Essence of Machine Learning (and Deep Learning)

Hoa M. Le
Data Science Lab, HUST
hoamle.github.io

1

Examples

- https://www.youtube.com/watch?v=BmkA1ZsG2P4
- http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

2

Machine Learning is about …

… a computer program (machine) *learns* to do a task (problem) from experience (data)

- *learning* ≜ improved *performance* with more experience

- Tom Mitchell

⇑

predictive modelling with **sample data**

⇑

"heurestics" & statistical modelling

note 1: "heurestic" as in "intuitive, but not (yet!) rigorously proven by mathematical tools at some extend"

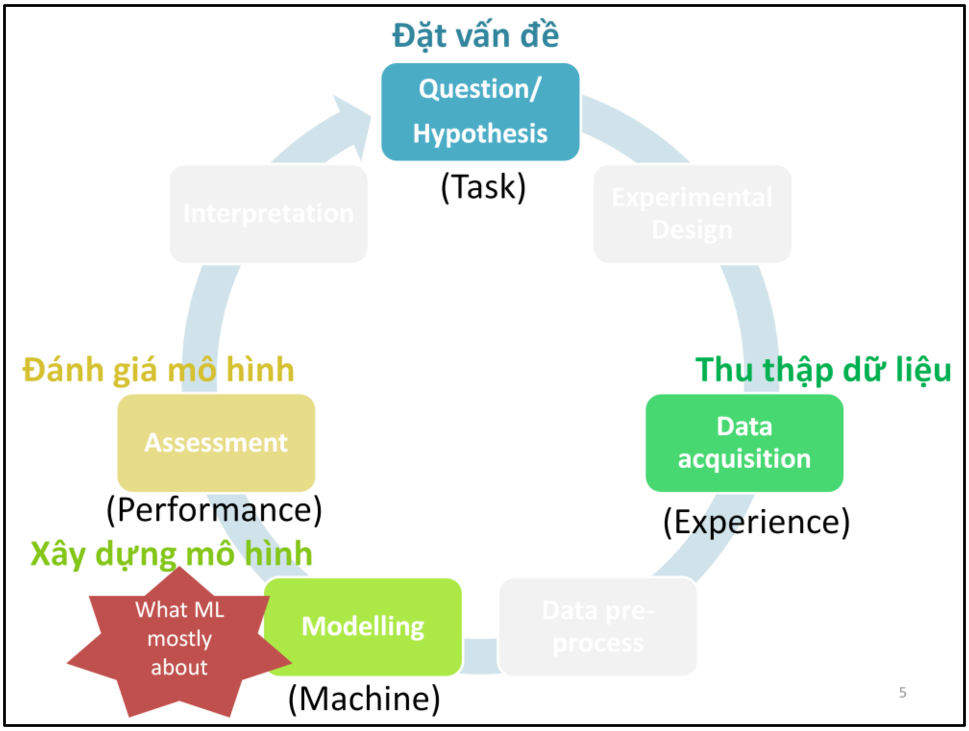
note 2: predictive modelling can also be in the form of rule-based systems, models in physics, etc

