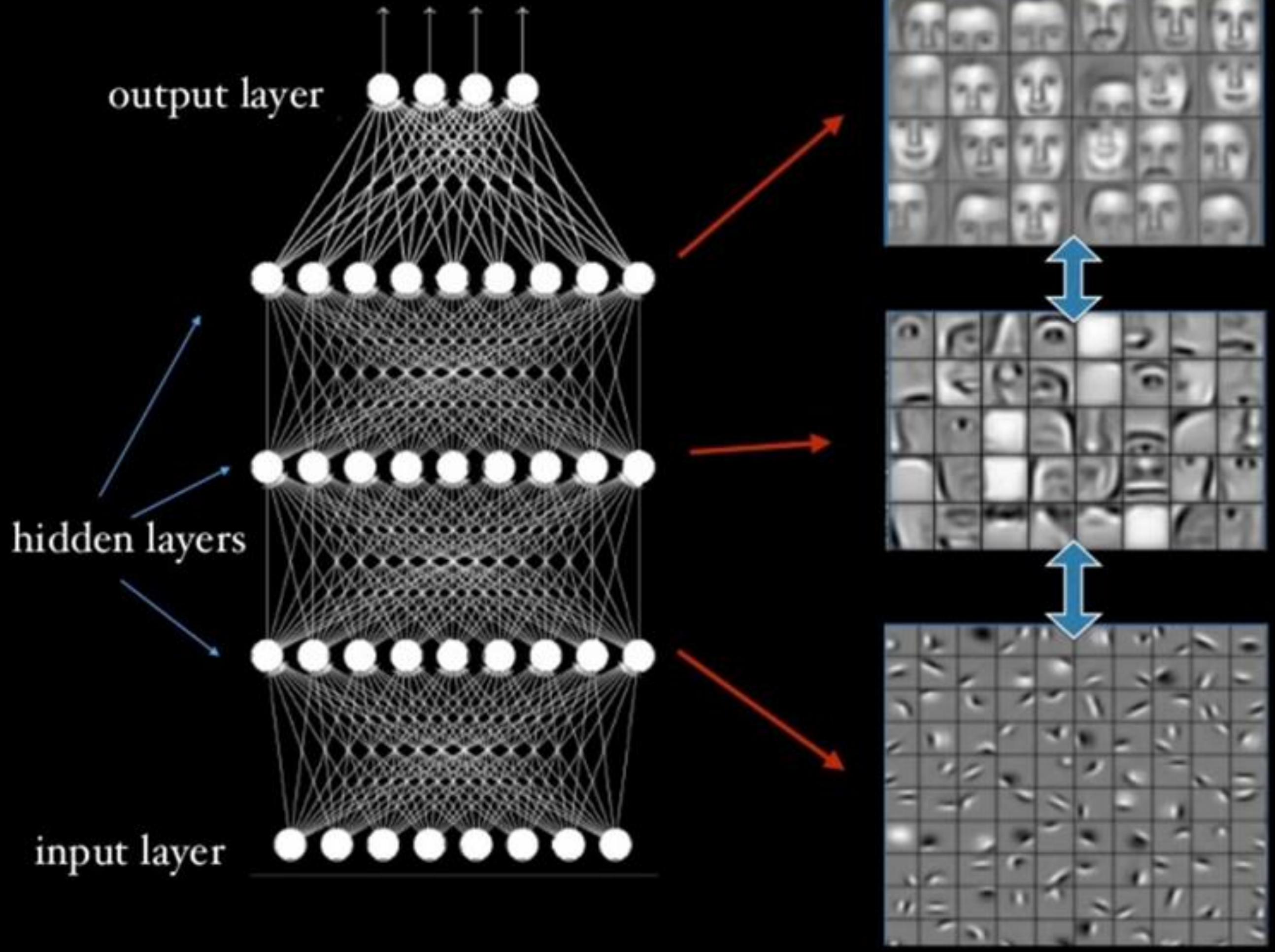
SOCIAL HISTORY: The patient lives in rehab, married. Unclear smoking history from the admission note...

HOSPITAL COURSE: ... It was recommended that she receive ...We also added Lactinax, oral form of Lactobacillus acidophilus to attempt a repopulation of her gut.

SH: widow,lives alone,2 children,no tob/alcohol.



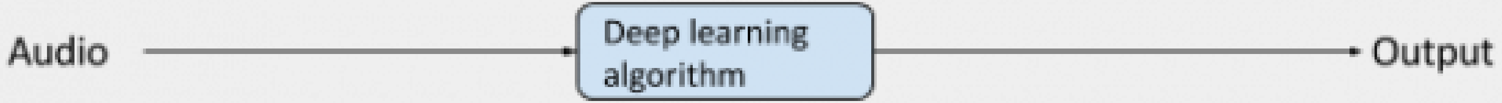
Feature Hierarchies: Vision



Speech recognition

Traditional model:

End-to-end learning:



此即所謂的End-to-end learning
避免中間的feature engineering

Stanford ML Group

Cardiologist-Level Arrhythmia Detection With Convolutional Neural Networks

Pranav Rajpurkar*, Awni Hannun*, Masoumeh Haghpanahi, Codie Bourn, and Andrew Ng

A collaboration between Stanford University and iRhythm Technologies

We develop a model which can diagnose irregular heart rhythms, also known as arrhythmias, from single-lead ECG signals better than a cardiologist.

Key to exceeding expert performance is a deep convolutional network which can map a sequence of ECG samples to a sequence of arrhythmia annotations along with a novel dataset two orders of magnitude larger than previous datasets of its kind.

ARTICLE PREVIEW

view full access options

NATURE | LETTER

日本語要約

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Affiliations | Contributions | Corresponding authors

Nature **542**, 115–118 (02 February 2017) | doi:10.1038/nature21056
Received 28 June 2016 | Accepted 14 December 2016 | Published online 25 January 2017
Corrigendum (June, 2017)

Editor's summary

العربية

Andre Esteva *et al.* used 129,450 clinical images of skin disease to train a deep convolutional neural network to classify skin lesions. The result is an algorithm that can classify lesions from photog…

Associated links

News & Views
Medicine: The final frontier in cancer diagnosis
by Leachman and Merlino

Related video

Digital doctor: AI singles out skin cancer from photos

Black-box methods?

Not knowing how the prediction came from...

How to trust the model is making reasonable predictions in general?

The Mythos of Model Interpretability

Zachary C. Lipton [1]

Statistical Science
2010, Vol. 25, No. 3, 289–310
DOI: 10.1214/10-STS330

To Explain or to Predict?

Galit Shmueli

EXPLAINABLE ARTIFICIAL INTELLIGENCE: UNDERSTANDING, VISUALIZING AND INTERPRETING DEEP LEARNING MODELS

Wojciech Samek[1], *Thomas Wiegand*[1,2], *Klaus-Robert Müller*[2,3,4]

[1]Dept. of Video Coding & Analytics, Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany
[2]Dept. of Computer Science,Technische Universität Berlin, 10587 Berlin, Germany
[3]Dept. of Brain & Cognitive Engineering, Korea University, Seoul 136-713, South Korea
[4]Max Planck Institute for Informatics, Saarbrücken 66123, Germany

"Why Should I Trust You?" Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

Trust

- Not just how often it is right
- But also which examples it is right ?

Causality

Interpretable

Being fair & Ethical

Transferability

- capacity to generalize to unfamiliar situations

Properties of interpretable models

1. Transparency

How does the model work?

