



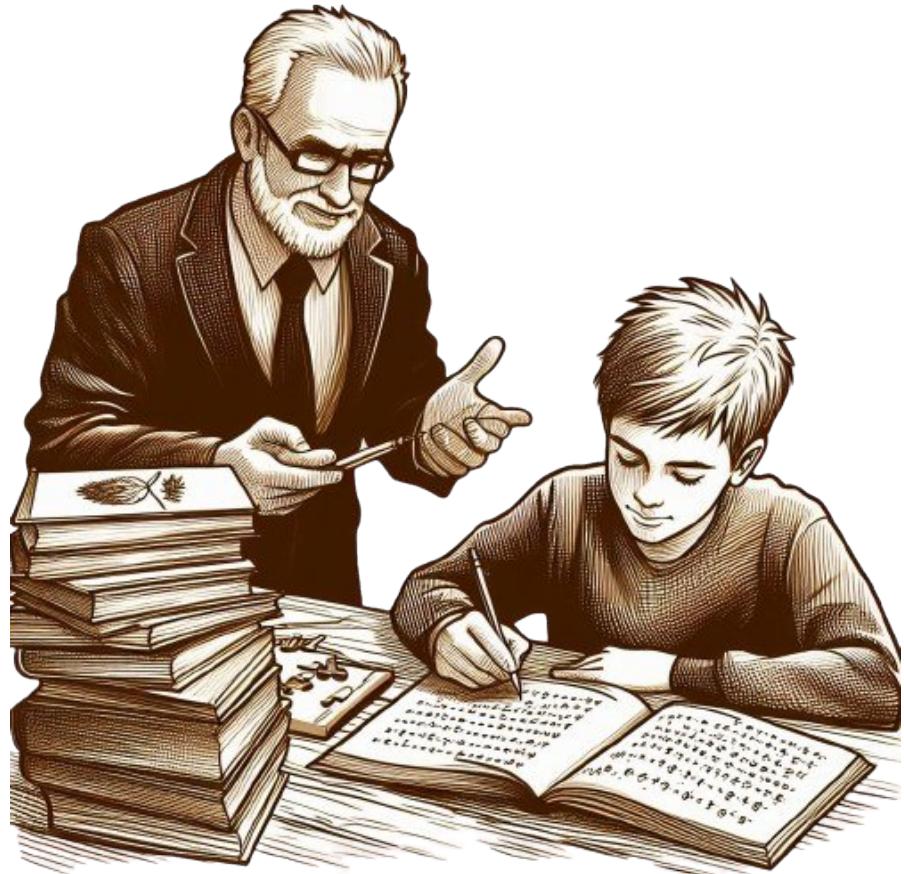
Hands-on Introduction to Deep Learning

Methodology



IDRIS

- 1 Quel travail faire pour améliorer les données utilisées pour l'entraînement ?
- 2 Comment évaluer un modèle ?
- 3 Est-il possible de rendre l'entraînement plus robuste ?
- 4 Peut-on profiter d'un modèle déjà entraîné ?
- 5 Bonus : Quelques bonnes pratiques ?



Learning from exercise with a
teacher to guide us

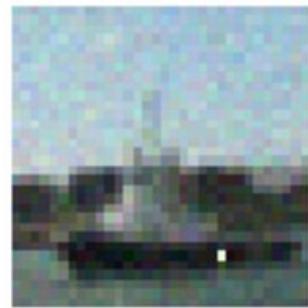


Applying what we learn
to the real world

Training & Generalization

CIFAR-10

AllConv



SHIP
CAR (99.7%)

NiN



HORSE
FROG (99.9%)

VGG16



DEER
AIRPLANE (85.3%)

