SAISON 22/23

Formation

Introduction Deep Learning

Séquence 05a

Stratégies d'évaluation des modèles

FIDLE

https://fidle.cnrs.fr

MIAI Grenoble Alpes

cnrs

UGA Université Grenoble Alpes

EFELIA - MIAI

ECOLE FRANÇAISE DE L'INTELLIGENCE ARTIFICIELLE

DEVLOG RÉSEAU

Resinfo

sari

ANITI

GRICAD

SIMaP

Cette session va être enregistrée.
Retrouvez-nous sur notre chaine YouTube :-)

This session will be recorded.
Find us on our YouTube channel :-)

https://fidle.cnrs.fr/youtube

SAISON 22/23

Formation

Introduction Deep Learning

Séquence 05a

Stratégies d'évaluation des modèles

FIDLE

https://fidle.cnrs.fr

MIAI Grenoble Alpes · cnrs · UGA Université Grenoble Alpes · EFELIA - MIAI ECOLE FRANÇAISE DE L'INTELLIGENCE ARTIFICIELLE

DEVLOG · Resinfo · sari · ANITI · GRICAD · SIMaP

Resources

https://fidle.cnrs.fr

Powered by CNRS CRIC, and UGA DGDSI
of Grenoble, Thanks !

- Course materials (pdf)
- Practical work environment*
- Corrected notebooks
- Videos (YouTube)

(*) Procedure via Docket or pip
Remember to get the latest version !

Resources

You can also subscribe to :

FIDLE http://fidle.cnrs.fr/listeinfo

DEVLOG RÉSEAU DU DEVELOPPEMENT LOGICIEL Recherche et Enseignement Supérieur https://listes.services.cnrs.fr/wws/info/devlog[1]

GROUPE CALCUL https://listes.math.cnrs.fr/wws/info/calcul[2]

(1) List of ESR* developers,
(2) List of ESR* « calcul » group
Where ESR is Enseignement Supérieur et Recherche, french universities and public academic research organizations

Program

20 Séquences
du 17 novembre au 14 mai 2023

1. **History, Fundamental Concepts**
   - **Basic Regression** DNN
   - **Basic Classification** DNN
2. 3. **Hight Dimensionnal Data** CNN
4. Demystify **mathematics** for neural networks.
5. **Training strategies** Evaluation — **Sparse data (text)** Embedding
6. **Sequences data** RNN
7. PyTorch — A small detour with **PyTorch**.
8. **«Attention is All You Need»** Transformers
9. New ! **Graph Neural Network** GNN
10. **Autoencoder networks** AE
11. **Variational Antoencoder** VAE
12. **Project session** «My project in 180 s»
13. **Generative Adversarial Networks** GAN
14. New ! **Diffusion Model** Text to image
15. **AI, Law, Society** and **Ethics**
16. New ! **Model and training optimization** Resource efficiency
17. **Jean-Zay** GPU acceleration
18. New ! **Physics-Informed Neural Networks** PINNS
19. **Deep Reinforcement Learning** RL
20. **What will be tomorrow's AI** Review & perspectives !