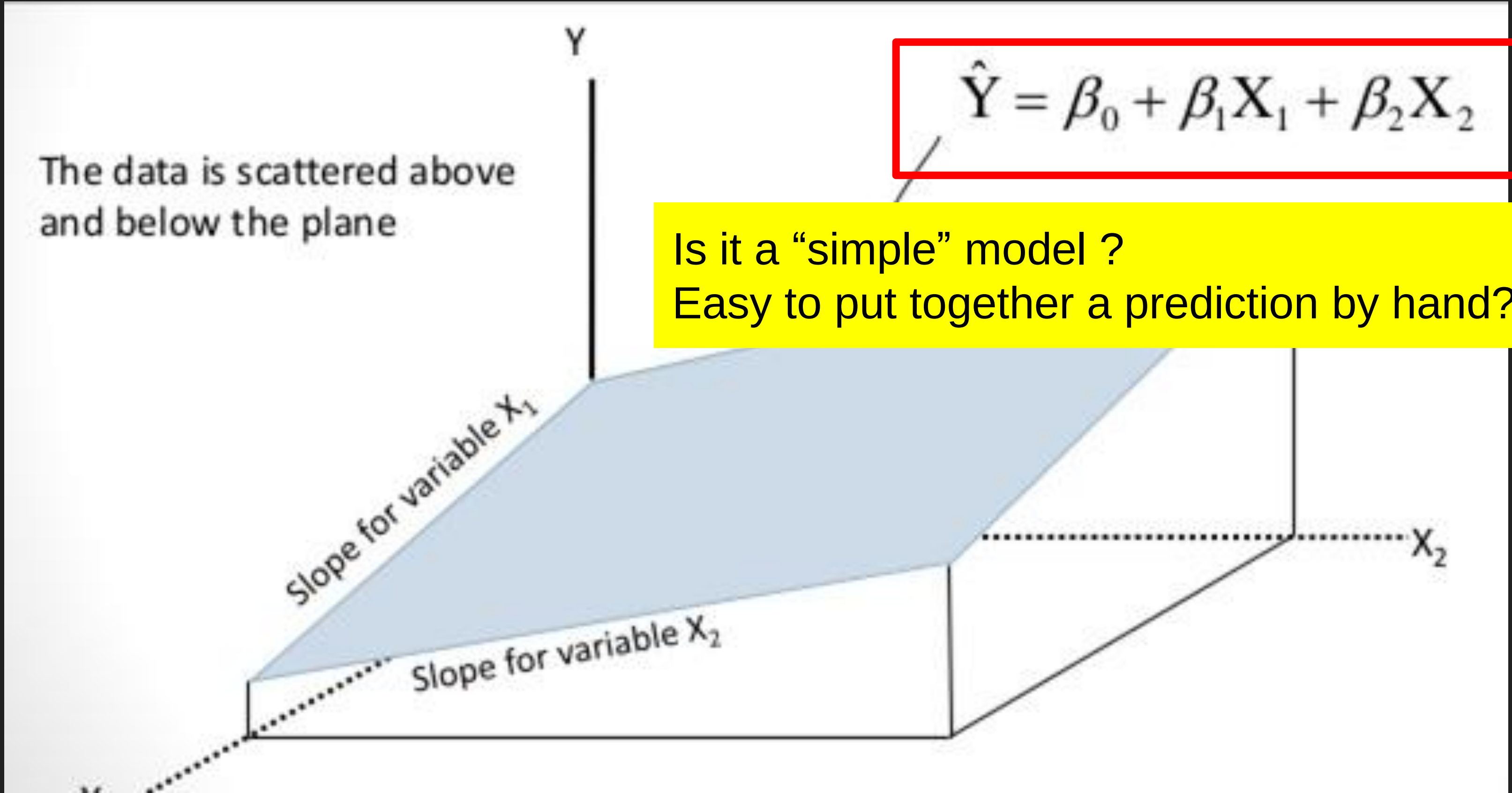
2. Post-hoc interpretability

Besides the prediction, what else can the model tell me?

→ Learning the model locally around the prediction

Transparency

- Simulatability (entire model level)

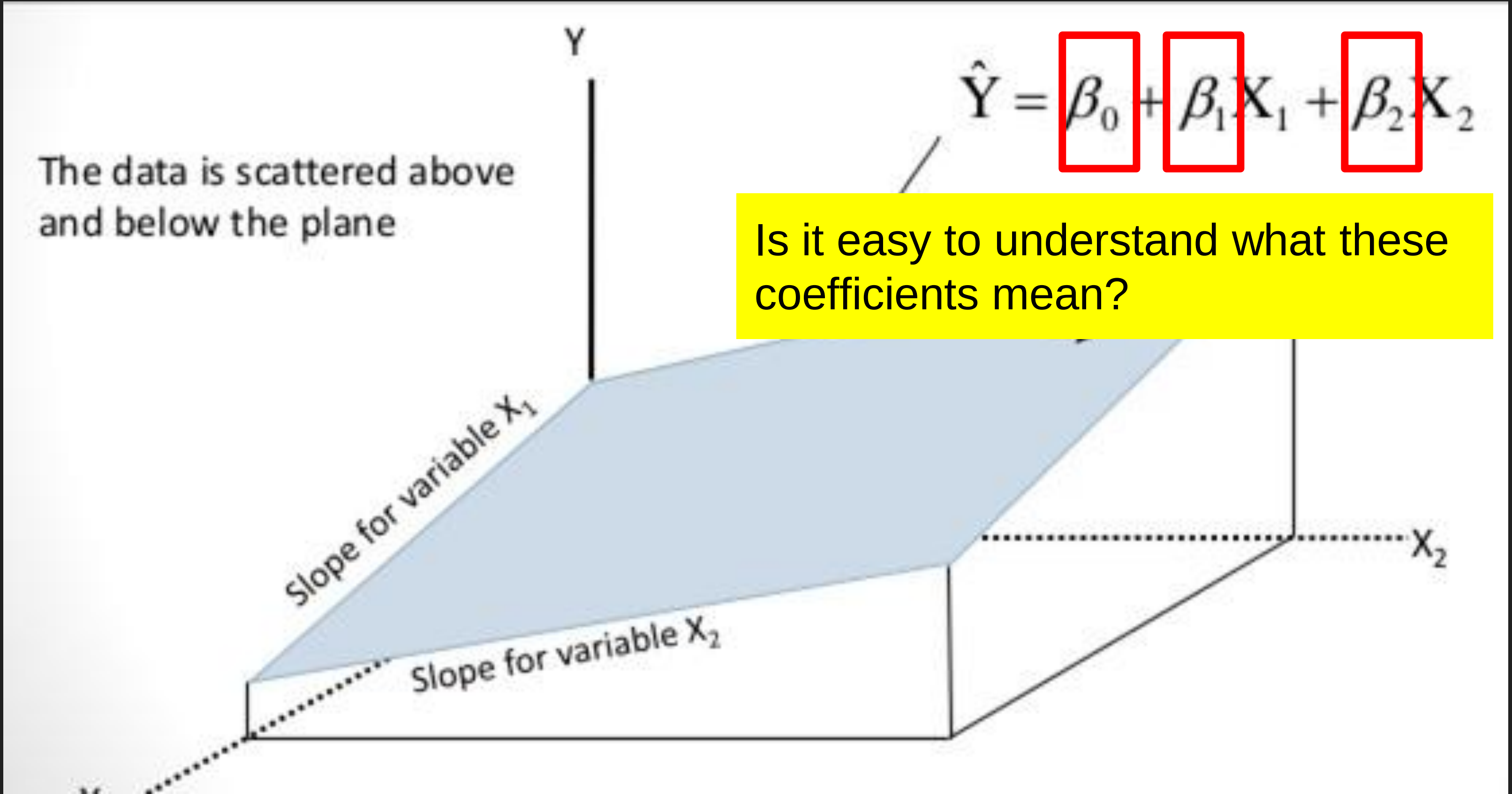


$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Is it a “simple” model ?
Easy to put together a prediction by hand?

Transparency

- Decomposability (parameters level)

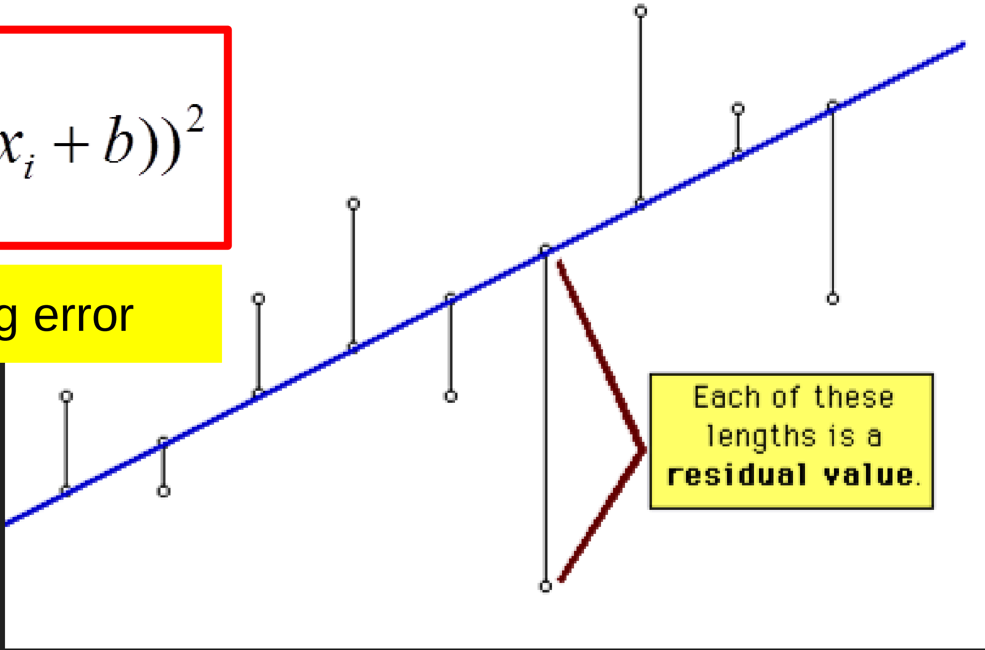


Transparency

- Algorithmic transparency
  (algorithm level)

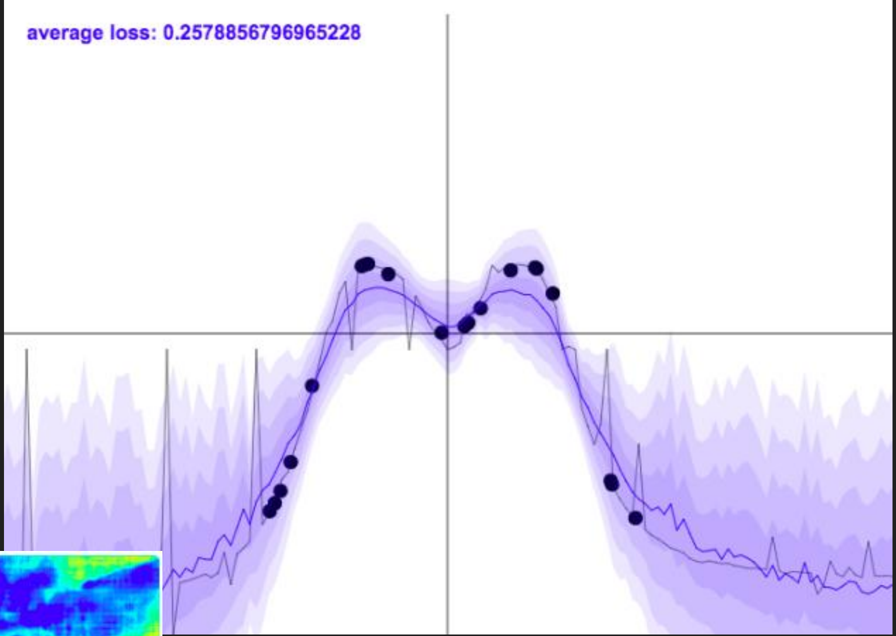
$$\text{Error}_{(\text{m,b})} = \frac{1}{N}\sum_{i=1}^{N}(y_i - (mx_i + b))^2$$