ImageNet

BVLC AlexNet

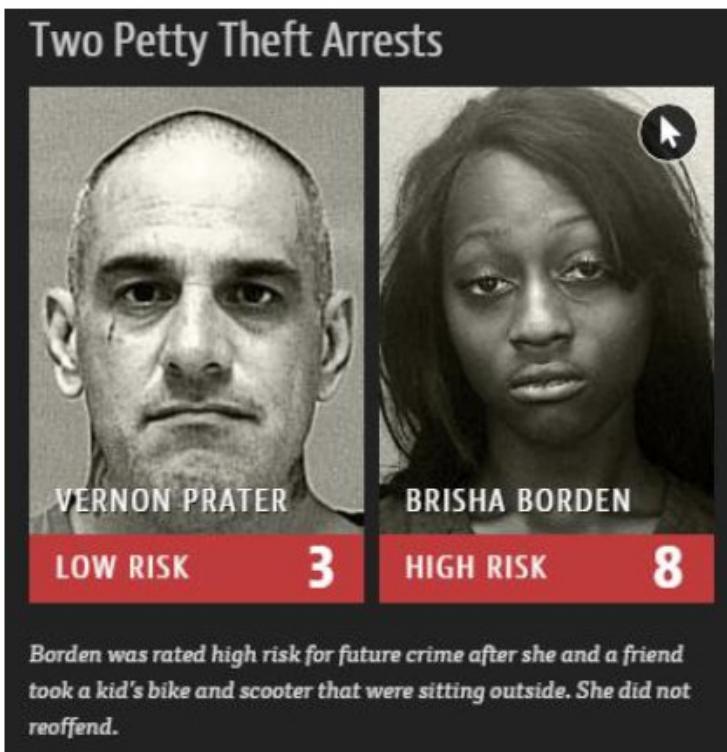


Cup (16.48%)
Soup Bowl (16.74%)



Bassinet (16.59%)
Paper Towel (16.21%)

COMPAS



Source : Propublica

ALGORITHME DE RECRUTEMENT

Quand le logiciel de recrutement d'Amazon discrimine les femmes

En 2014, le géant du e-commerce a voulu confier ses candidatures à un algorithme, mais celui-ci a commencé à écarter les profils féminins.

Source : *Les Echos*

Amazon a dû désactiver une IA qui discriminait les candidatures de femmes à l'embauche

Source : Numerama

Generalization Lack Issue : Discrimination

Dataset is the guilty ?!



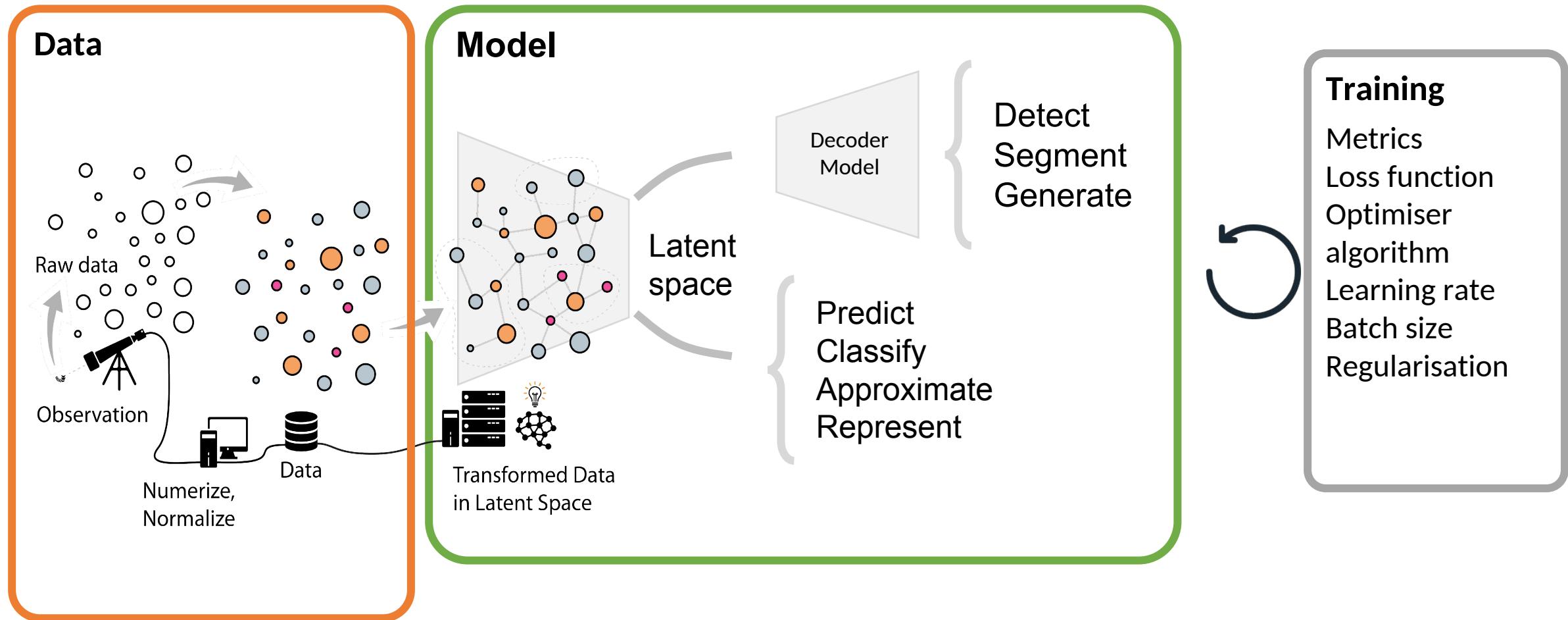
Augment It, Transform It !!!