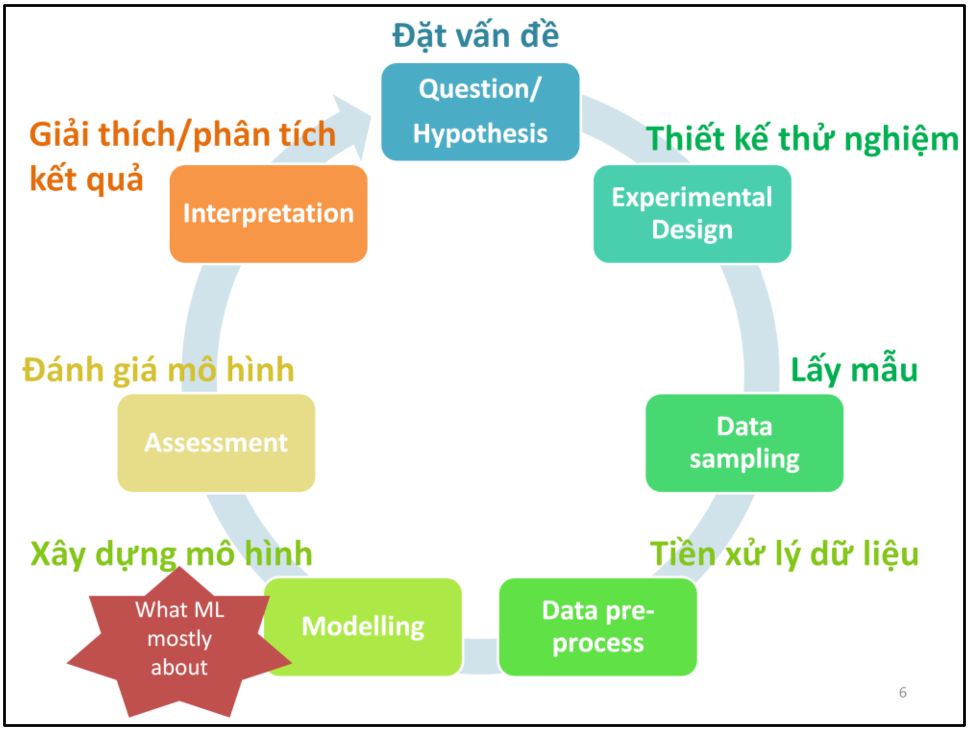
3

BUILD
A MACHINE LEARNING SOLUTION

the Pipeline

4


Đặt vấn đề
Question/
Hypothesis
(Task)
Experimental
Design
Thu thập dữ liệu
Data
acquisition
(Experience)
Data pre-
process
Modelling
(Machine)
Xây dựng mô hình
What ML
mostly
about
Đánh giá mô hình
Assessment
(Performance)
Interpretation
5



Data sampling and Result interpretation are often neglected in a ML 101 course for brevity. However, remember that they are NOT disposable. They are also arguably 2 most important steps in the pipeline.

Đặt vấn đề

Question/ Hypothesis

Q.a. What are there in an **abitrary photo?**
Q.b. What is there in an **abitrary photo?**
Q.c. Is there any puppy an **abitrary photo?**

cat
flower
dog
jet
ground
grass
...

Other questions:
- *Where are the puppies in a photo?*
- *How confident can I assure that there is a cat a photo?*
- *For what reasons can I know that there is a cat in a photo?*

7

a. Multi-label Multi-class classification
b. Standard Multi-class classification (single-label)
c. Binary classification (single-class)