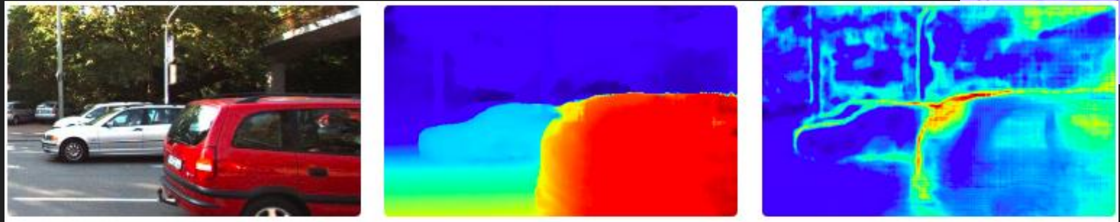
The model learns by minimizing error



Post-hoc interpretability

- Visualizing uncertainty

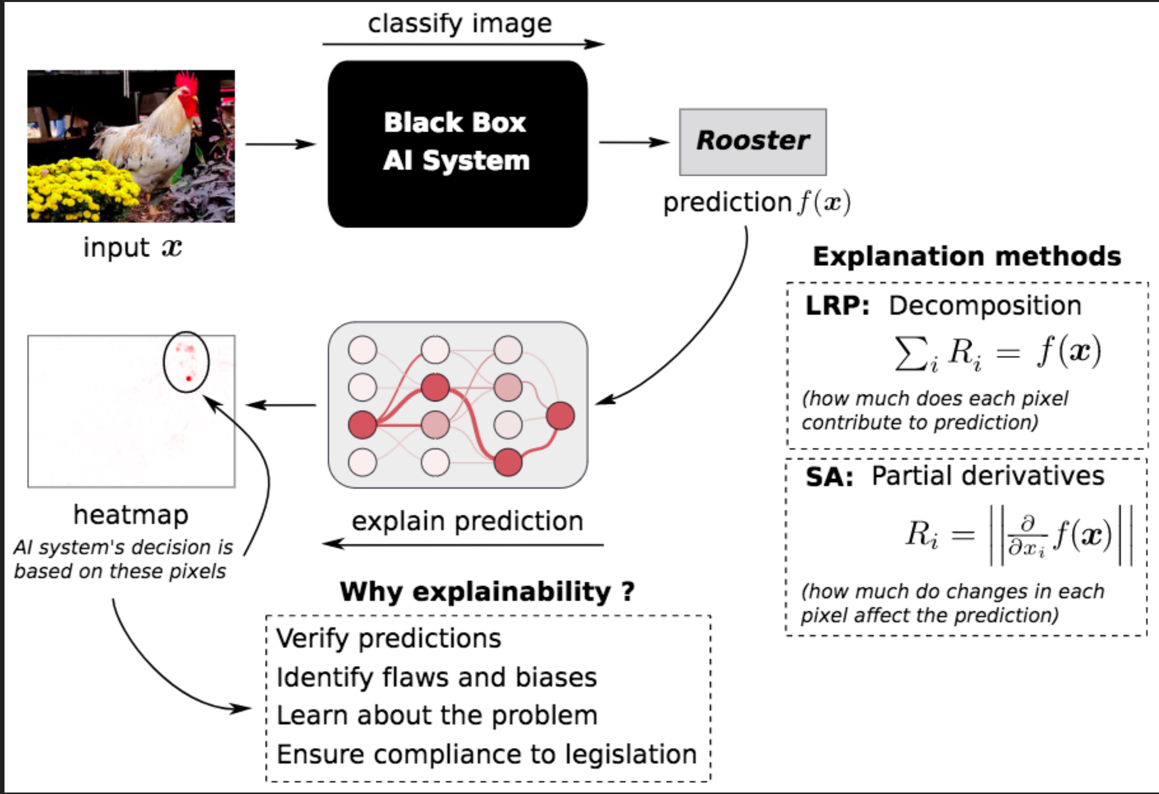




An example of why it is really important to understand uncertainty for depth estimation. The first image is an example input into a Bayesian neural network which estimates depth, as shown by the second image. The third image shows the estimated uncertainty. You can see the model predicts the wrong depth on difficult surfaces, such as the red car's reflective and transparent windows. Thankfully, the Bayesian deep learning model is also aware it is wrong and exhibits increased uncertainty.

We Need Bayesian Deep Learning for Safe AI
https://goo.gl/oQAepZ
https://goo.gl/XEKbM9

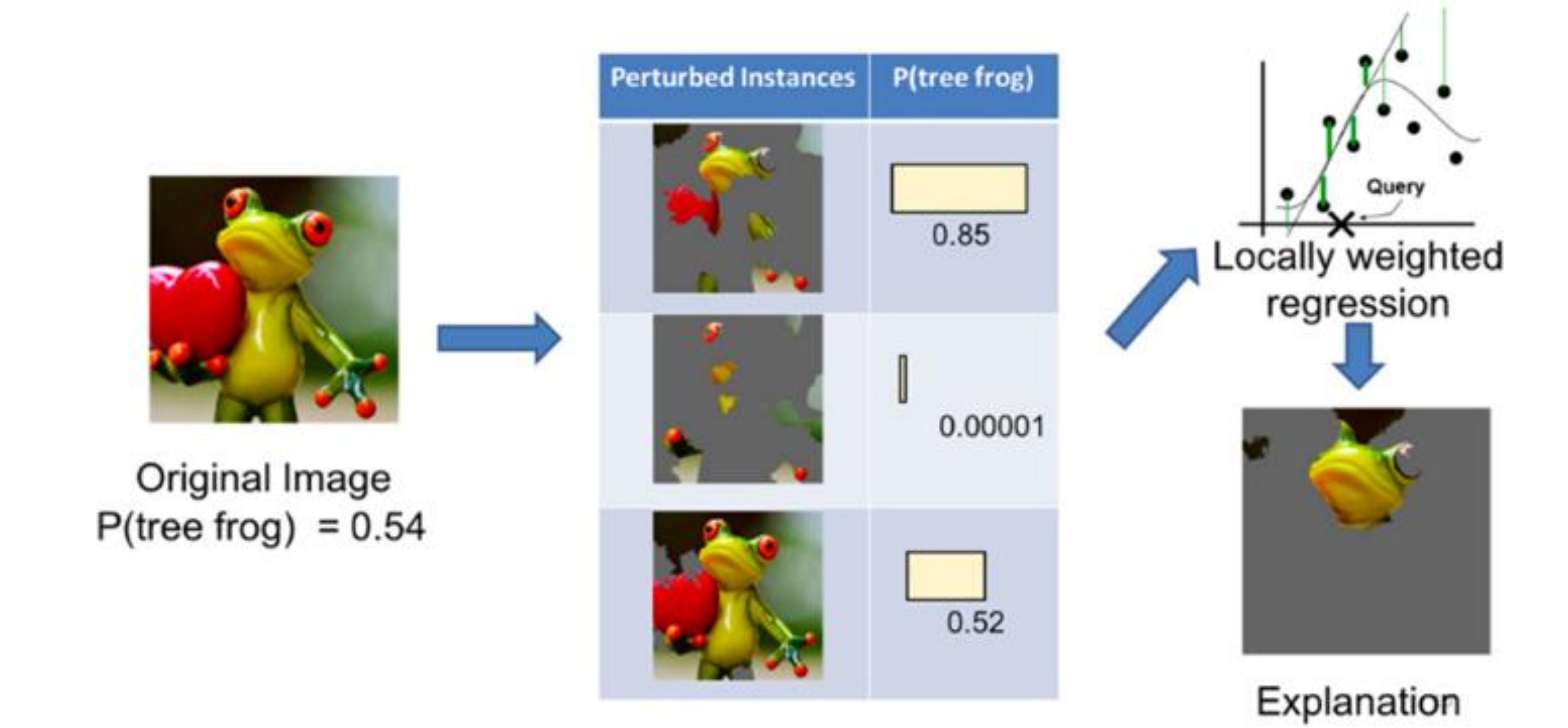
Post-hoc interpretability
- Visualizing relevance



http://www.explain-ai.org/
https://arxiv.org/abs/1708.08296

Post-hoc interpretability

- Local Interpretable Model-agnostic Explanations (LIME)