For a Better Generalization...

Super Regularizer !!



Use his Super Powers !!



Deep Learning Pipeline Principle

1

Quel travail faire pour améliorer les données utilisées pour l'entraînement ?

2

Comment évaluer un modèle ?

3

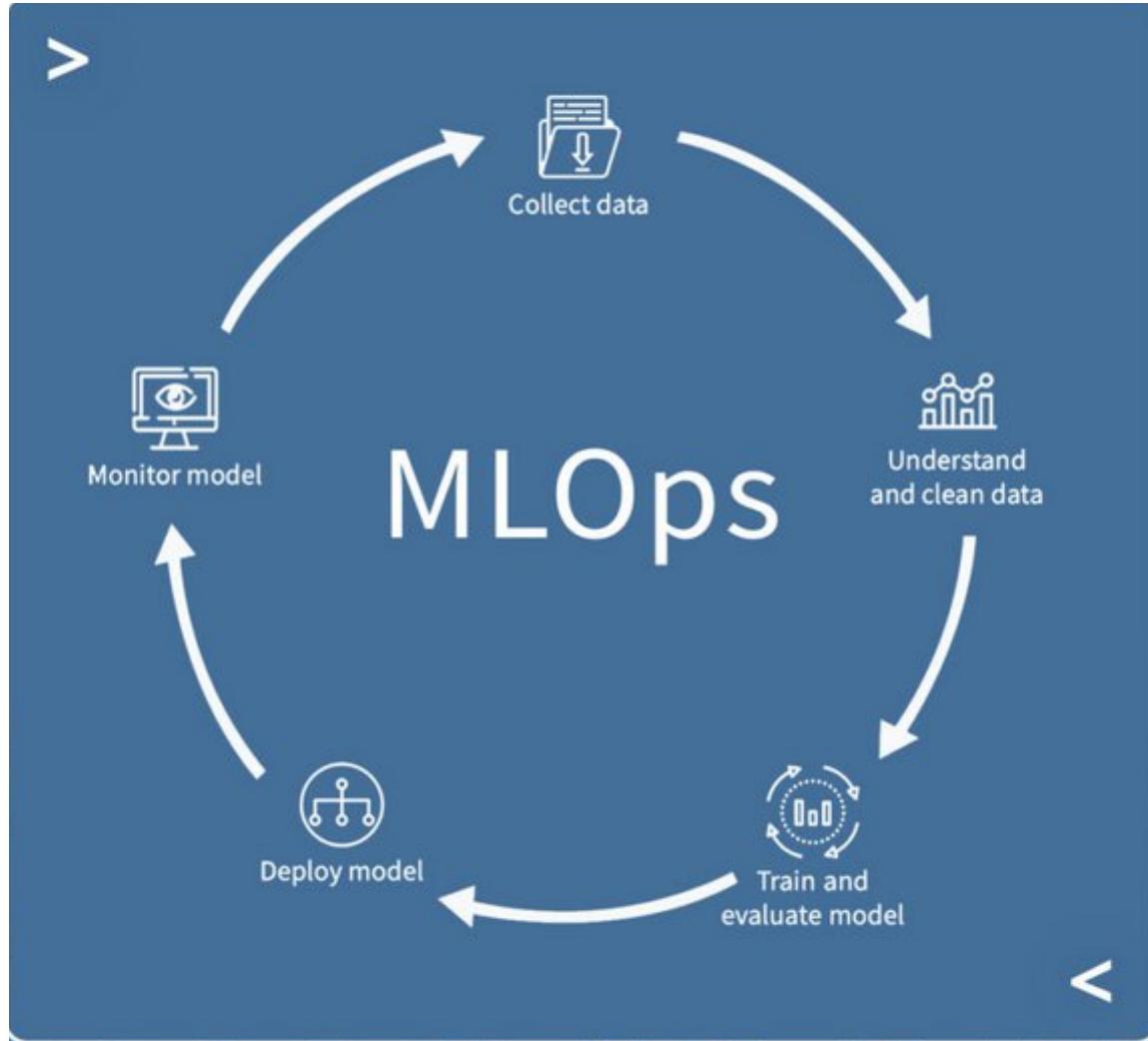
Est-il possible de rendre l'entraînement plus robuste ?

4

Peut-on profiter d'un modèle déjà entraîné ?