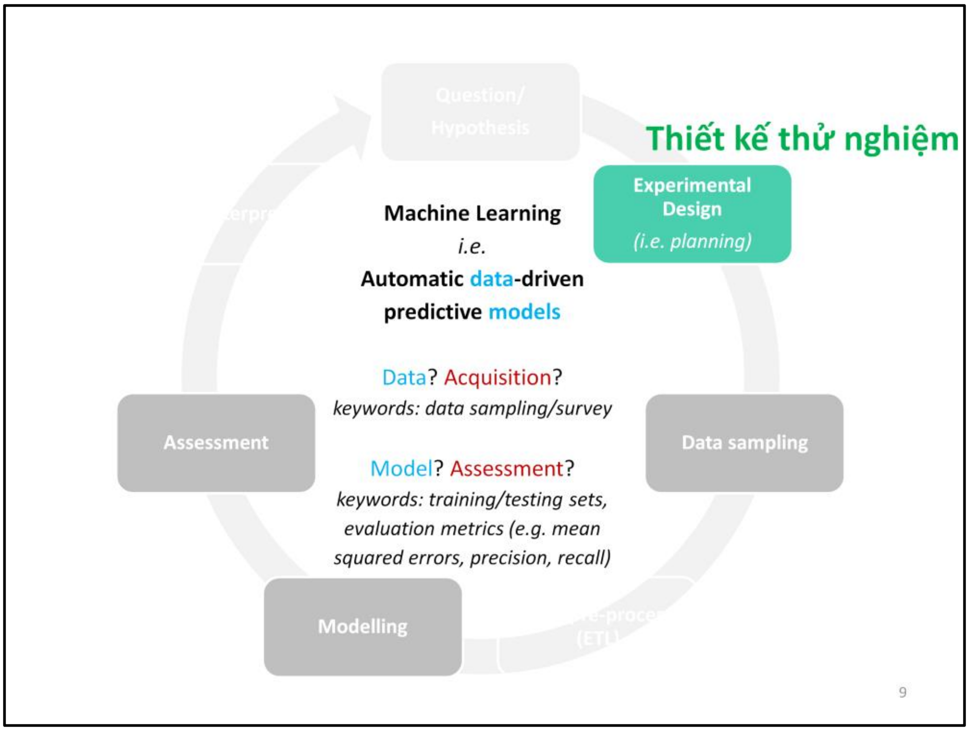
Thiết kế thử nghiệm

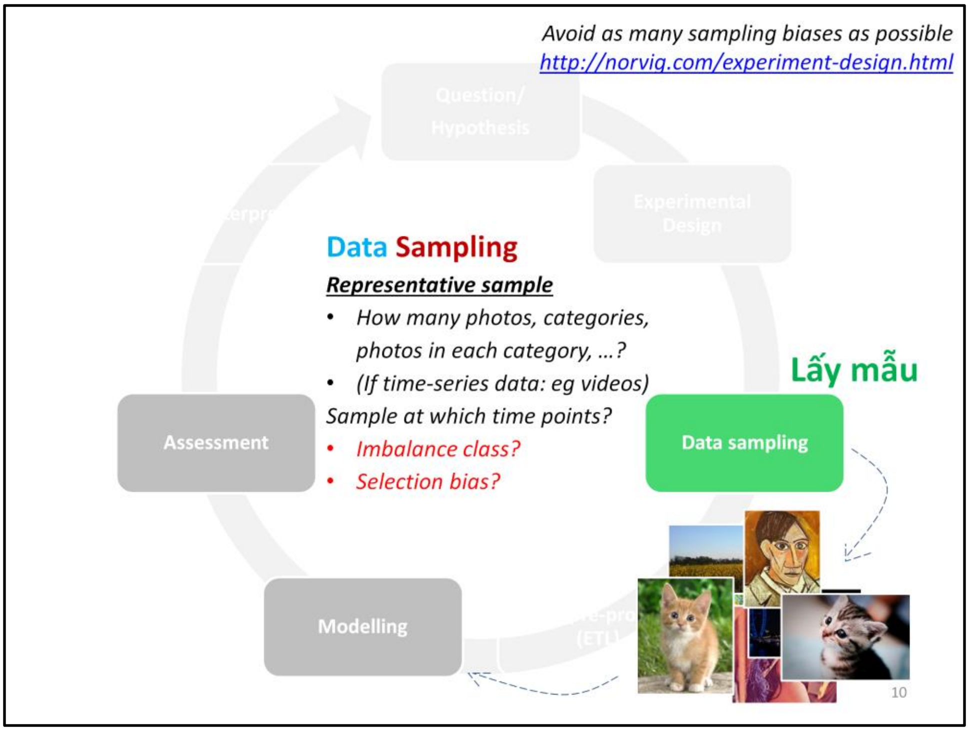
Experimental Design
(i.e. planning)

Machine Learning
i.e.
Automatic data-driven predictive models

Data? Acquisition?
keywords: data sampling/survey

Model? Assessment?
keywords: training/testing sets, mean squared errors, precision, recall, …


Question/
Hypothesis
Thiết kế thử nghiệm
Experimental
Design
(i.e. planning)
Machine Learning
i.e.
Automatic data-driven
predictive models
Data? Acquisition?
keywords: data sampling/survey
Model? Assessment?
keywords: training/testing sets,
evaluation metrics (e.g. mean
squared errors, precision, recall)
Assessment
Data sampling
Modelling
9



<u>How representative the sample is => how generalisable the model is</u>

Rule-of-thumbs: “garbage in, garbage out”

Misconception: “We are in big data era, data sampling is no longer a concern”
- Big data ≠ sufficient data. Many domains (e.g. biomedicine, social sciences) typically have small/tiny sample size in most of their problems. Even in computer vision, there exist problems that do not readily have big data, e.g. humour detection, lip-reading.

Which metrics to use depend on which problem
http://scikit-learn.org/stable/modules/model_evaluation.html

Model Assessment

Evaluation metrics

- *Accuracy*
- *Precision, Recall*
- *Area Under Curve (AUC)*
- *Mean squared errors (MSE)*
- *…*

(If hypothesis testing problem)

- *t-statistic, z-statistic, χ^2-statistic, …*

Đánh giá mô hình

Assessment

Data sampling

Modelling

```
cat
flower
dog
jet
ground
grass
```

11

If training/testing set split is well designed with sufficient examples, we might not need to repeat many experiments.