https://arxiv.org/pdf/1602.04938.pdf

Explanatory power =/= Predictive power

Construct X

X causes Y by mechanism F where $F(X) = Y$

Construct Y

Measurement x

model f maps x to y via $f(x) = y$

Measurement y

Construct X

X causes Y by mechanism F where $F(X) = Y$

Construct Y

Explanation

match f to F

Measurement x

model f maps x to y via $f(x) = y$

Measurement y

Construct X

X causes Y by
mechanism F
where F(X) = Y

Construct Y

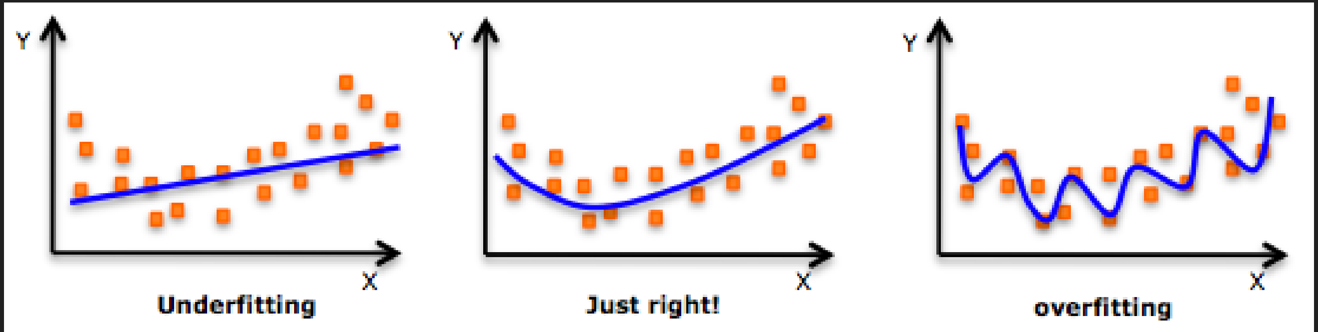
Prediction
use f to generate y

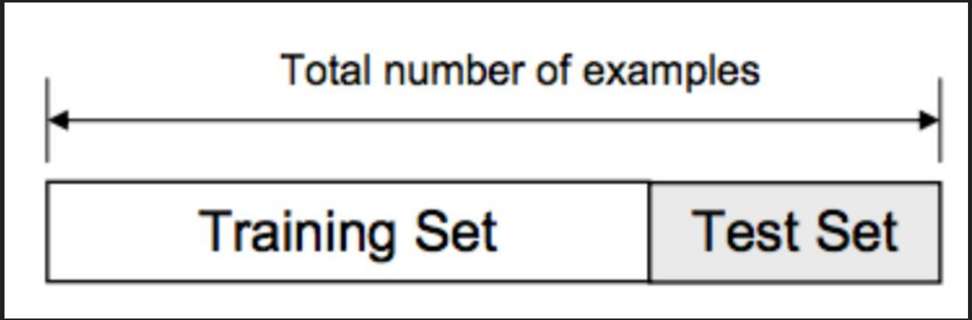
Measurement x

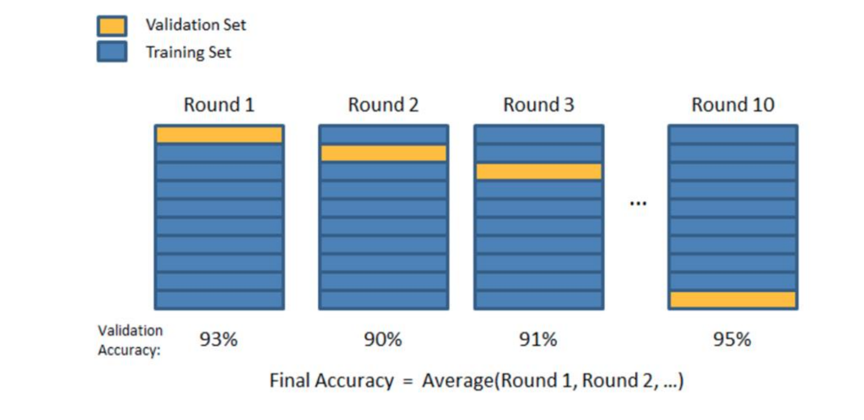
model f maps x to y
via f(x) = y

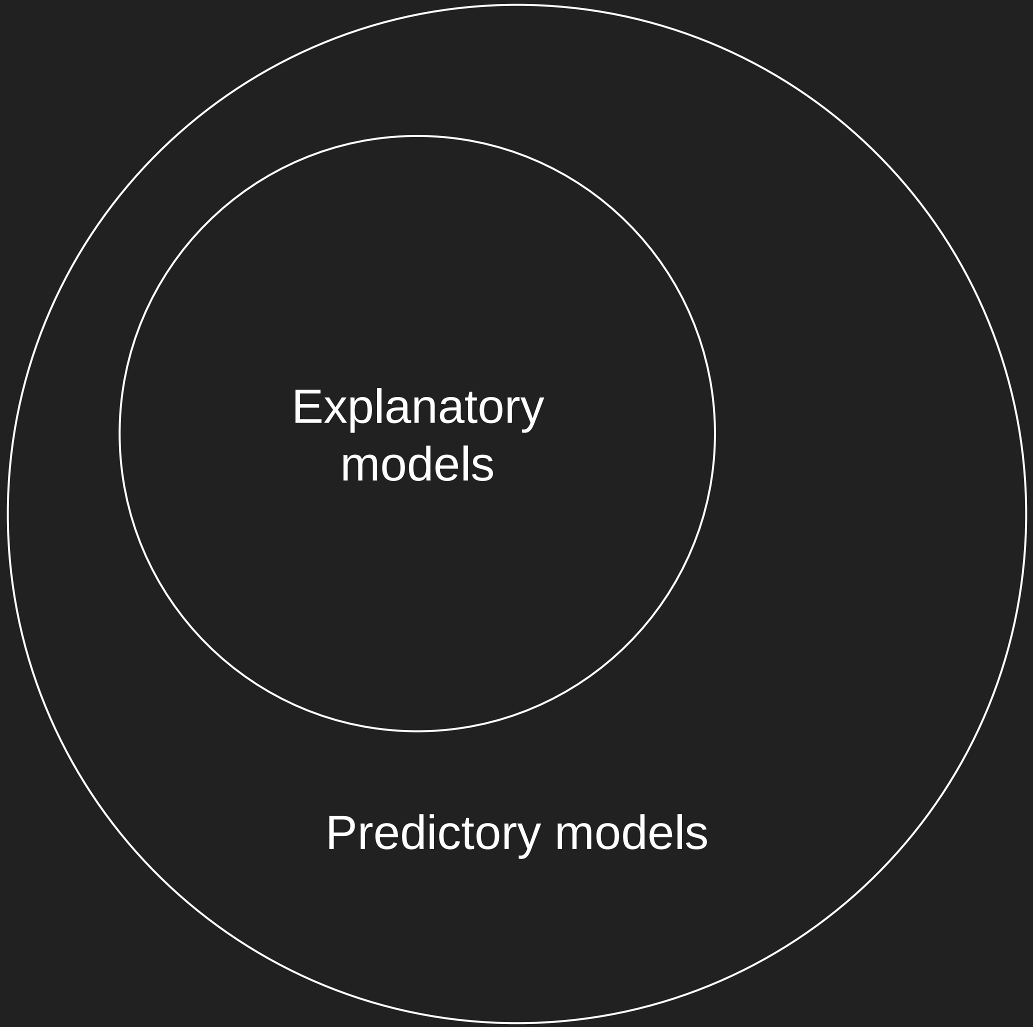
Measurement y

Predictive modeling
- Cross validation








Explanatory models
Predictory models

Measurements are not accurate representations of their underlying constructs

Results in difference between prediction vs. explanation

$$
\begin{aligned}
\mathrm{EPE} &= \overbrace{E\{Y-\hat{f}(x)\}^2}^{\text{Reduce overall error} \to \text{Prediction goal}} \\
&= E\{Y-f(x)\}^2 + \{E(\hat{f}(x)) - f(x)\}^2 \\
&\quad + E\{\hat{f}(x) - E(\hat{f}(x))\}^2 \\
&= \mathrm{Var}(Y) + \underbrace{\mathrm{Bias}^2}_{\text{Reduce bias} \to \text{Explanation goal}} + \mathrm{Var}(\hat{f}(x)).
\end{aligned}
$$

https://arxiv.org/pdf/1101.0891.pdf

“A good model”

Bias = 0, while exhibiting minimal overall error

$$\underbrace{E\left(y_0 - \hat{f}(x_0)\right)^2}_{\text{minimal}} = \mathrm{Var}(\hat{f}(x_0)) + \underbrace{[\mathrm{Bias}(\hat{f}(x_0))]^2}_{=0} + \mathrm{Var}(\epsilon).$$

It is possible to reduce variance by increasing bias
→ And still resulting in reduced overall error

Loss of model explanatory power → Increased predictive power

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$

https://arxiv.org/pdf/1101.0891.pdf

Explanatory power

(Given theory, how does the sample data fit?)

Two dimensions

Predictive power

(How does model perform in out-samples?)