FIDLE

SAISON 22/23

5a 5b

Training strategies
Evaluation

Sparse data (text)
Embedding

5a

Training strategies
Evaluation

5b

Sparse data (text)
Embedding

Full description on https://fidle.cnrs.fr/programme

Training strategies

5a 5b

Training strategies
Models evaluation

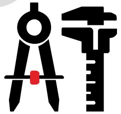
5.1 Training strategies

- Basic learning process and bias
- Hold-out evaluation
- K-fold evaluation

5.2 Finding the right metric

- Implementation of a simple case

5a 5b



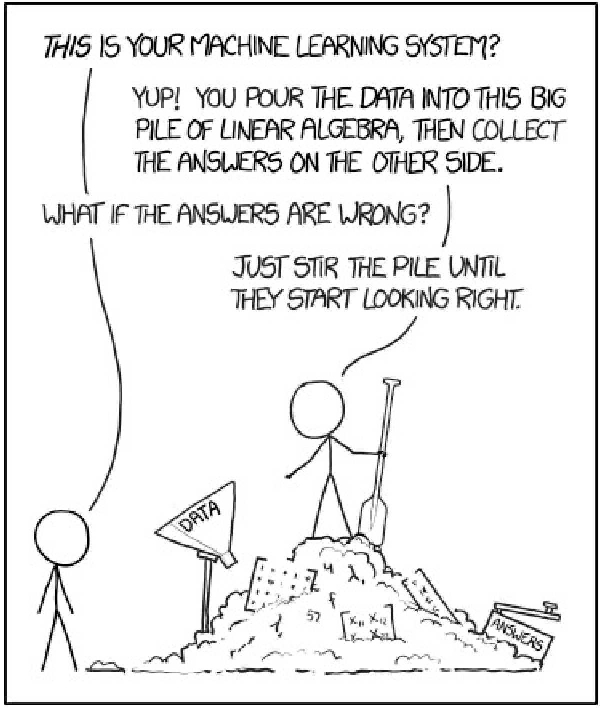
Training strategies
Models evaluation

5.1 **Training strategies**

- Basic learning process and bias
- Hold-out evaluation
- K-fold evaluation

5.2 **Finding the right metric**

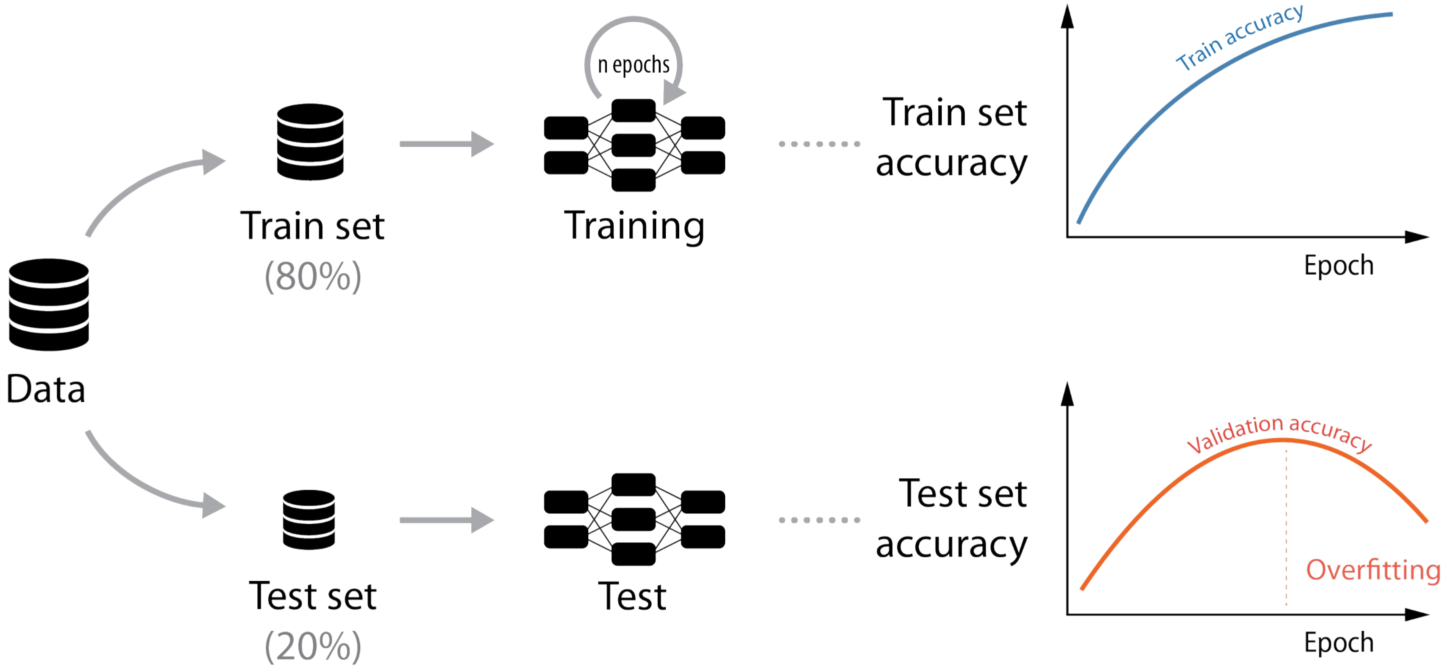
- Implementation of a simple case



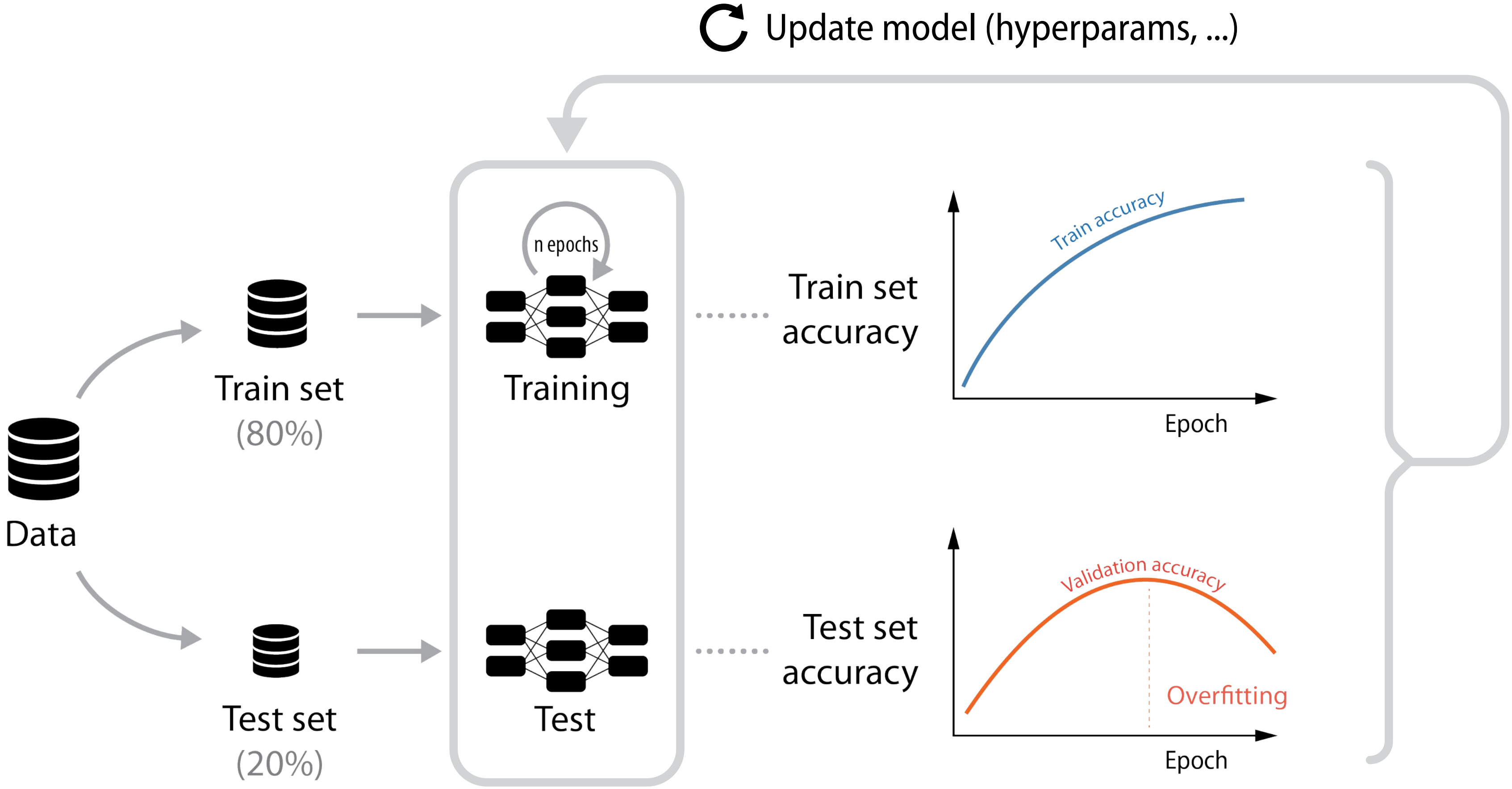
Learning is not so easy!