5

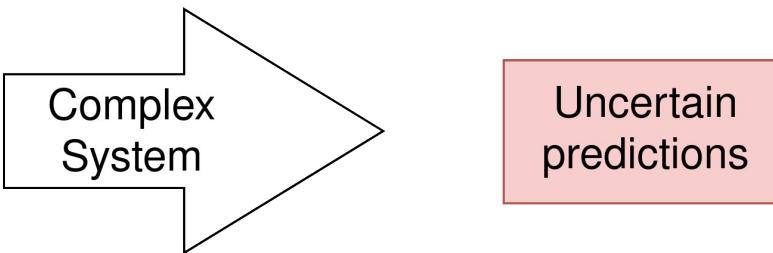
Bonus : Quelques bonnes pratiques ?



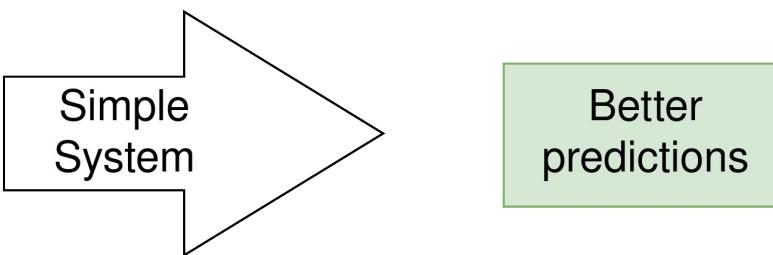
50 to 80% time spent on data

Model development

Gender
Height
Eye color
Fingers length
Weight



Height
Weight



Diabetes risk prediction system

Features selection - Example

All the features



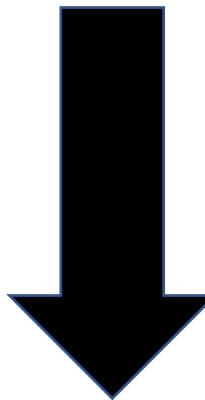
Selection

Selected features