Good experimental design

(Randomized control trials
→ Remove confounding)

+

In-sample model fitting

→ Test hypothesis

Observational data

(Complex interactions that are difficult to hypothesize or measure in isolation)

+

Out-sample, predictive modeling

Allows modeling non-linear relationships

Table 1. Differences Between Explanatory Statistical Modeling and Predictive Analytics

Step	Explanatory	Predictive
Analysis Goal	Explanatory statistical models are used for testing causal hypotheses.	Predictive models are used for predicting new observations and assessing predictability levels.
Variables of Interest	Operationalized variables are used only as instruments to study the underlying conceptual constructs and the relationships between them.	The observed, measurable variables are the focus.
Model Building Optimized Function	In explanatory modeling the focus is on minimizing model bias. Main risks are type I and II errors.	In predictive modeling the focus is on minimizing the combined bias and variance. The main risk is over-fitting.
Model Building Constraints	Empirical model must be interpretable, must support statistical testing of the hypotheses of interest, must adhere to theoretical model (e.g., in terms of form, variables, specification).	Must use variables that are available at time of model deployment.
Model Evaluation	Explanatory power is measured by strength-of-fit measures and tests (e.g., R^2 and statistical significance of coefficients).	Predictive power is measured by accuracy of out-of-sample predictions.

https://ai.arizona.edu/sites/ai/files/MIS611D/shmueli-2011-predictiveanalytics-is-research.pdf

Explanatory modeling

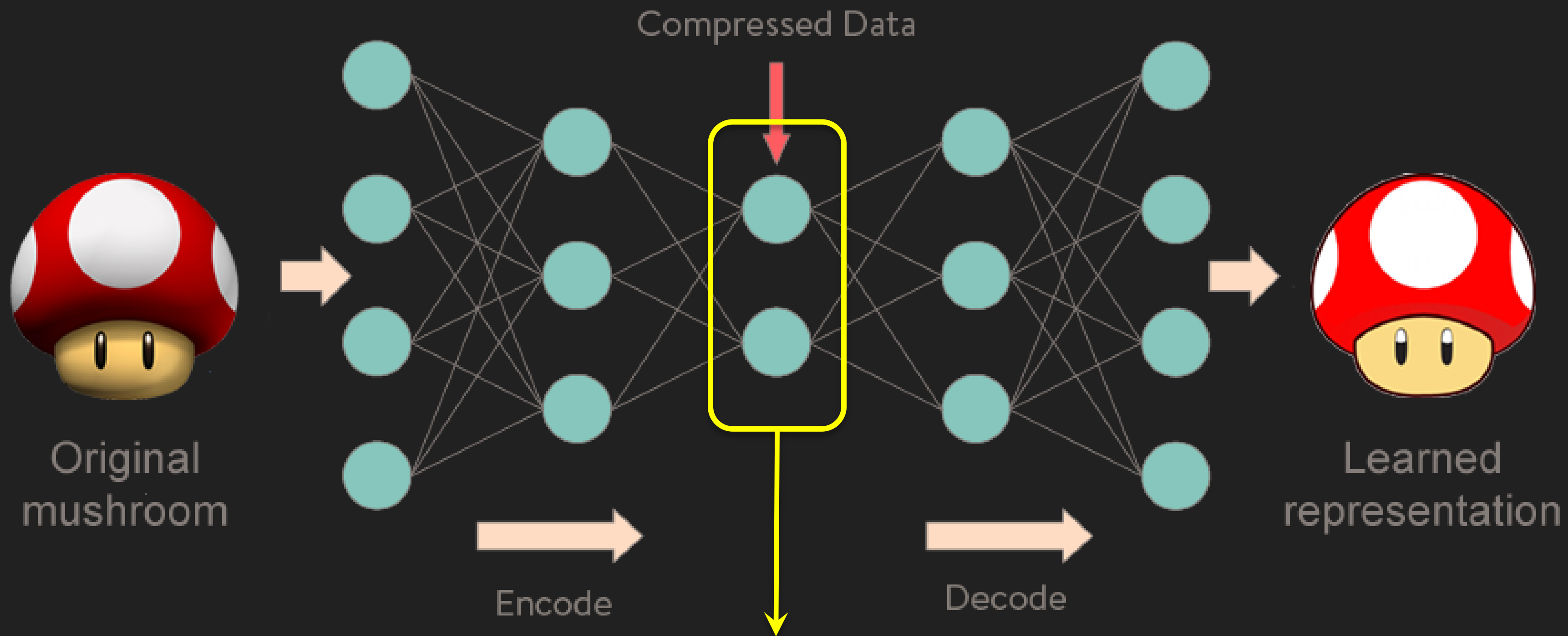
- What to act/intervene on?
- A/B testing

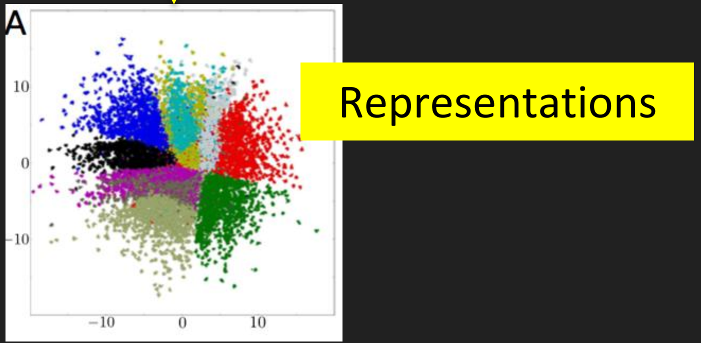
Predictory modeling

- What will happen?
- Early markers, Pre-selection

Descriptive modeling

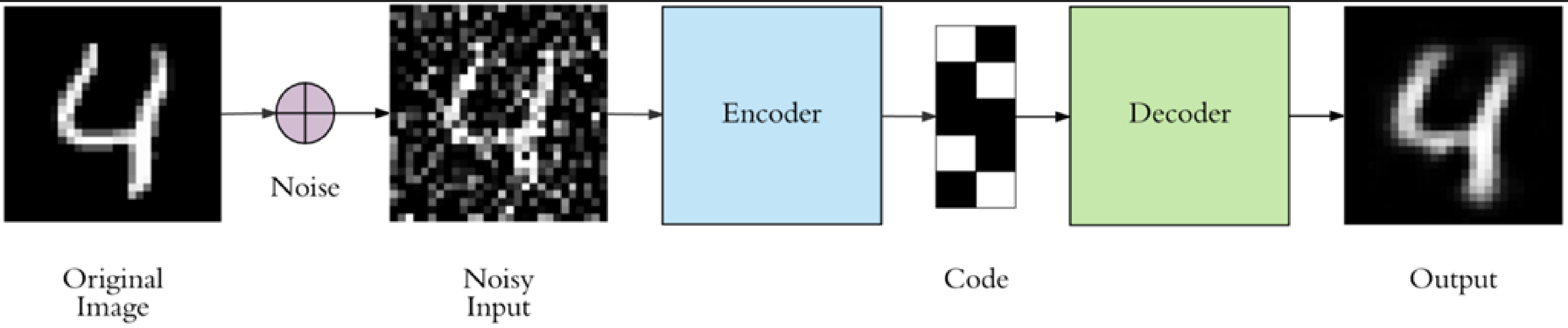
- summarizing or representing the data structure in a compact manner
- Learning useful representations