https://imgs.xkcd.com/comics/machine_learning.png

Basic learning process

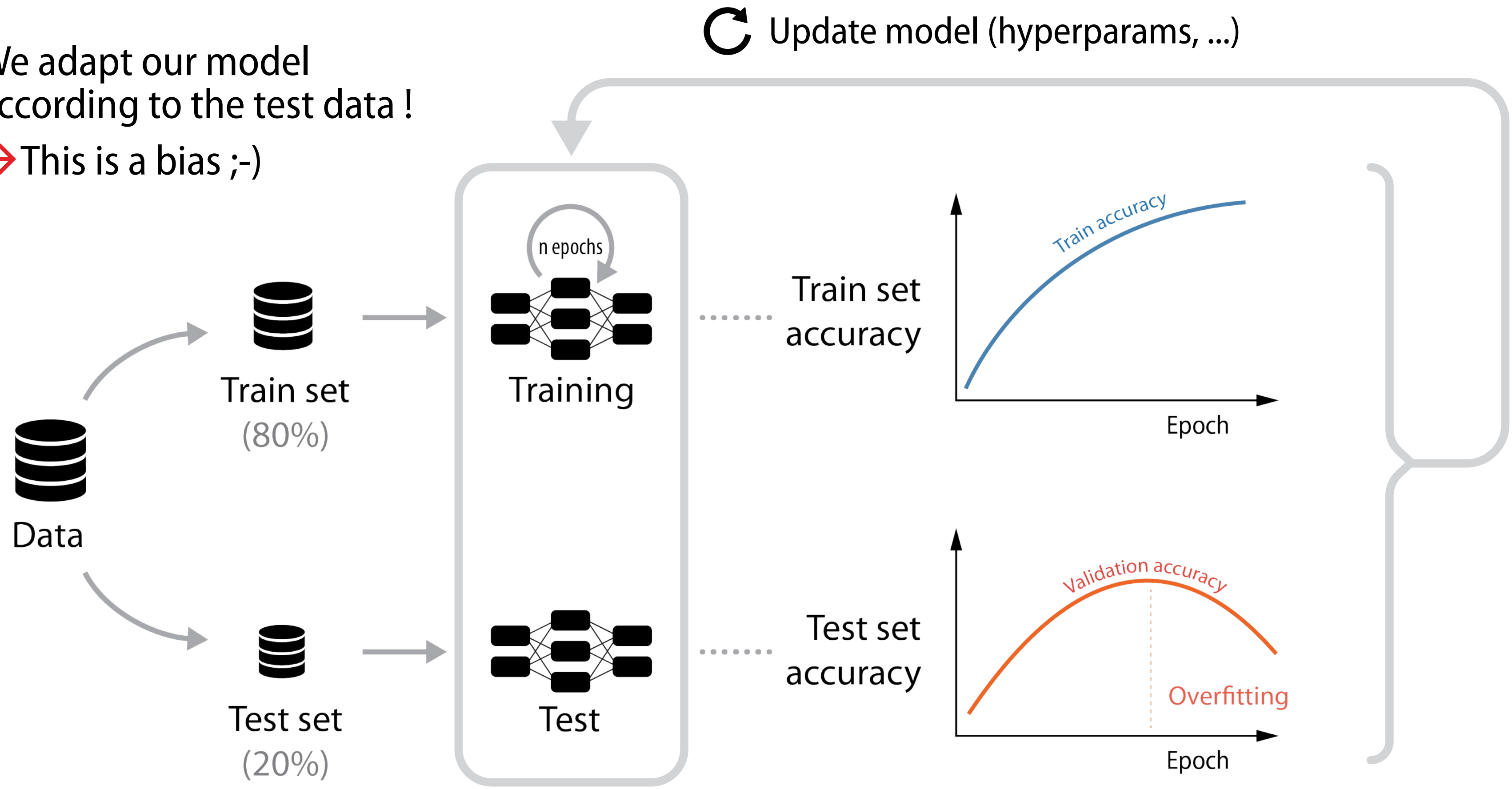


Basic learning process



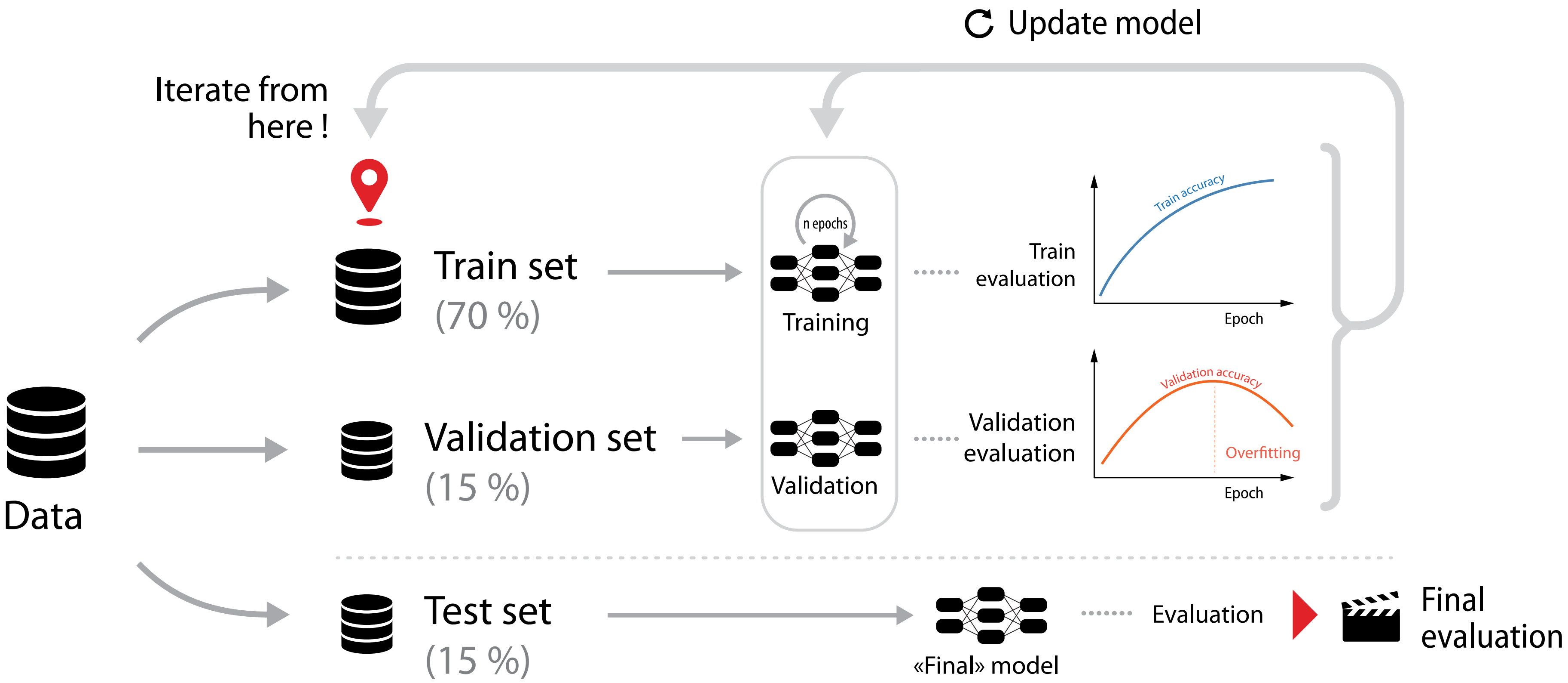
Basic learning process

We adapt our model according to the test data !
→ This is a bias ;-)



Hold-out evaluation

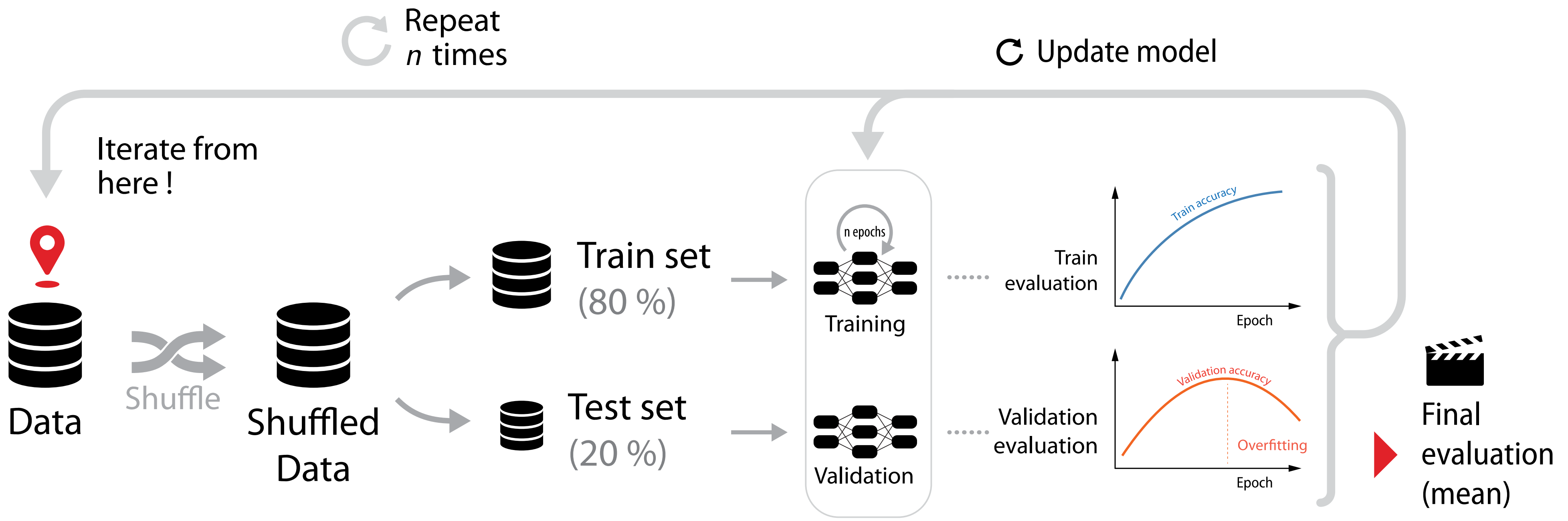