Model Assessment

Evaluation setup

Evaluation (i.e.report results) on *unseen* data

- Training/testing set split: follows data sampling principles
- Repeat experiment: gives measurable confidence to the reported results

Đánh giá mô hình

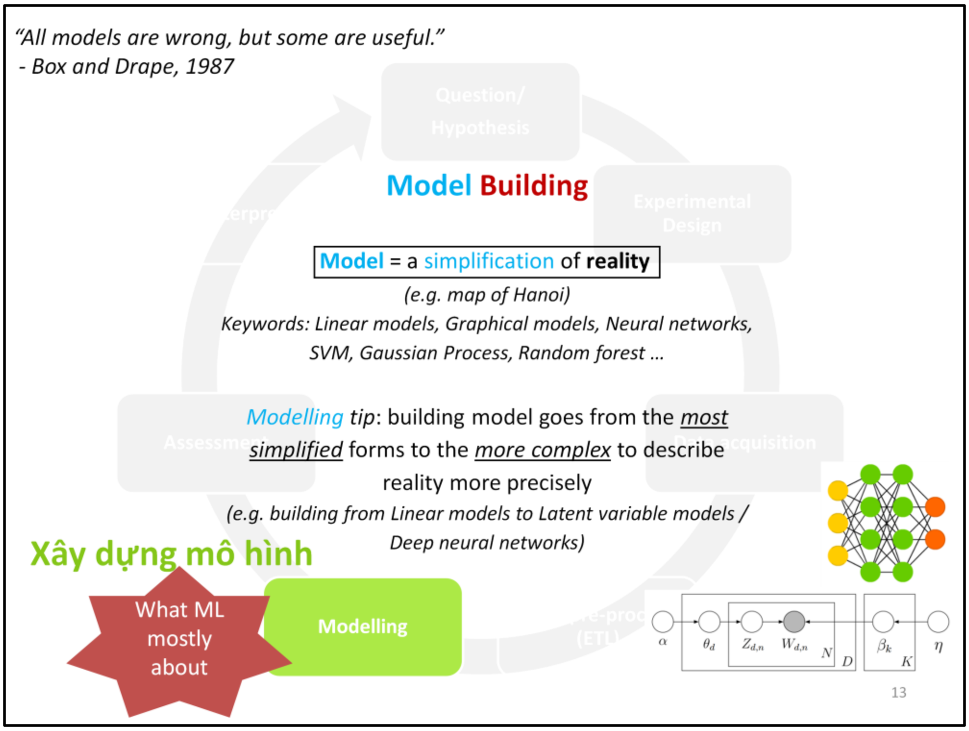
Assessment

Data sampling

Modelling

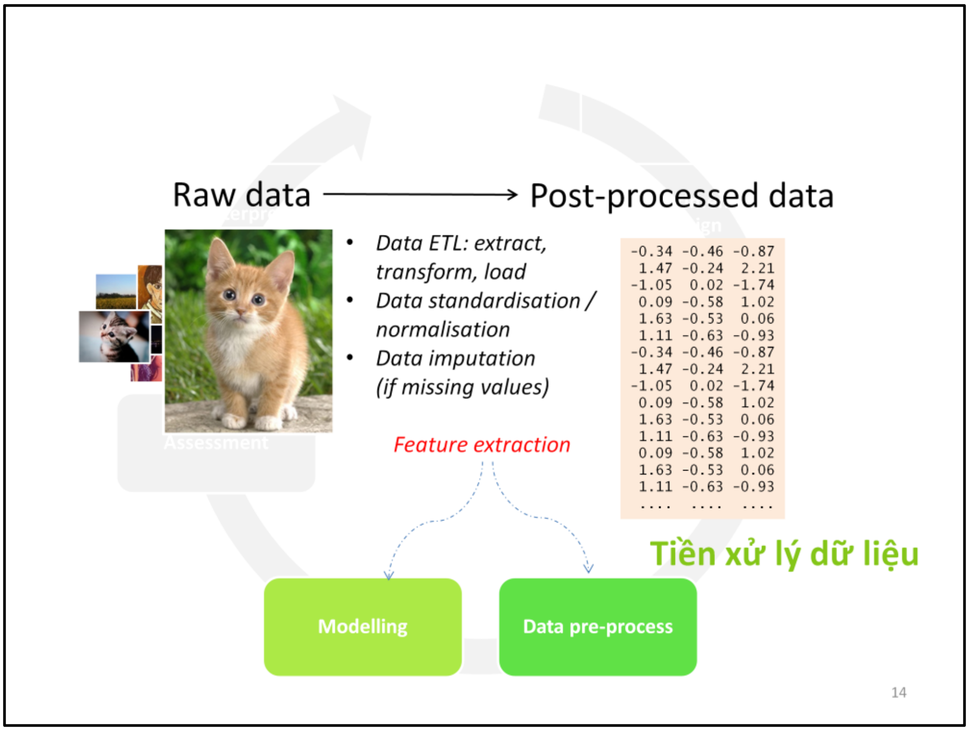
```
cat
flower
dog
jet
ground
grass
```

12

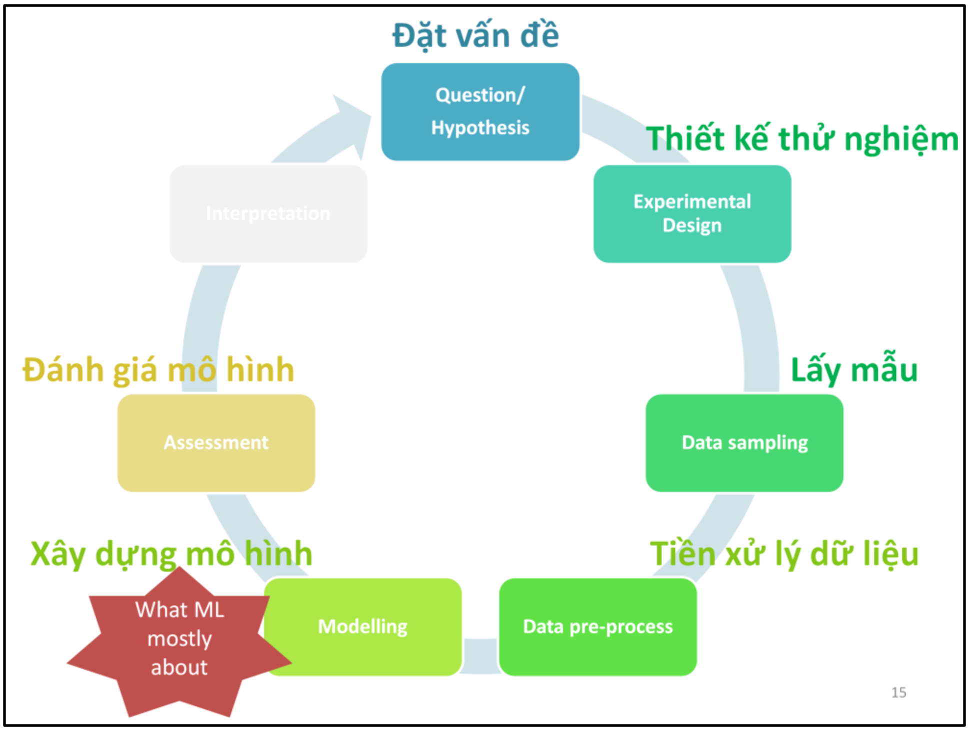


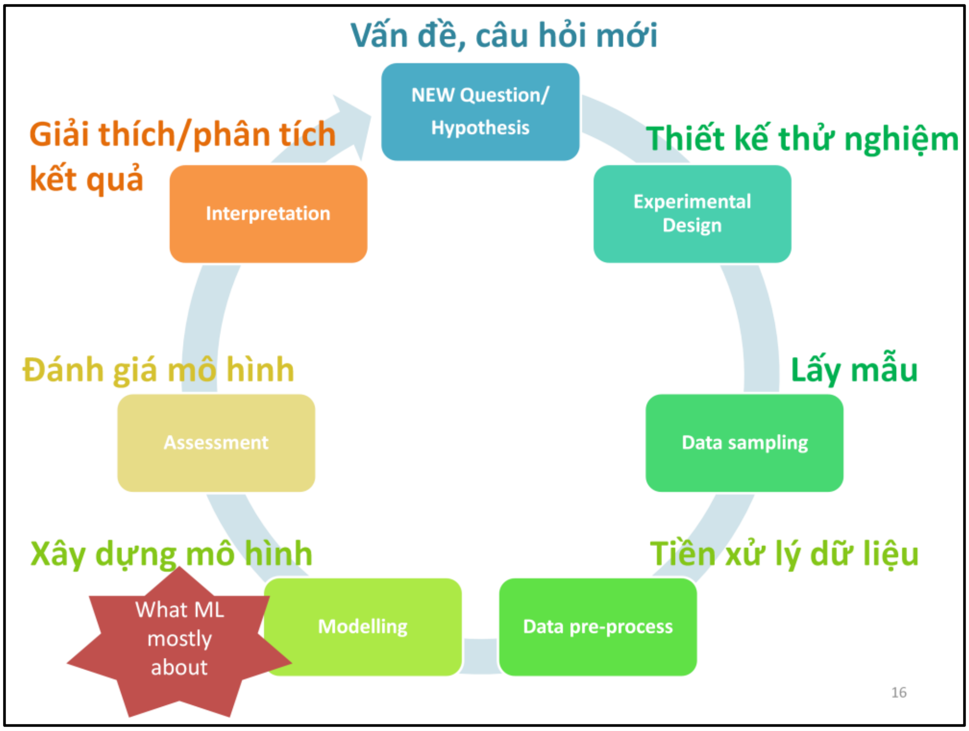
Graphics:

- http://www.asimovinstitute.org/neural-network-zoo/
- LDA (Blei's KDD 2011 tutorial)



The idea of Deep Learning is to incorporate feature extraction stage into the model, for which how the features are extracted is also *learnt from the data*.


Đặt vấn đề
Question/ Hypothesis
Thiết kế thử nghiệm
Experimental Design
Lấy mẫu
Data sampling
Tiền xử lý dữ liệu
Data pre-process
Xây dựng mô hình
Modelling
What ML mostly about
Đánh giá mô hình
Assessment
Interpretation
15

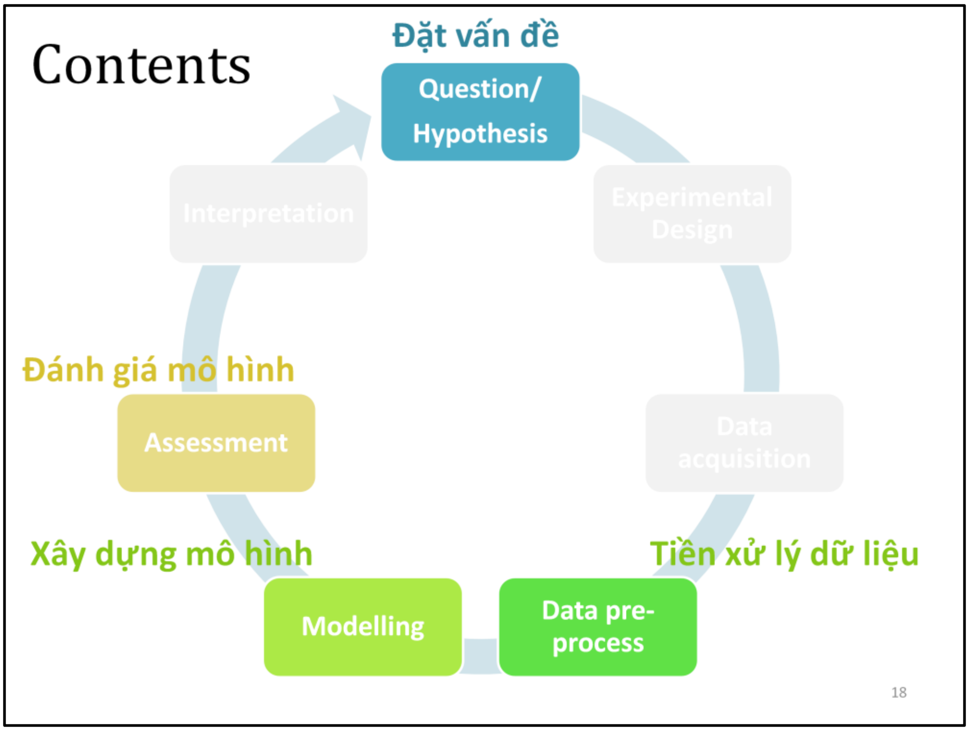


The most valuable outputs from interpreting/analysing the results are better/new insights to the current problem, which motivates further improvements for that problem of interest, and/or novel approaches to related problems/domains.
⇒ driving force for developments

PRINCIPLES OF MODELLING

Statistical reasoning (*)

*(*) A machine learning algorithm does not necessarily have a probabilistic interpretation, or developed from a statistical framework. Nevertheless, statistical reasoning provides a rigorous mathematical tool for estimation and inference to make optimal decision (e.g. prediction, action) under **uncertainty,** which is one of the ultimate objectives in ML.*