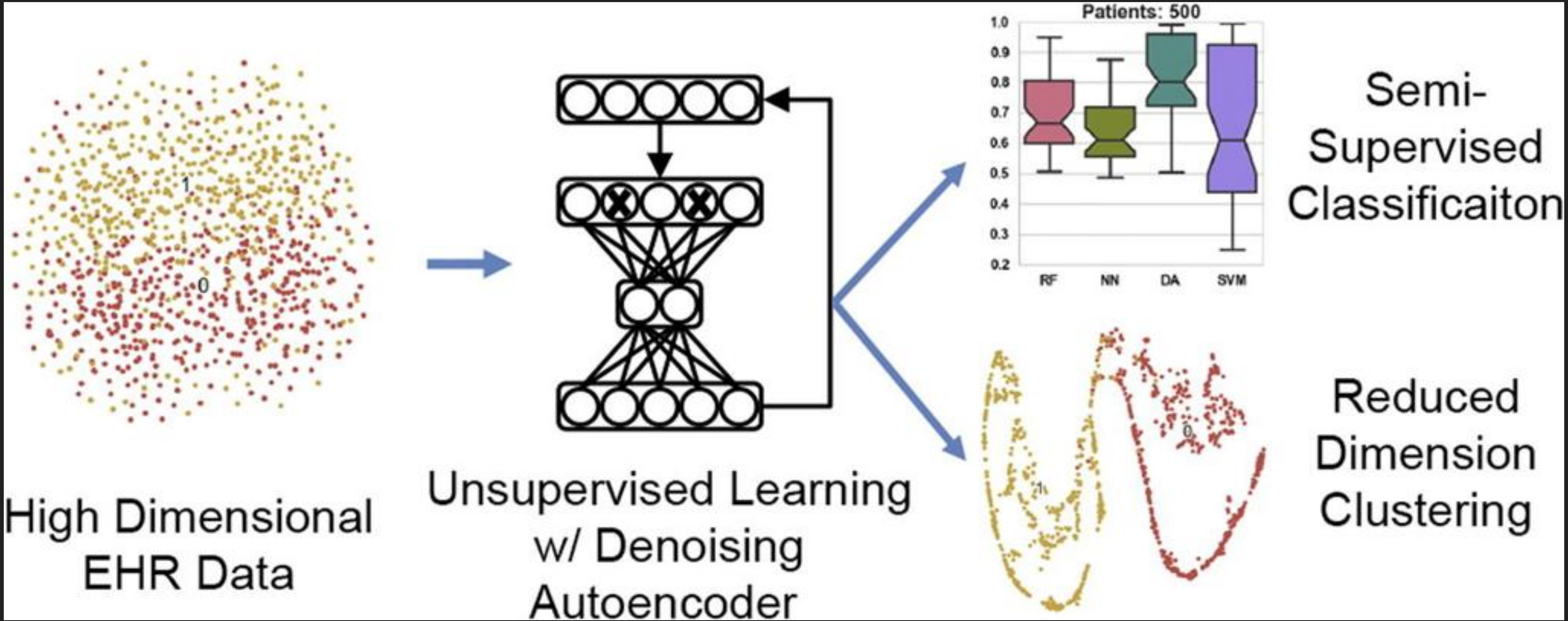




Autoencoders

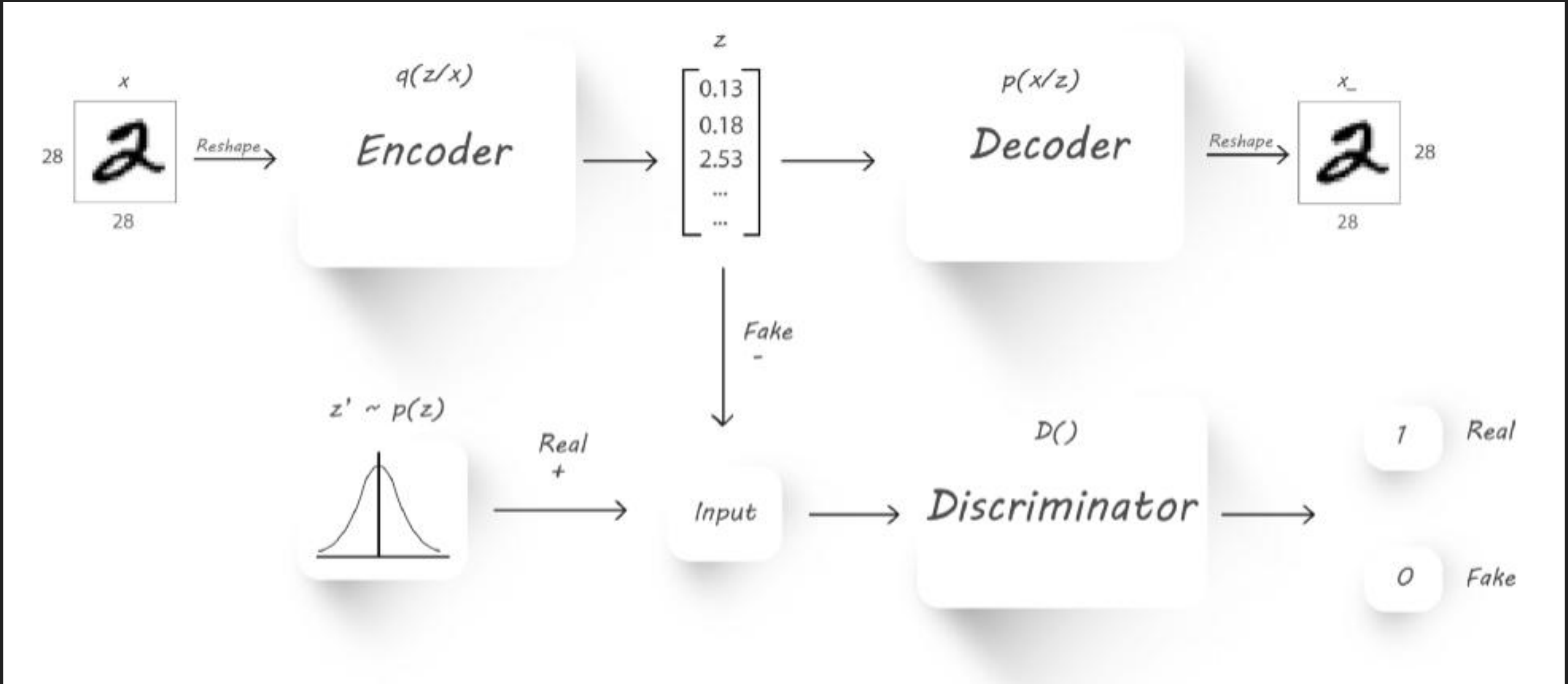
for learning useful representations


Noise
Encoder
Decoder
Original
Image
Noisy
Input
Code
Output

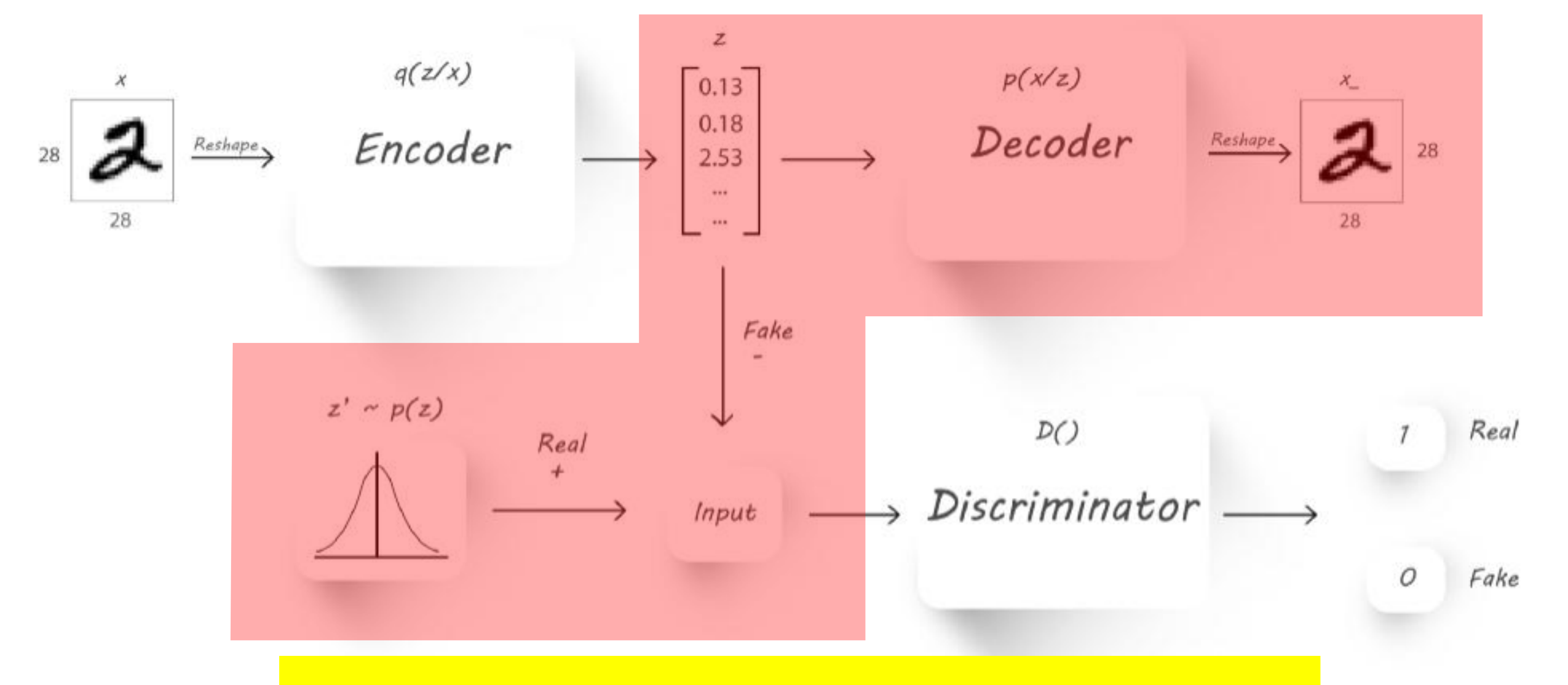


Semi-supervised learning of the electronic health record for phenotype stratification

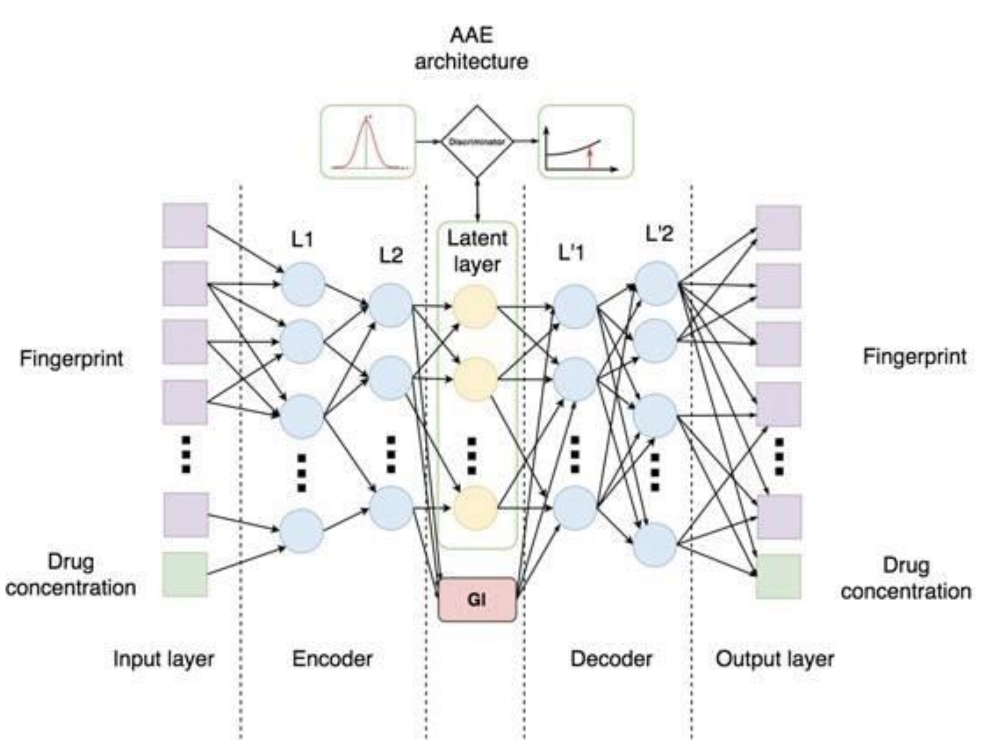
Brett K. Beaulieu-Jones [a,b], Casey S. Greene [b,c,d,*], the Pooled Resource Open-Access ALS Clinical Trials Consortium [1]



https://towardsdatascience.com/a-wizards-guide-to-adversarial-autoencoders-part-2-exploring-latent-space-with-adversarial-2d53a6f8a4f9



A model that can generative new samples



Architecture of Adversarial Autoencoder (AAE) used in this study

Oncotarget. 2017 Feb 14;8(7):10883-10890. doi: 10.18632/oncotarget.14073.