Validation simple



Suitable for large datasets, allowing for large Validation and Test sets.
If the Validation/Test sets are too small, the final evaluation will be statistically unstable.

Iterative hold-out evaluation with shuffling

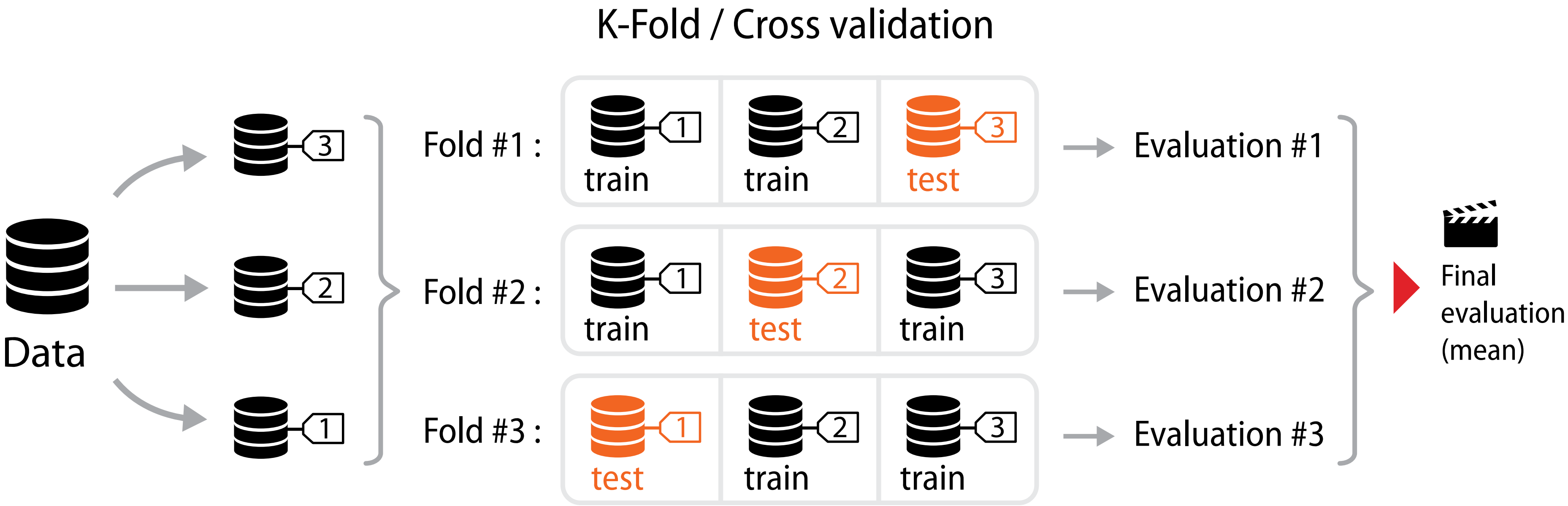
Validation simple, itérative avec brassage des données