Contents
Đặt vấn đề
Question/ Hypothesis
Experimental Design
Data acquisition
Tiền xử lý dữ liệu
Data pre-process
Xây dựng mô hình
Modelling
Đánh giá mô hình
Assessment
Interpretation
18

ML problem: Classification

Question **Is there any cat in an abitrary photo?**

Experience: dataset of {image, label} pairs $\mathcal{D} = \{x_n, y_n\}_{n=1}^{N}$

Modelling predict $\hat{y}_n$ – *cat existence* – *given arbitrary* x_n

Cat?
Not cat?

supervised learning

Image
x_n
$\mathbb{N}^{400\times600\times3}$

Prediction
$\hat{y}_n$
$\{\text{True}, \text{False}\}$

(single-class)

binary classification problem

Assessment

Accuracy = $\frac{1}{N}\sum_n \mathbb{I}(\hat{y}_n = y_n)$

Precision, Recall, F1-score

Area Under Curve (AUC)

...

Example models:
Logistic regression (linear model)
Neural Net with sigmoid output (nonlinear model)

ML problem: Classification

Question — **What is there in an abitrary photo?**

Experience: dataset of {image, label} pairs $\mathcal{D} = \{x_n, y_n\}_{n=1}^{N}$

Modelling — predict $\hat{y}_n$ – *object identity* – *given arbitrary* x_n

cat
flower
dog
jet
ground
grass

Image
x_n
$\mathbb{N}^{400\times600\times3}$

Prediction
$\hat{y}_n$
$\{1,2,3,4,5,6\}$

supervised learning

(multi-class)

categorical classification problem

Assessment

Accuracy = $\frac{1}{N}\sum_n \mathbb{I}(\hat{y}_n = y_n)$

Precision, Recall, F1-score

Area Under Curve (AUC)

...

Example models:
Softmax classification (linear model)
Neural Net with softmax output (nonlinear model)

20

ML problem: Regression

Question **How much is the price of a house given …**

Experience: dataset of {(area, location, #rooms), price} pairs $\mathcal{D} = \{x_n, y_n\}_{n=1}^{N}$

Modelling predict $\hat{y}_n$ – *house price* – *given arbitrary* x_n

Area	100m²
Location	24.7⁰N 183.0⁰E
#Rooms	3

→ \$150,000

Features/Predictors

x_n

$\mathbb{R} \times \mathbb{R}^2 \times \mathbb{N}$

Prediction

$\hat{y}_n$

$\mathbb{R}$

supervised learning

regression problem

Assessment squared_errors = $\frac{1}{N}\sum_n(\hat{y}_n - y_n)^2$

Example models/algorithms:
Linear regression (linear model)
Neural Net with linear output (nonlinear model)
Curve fitting algorithm

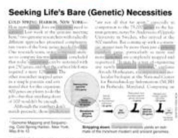
21

ML problem: Clustering

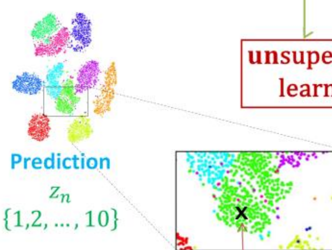
Question **What is the "topic" that a news article is talking about?**
Experience: dataset of article content *only* $\mathcal{D} = \{x_n\}_{n=1}^{N}$

Modelling predict $\boxed{z_n}$ – *"topic" (cluster) identity* – *given arbitrary* x_n

unsupervised learning


Seeking Life's Bare (Genetic) Necessities

Article (text)
x_n
$\mathbb{N}^{1500}$



Prediction
z_n
$\{1,2,\ldots,10\}$

x_n
$z_n =$ green

Assessment $\text{mean\_distance\_to\_clusters} = \frac{1}{N}\sum_n \left(x_n - \mu_{z_n}\right)^2$

Note: "topic" = group/cluster in this context, and is *<u>not</u>* *pre-defined*
We will meet the term "topic" again when visiting Topic models

Example models/algorithms:
k-means algorithm
Generative models: Mixture models, Topic models

Graphics:

- David Blei, KDD 2011
- https://lvdmaaten.github.io/tsne/examples/mnist_tsne.jpg

A ML problem can also be:

- both supervised and unsupervised *(semi-supervised)*
- combination of regression and classification sub-problems *e.g. image localisation*

Classification: C classes
Input: Image
Output: Class label
Evaluation metric: Accuracy