The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology.

Kadurin A[1,2,3,4], Aliper A[2], Kazennov A[2,5], Mamoshina P[2,6], Vanhaelen Q[2], Khrabrov K[1], Zhavoronkov A[2,7,5].

Feature	Description
Mean	Mean.
Var	Variance.
ACF1	First order of autocorrelation.
Trend	Strength of trend.
Linearity	Strength of linearity.
Curvature	Strength of curvature
Season	Strength of seasonality.
Peak	Strength of peaks.
Trough	Strength of trough.
Entropy	Spectral entropy.
Lumpiness	Changing variance in remainder.
Spikiness	Strength of spikiness
Lshift	Level shift using rolling window.
Vchange	Variance change.
Fspots	Flat spots using disretization.
Cpoints	The number of crossing points.
KLscore	Kullback-Leibler score.
Change.idx	Index of the maximum KL score.

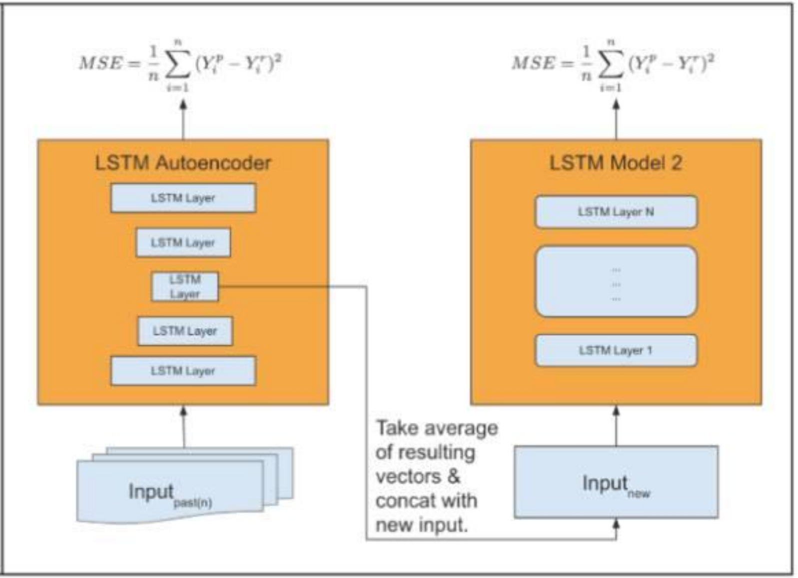
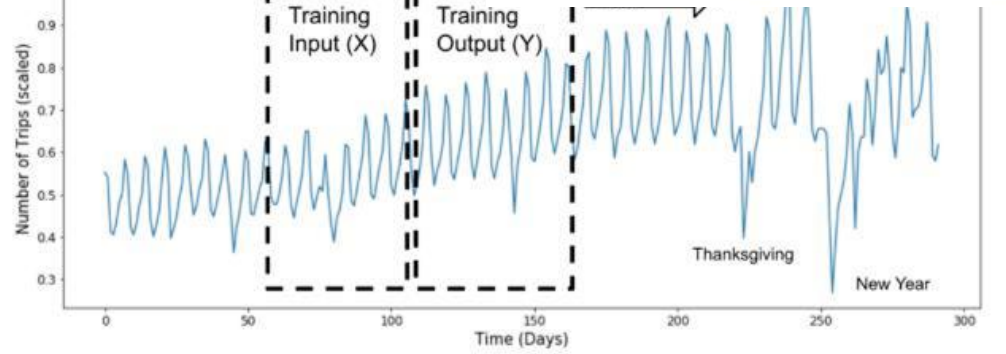


Figure 4: Our model is composed of manually derived time series features (left) and our proposed LSTM architecture with an automatic feature extraction model (right).[5]



https://eng.uber.com/neural-networks/

然而，問題點中了
實作不一定要很複雜
而效果仍可很好

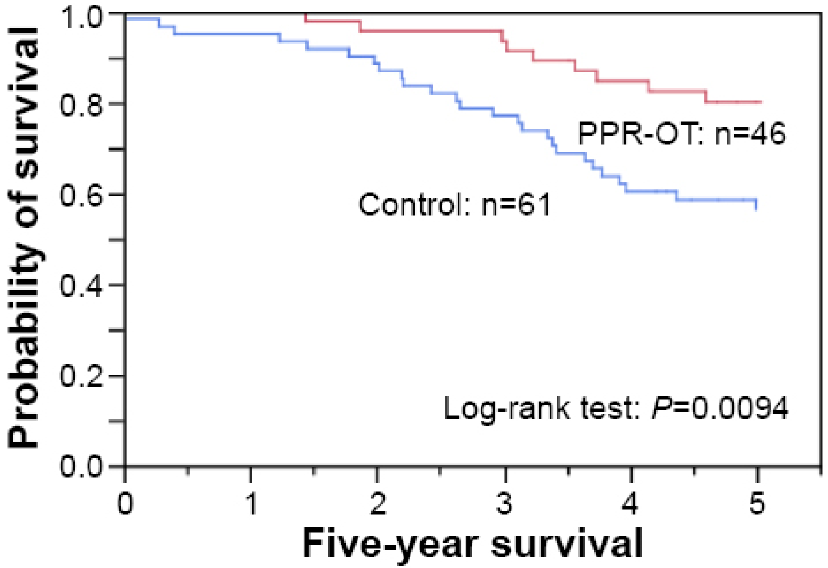


Figure 3 Effect of the personalized pulmonary rehabilitation program that included occupational therapy (PPR-OT) on the 5-year survival (all-cause mortality) of patients with COPD after CPET in the retrospective study.

Abbreviations: COPD, chronic obstructive pulmonary disease; CPET, cardiopulmonary exercise testing.