Suitable for medium sized datasets : generating a shuffled dataset can be expensive...
The number of iterations depends on the data.

K-fold cross validation

Validation croisée



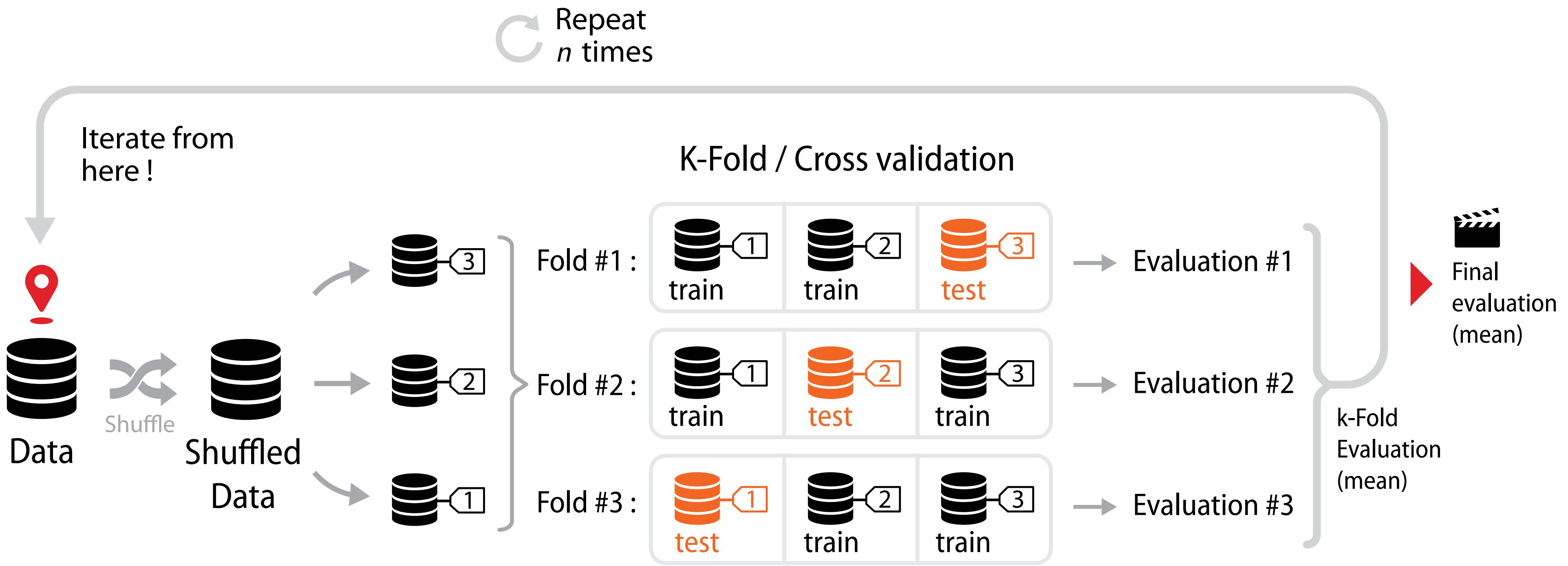
!

In this example, k=3
In practice, we will rather use a k=5, k=8, ...

Very interesting strategy for small datasets.
If the amount of data is small, however, the result can remain unstable...

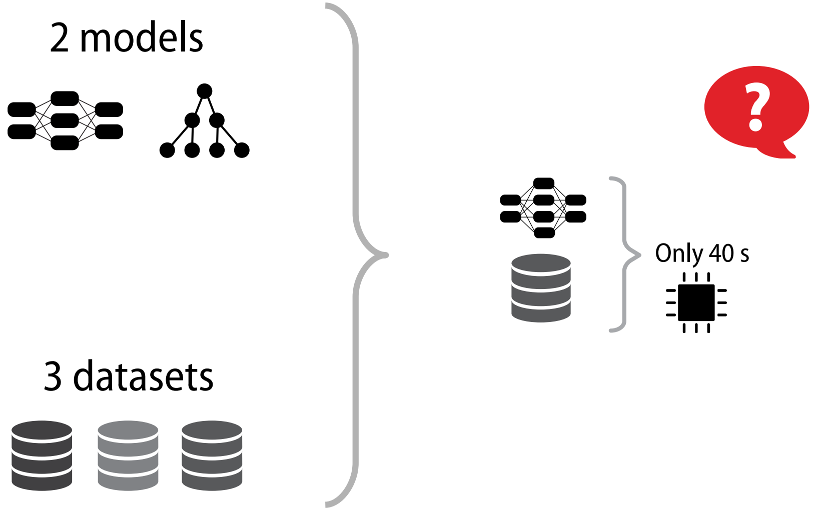
Iterated K-fold cross validation with shuffling