→ CAT

Localization:
Input: Image
Output: Box in the image (x, y, w, h)
Evaluation metric: Intersection over Union

→ (x, y, w, h)

Classification + Localization

CAT

23

Graphics: cs231n, lecture 8

Modelling

PRINCIPLES OF MODELLING

1. Model structure - constructs relationships (*stochastic and/or deterministic*) between model elements: data, parameters, and hyper-parameters.
   Keywords: graphical model
2. Learning principle - defines a framework to estimate unknown parameters (and unobserved i.e. hidden/latent variables)
   Keywords: Maximum Likelihood criterion, Bayesian inference, ++ others
3. Regularisation
   Keywords: over-fitting, Bayesian inference, ++ others
   Relevant keywords: L2-regularisation (Ridge), L1-regularisation (LASSO)

$\Rightarrow$ ALGORITHM - implements 1 + 2 + 3 to train the model
Keywords: (stochastic) gradient descent, Expectation-Maximisation (EM), Variational Inference (VI), sampling-based inference methods

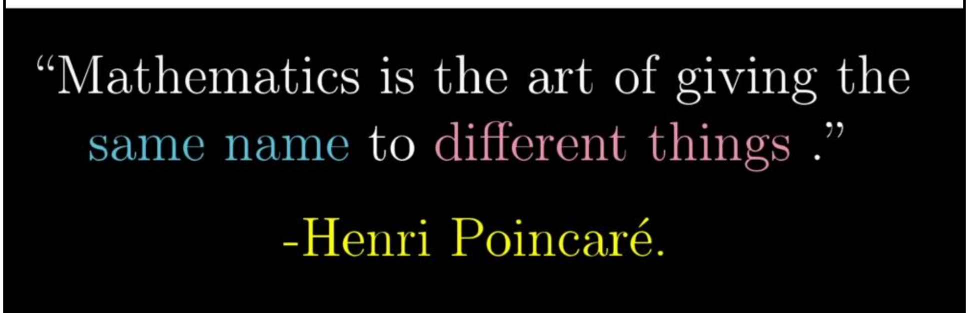
4. Model selection
   Keywords: cross-validation

24

Ref: Shakir Mohamed's Deep Learning summer school, 2016

Before we get going…

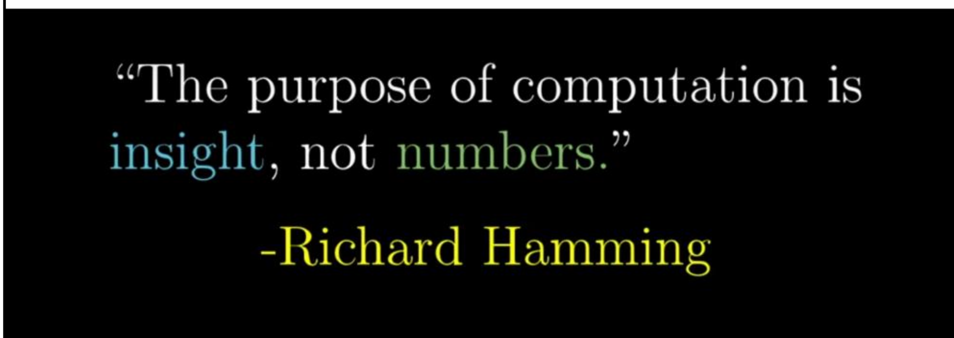
25



26

Graphics: [3Blue1Brown](https://youtu.be/P2LTAUO1TdA) https://youtu.be/P2LTAUO1TdA

Lesson learned from "modelling a real-life problem" (Lecture 1)



$$p(\mathbf{w}\mid\alpha,\beta) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \int \left(\prod_{i=1}^{k} \theta_i^{\alpha_i-1}\right)\left(\prod_{n=1}^{N} \sum_{i=1}^{k} \prod_{j=1}^{V} (\theta_i\beta_{ij})^{w_n^j}\right) d\theta,$$

$$p(D\mid\alpha,\beta) = \prod_{d=1}^{M} \int p(\theta_d\mid\alpha)\left(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn}\mid\theta_d)p(w_{dn}\mid z_{dn},\beta)\right) d\theta_d.$$

27

Graphics: [3Blue1Brown](https://youtu.be/Ip3X9LOh2dk?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab&t=10)
https://youtu.be/Ip3X9LOh2dk?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab&t=10

See more in "learning and inference tasks" (week 2-3)