肺部復健
有助於改善慢性阻塞性
肺疾病患者的存活率

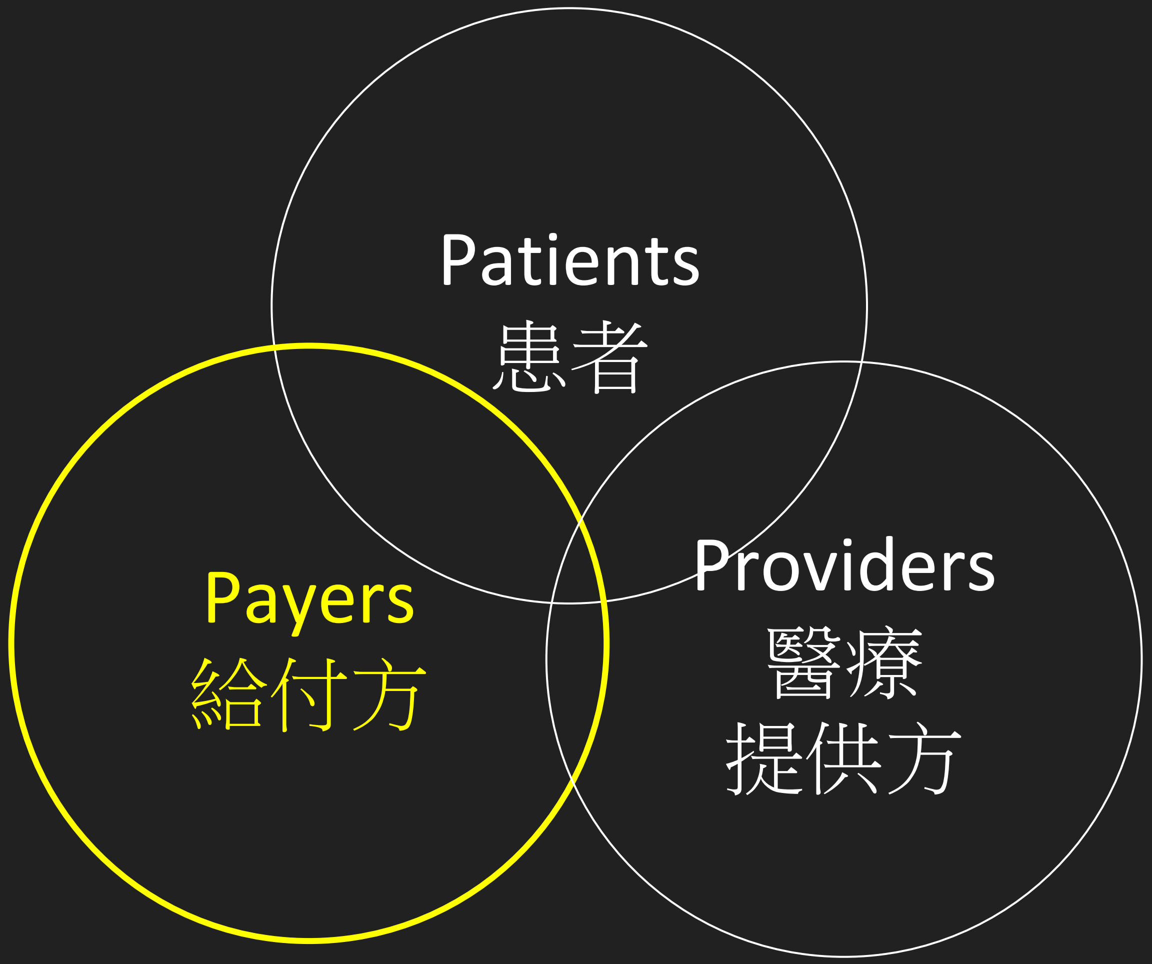
https://doi.org/10.2147/COPD.S86455

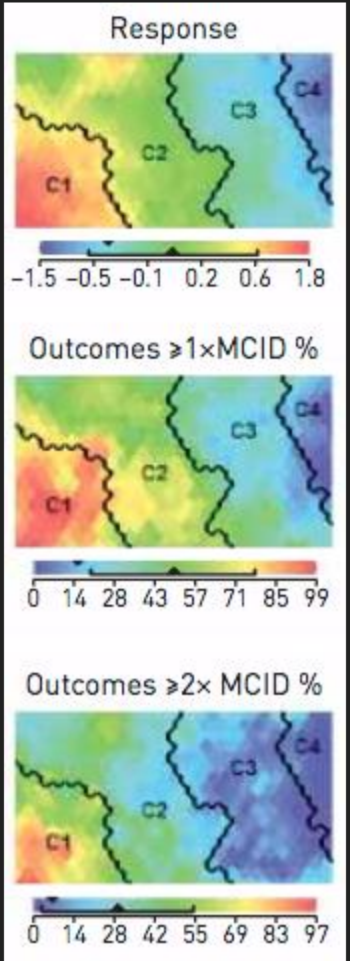
不同量表不盡相同，如何衡量成效？

喘的程度？

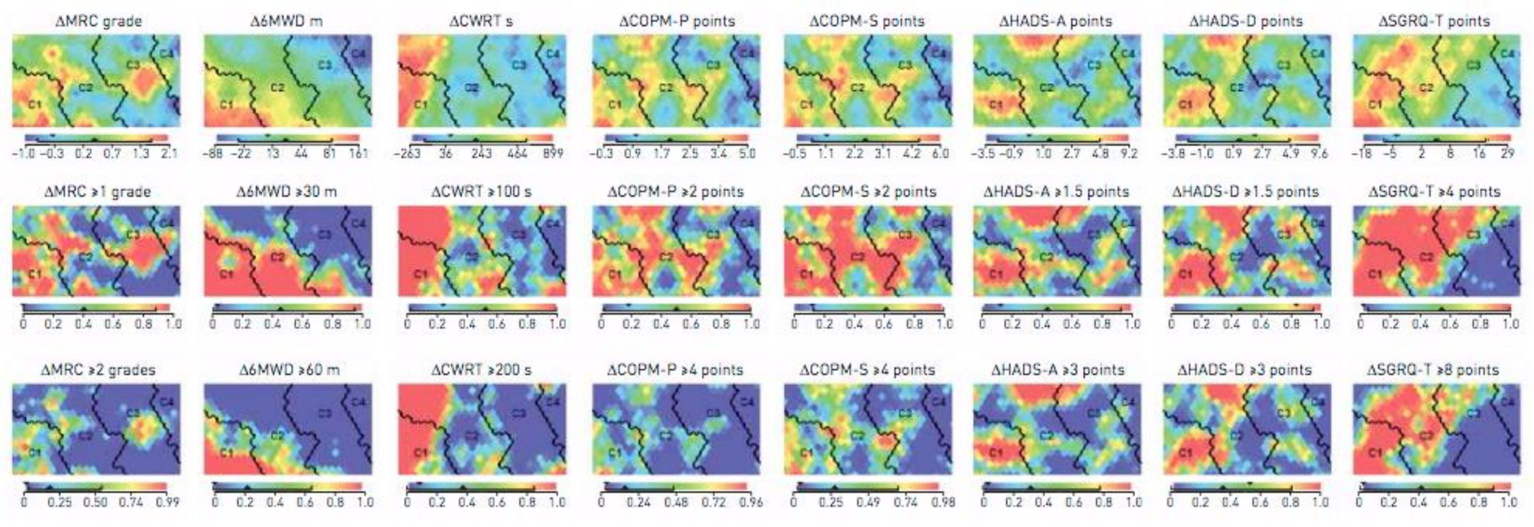
累的程度？

Symptom	Type and Name of Measure
Dyspnea	Short-term
	Borg
	VAS
	Situational
	MRC
	BDI
	SOBQ
	Impact
	CRQ (dyspnea subscale)
	PFSDQ
	PFSDQ-M (dyspnea subscale)
Fatigue	Short-term
	Borg
	VAS
	Impact
	CRQ (fatigue subscale)
	PFSDQ-M (fatigue subscale)
	FACIT-fatigue
	MFI
	CIS
Multiple symptoms	CAT
	SGRQ symptoms domain


Patients
患者
Payers
給付方
Providers
醫療
提供方



Profiling differential response to pulmonary rehabilitation in COPD by self-organizing maps



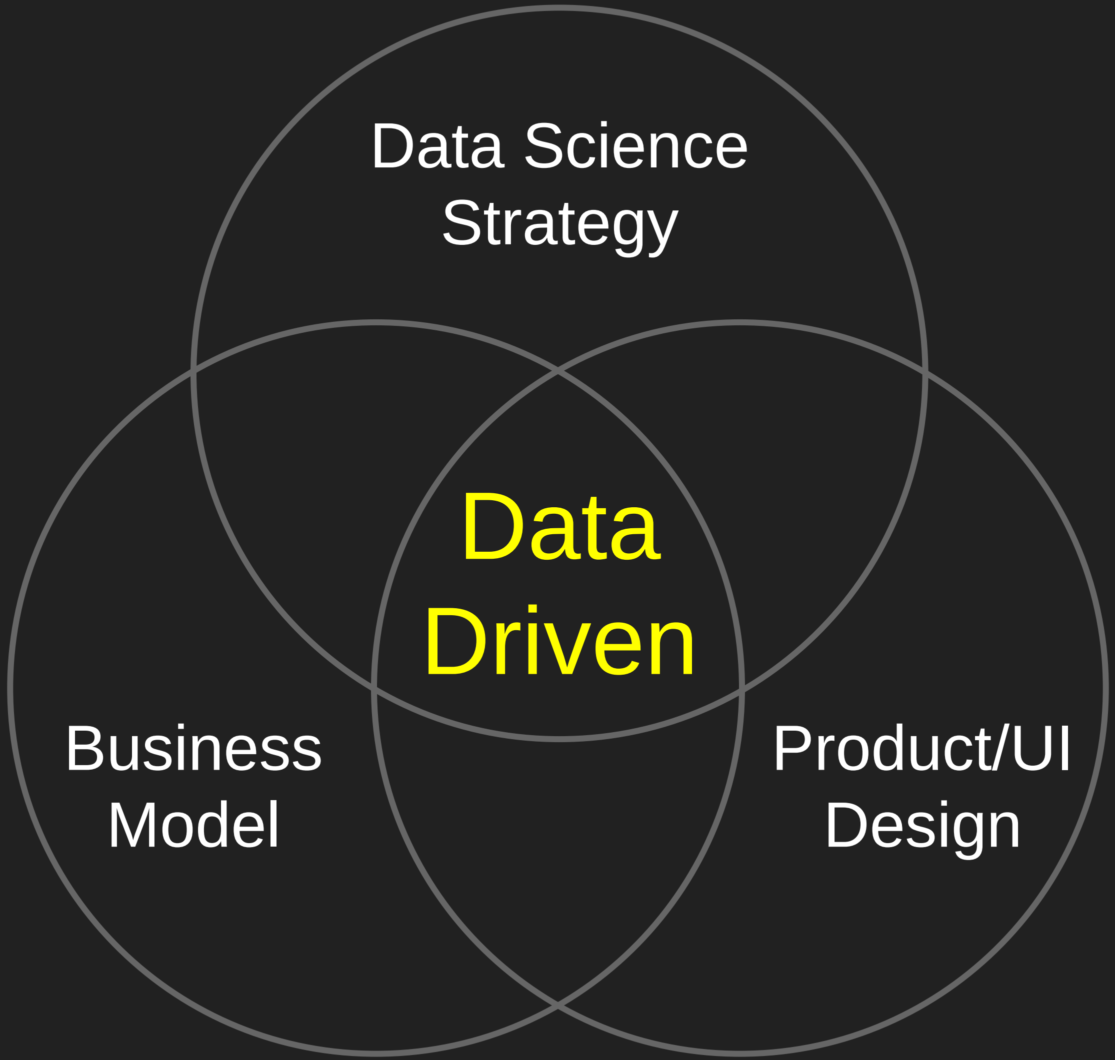
Eur Respir J 2015; 46: 1625–1635

預後/風險/安全性受重視程度
>> 純方便/使用者體驗

Algorithms are just means to synthesize information

But a good question with a relevant setting

is still a human task itself


Data Science Strategy
Business Model
Product/UI Design
Data Driven

Take home message

資料產出是策略性的，應直屬於決策單位

組織的資料力可以由 Data workflow 上的障礙來衡量

善用 UI design 讓使用者自我揭露 >> 事後建立複雜model

網絡/關聯性的資料具有戰略意義

三種角色的安排 : Data PM, Data Engineer, Data Scientist

兩類 Interpretability : Transparency, Post-hoc Interpretability

三種 modeling : Explanatory, Predictatory, Descriptive

Deep neural networks for representation learning

3P: Payer-Provider-Patients

預後/風險/安全性 > 方便/使用者體驗

Data driven = Business model + Product design + Data strategy