Validation croisée, itérative avec brassage des données



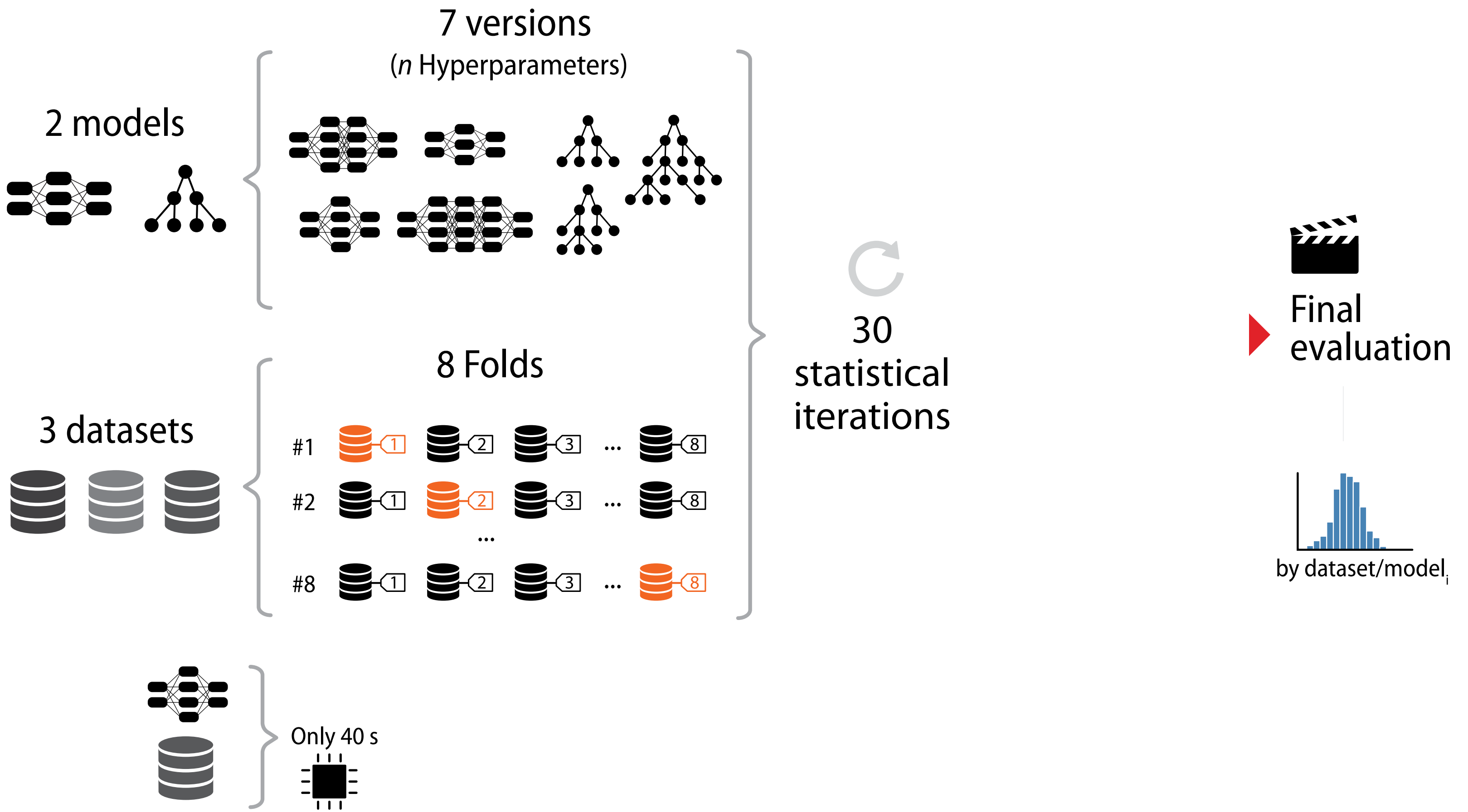
Probably the best strategy for small datasets...
...but if K and n are important, the combination can become very expensive !

Be careful with combinatories ;-)

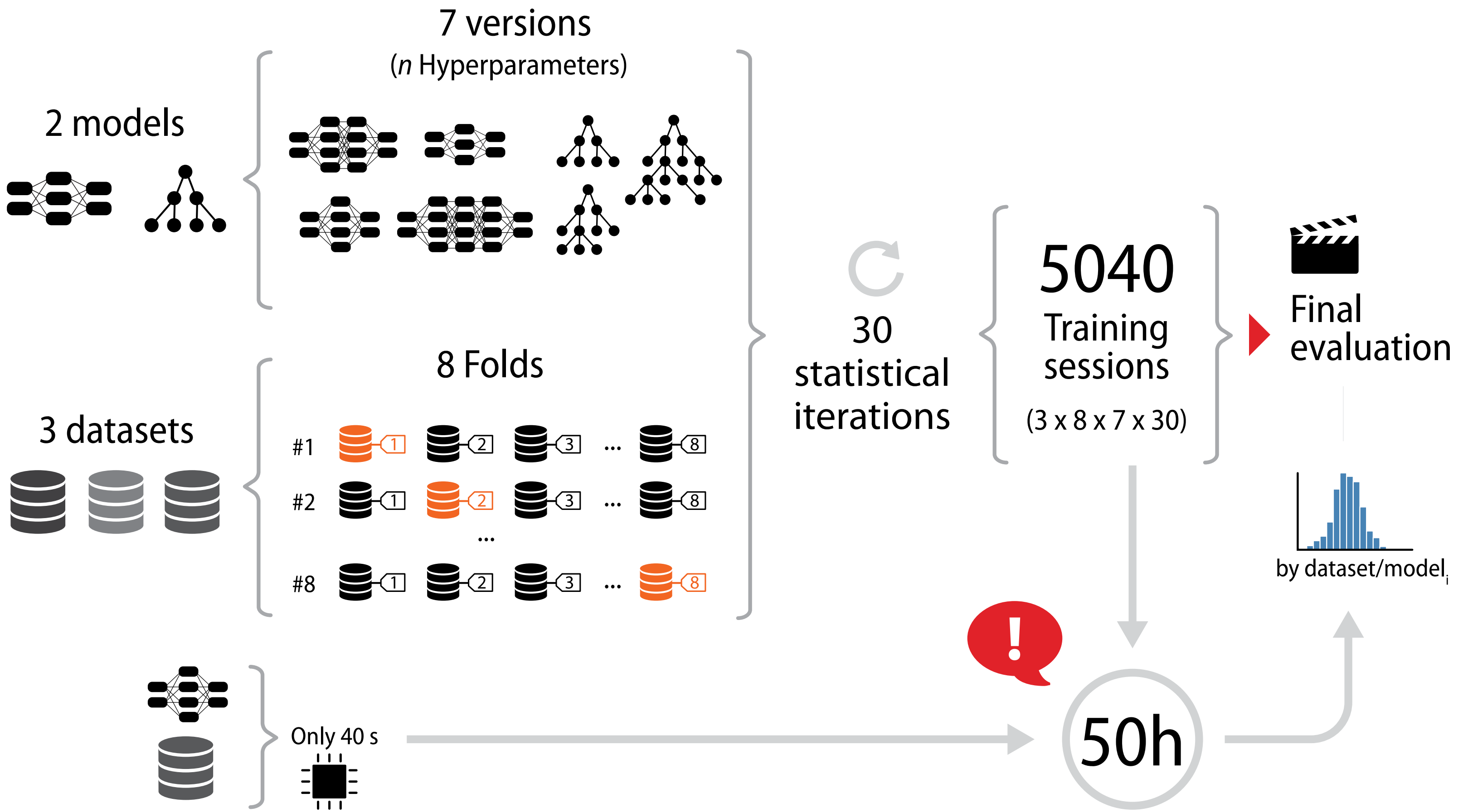


Which model is the best?
Which hyperparameters should I use?
Which dataset is the most usable?
Are my results significant?

Be careful with combinatories ;-)



Be careful with combinatories ;-)



A few questions to keep in mind !

?

Are my data subsets (train, test, ...) representative of my data?

Can I or should I shuffle my data? (time sequences, ordered data, ...)

Within the dataset, what is the share and impact of outliers?

Mes résultats sont-ils significatifs ?

How many folds, how many iterations do I need ?

How much data do I need?

How much will it cost ?

?

5a 5b

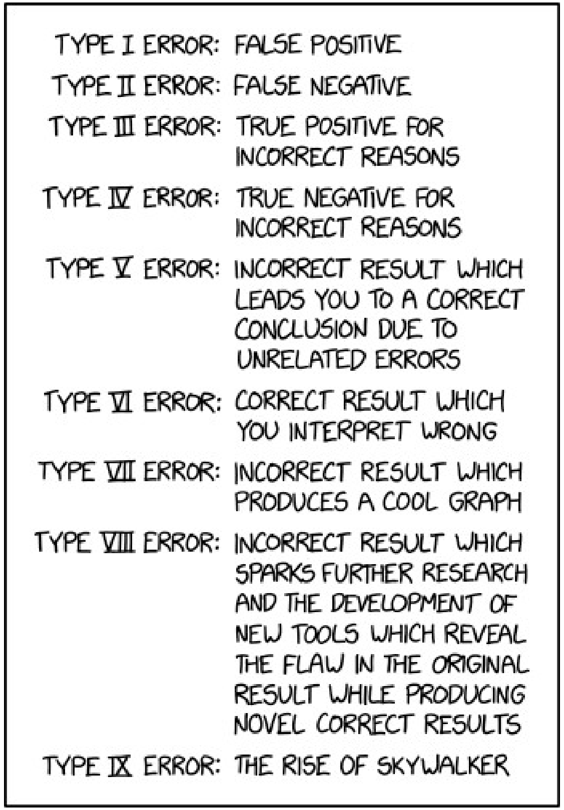
Training strategies
Models evaluation

5.1 Training strategies

- Basic learning process and bias
- Hold-out evaluation
- K-fold evaluation

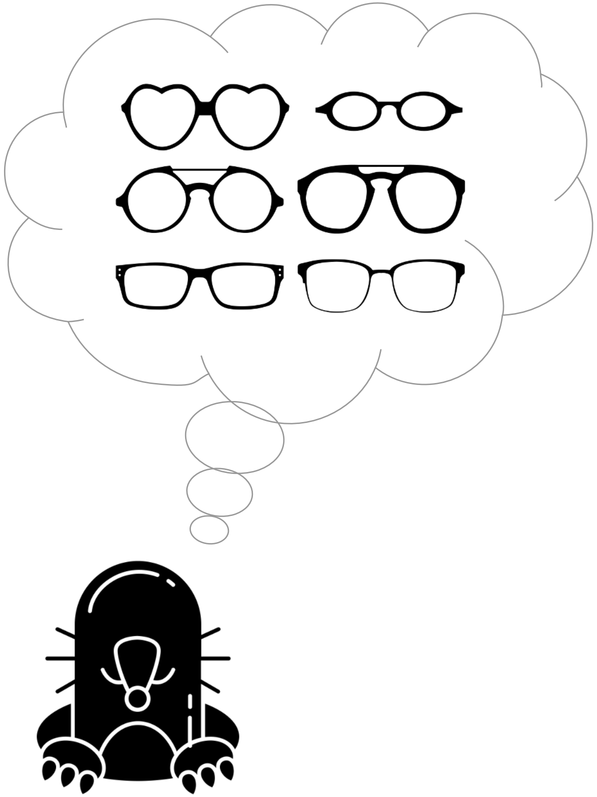
5.2 Finding the right metric

- Implementation of a simple case

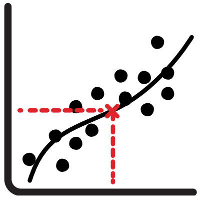


https://imgs.xkcd.com/comics/error_types.png

Evaluating a result is not easy!



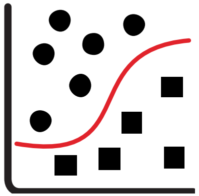
Finding the right metric ?



Regression

We try to predict a quantity (scalar, vector, ...)

→ Is my predicted value "good"?

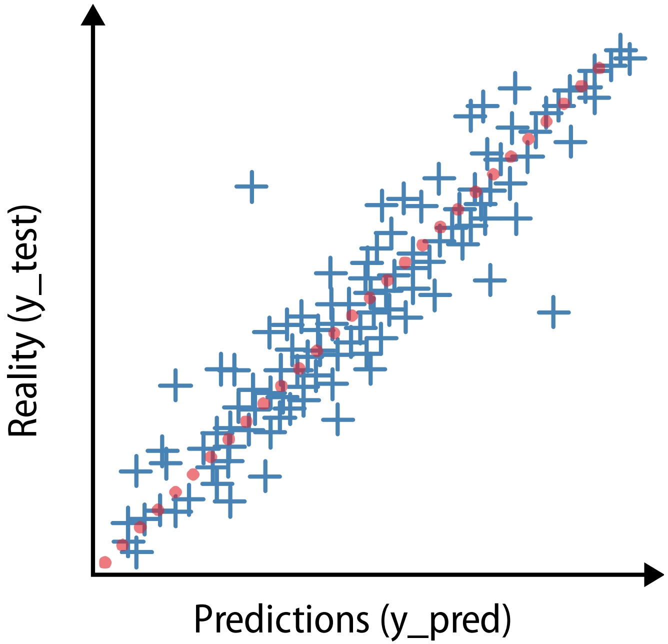


Classification

We try to predict a quality (class membership, ...)

→ Is my prediction "correct"?

Evaluation of a regression



$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}^{(i)} - y^{(i)}\right]^2$$

Mean Squared Error
Differentiable
Can be use as lost function
Increases very quickly

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}^{(i)} - y^{(i)}\right]^2}$$

Root Mean Squared Error
Same unit as y
Robust to outliers
Humans understandable

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}^{(i)} - y^{(i)}\right|$$

Mean absolute error
Same unit as y
More robust to outliers
Humans understandable

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=0}^{n}\frac{|y_i - \hat{y}_i|}{max(\epsilon, |y_i|)}$$

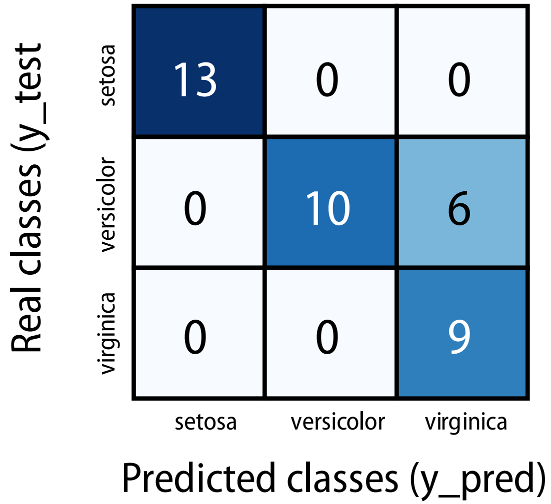
Mean Absolute Percentage Error
Humans understandable (%)
Problem when y is null !

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

R^2 score, coefficient of determination
Result in [0, 1]
Measures a correlation between 2 series, nothing else..

https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metics

Evaluation of a classification



$$accuracy = \frac{Total\ number\ of\ correct\ predictions}{Total\ number\ of\ prédictions}$$

Ability to make correct predictions

« exactitude » en fr

$$HammingLoss = \frac{Total\ number\ of\ wrong\ predictions}{Total\ number\ of\ prédictions}$$

Ability to make wrong predictions

$$precision_{class\,i} = \frac{Number\ of\ correct\ predictions\ for\ class\ i}{Total\ number\ of\ predictions\ for\ class\ i}$$

Ability to identify without error, the elements of the *class i*

$$recall_{class\,i} = \frac{Number\ of\ correct\ predictions\ for\ class\ i}{Total\ number\ of\ real\ class\ i}$$

Ability to identify all the elements of the *class i*

«sensibilité» en fr

$$F1_{class\,i} = 2 * \frac{recall_{class\,i} \cdot precision_{class\,i}}{recall_{class\,i} + precision_{class\,i}}$$

F1 is the harmonic mean of the model's precision and recall.

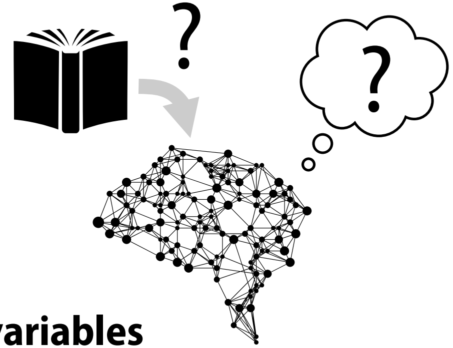
https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics

Next, on Fidle :

5b

Sparse data (text)
Embedding

Jeudi 15 décembre

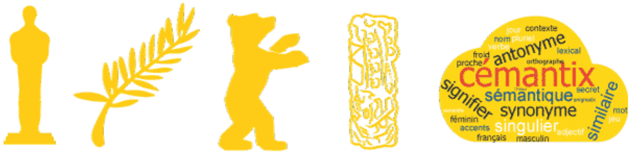


Séquence 5b :

Données creuses/textuelles de dimensions variables

Spécificités et gestion des données creuses/textuelles
Principes de l'Embedding (Keras, CBOW, Skip-Gram)
...ou comment réduire les dimensions !

Exemple proposé :
Analyse de sentiment avec une analyse de critique de films.



Next on Fidle :

5b
Sparse data
Embedding

Jeudi 15 décembre, 14h00

Séquence 5b :

Données creuses/textuelles de dimensions variables

To be continued...

Contact@fidle.cnrs.fr
FIDLE https://fidle.cnrs.fr
YouTube https://fidle.cnrs.fr/youtube

FIDLE