Features selection

All the features

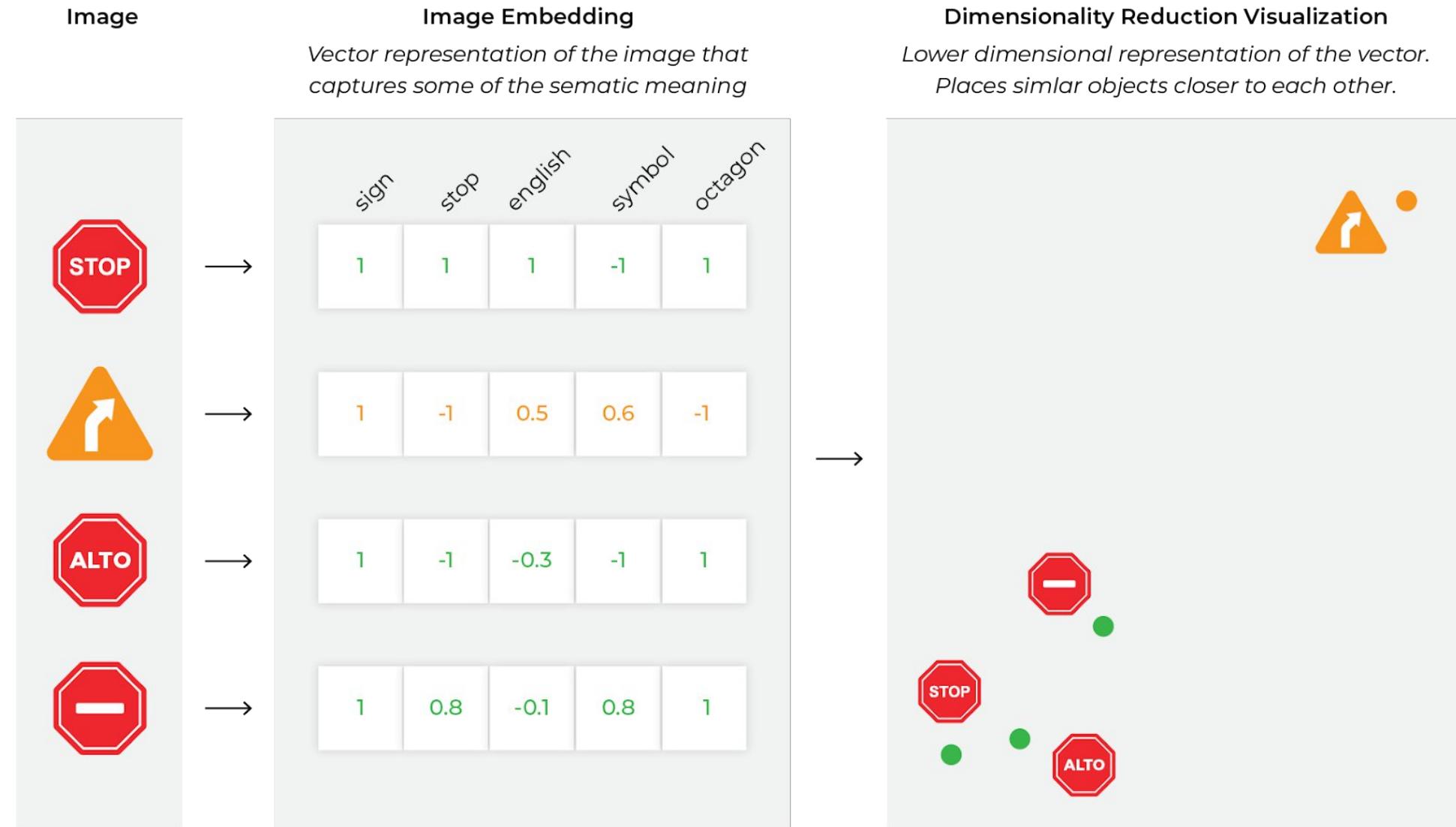


Extraction

Extracted features



Features extraction



Features extraction - Embedding example

All the features

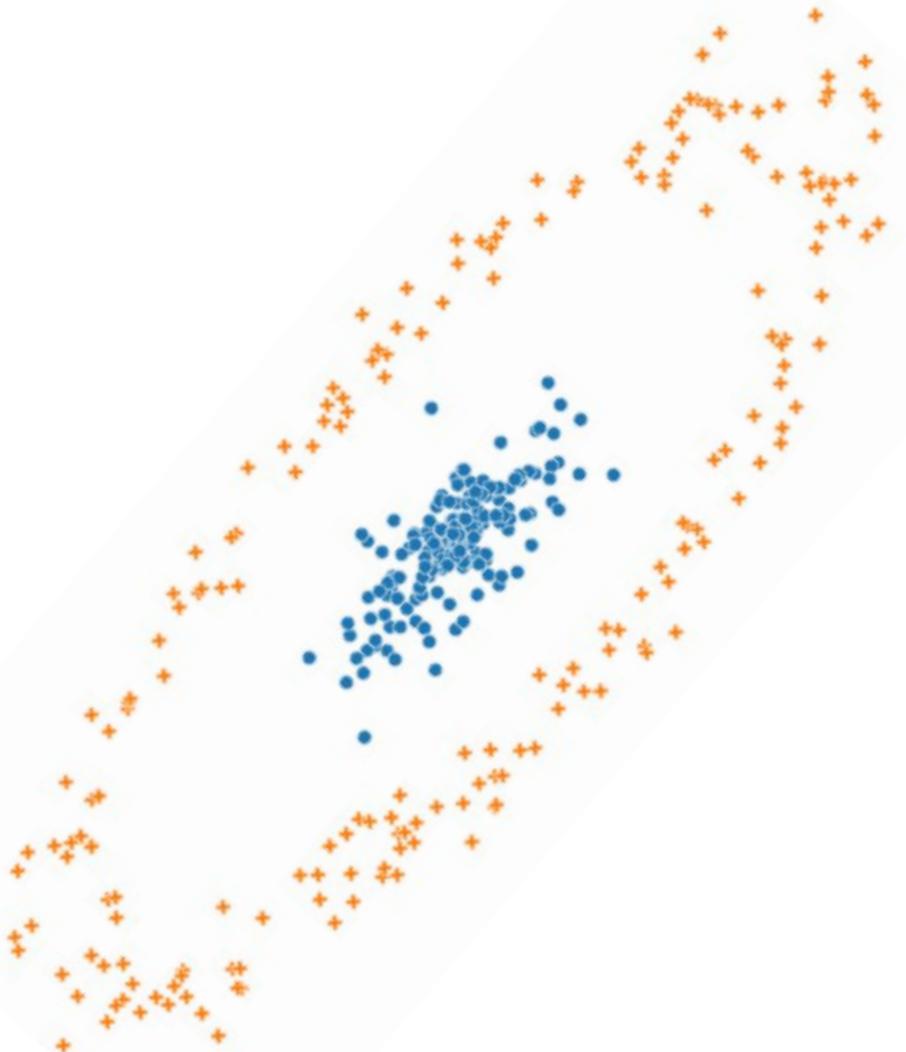


Transformation

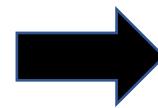
Transformed features



Feature transformation

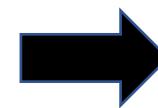


(x, y)



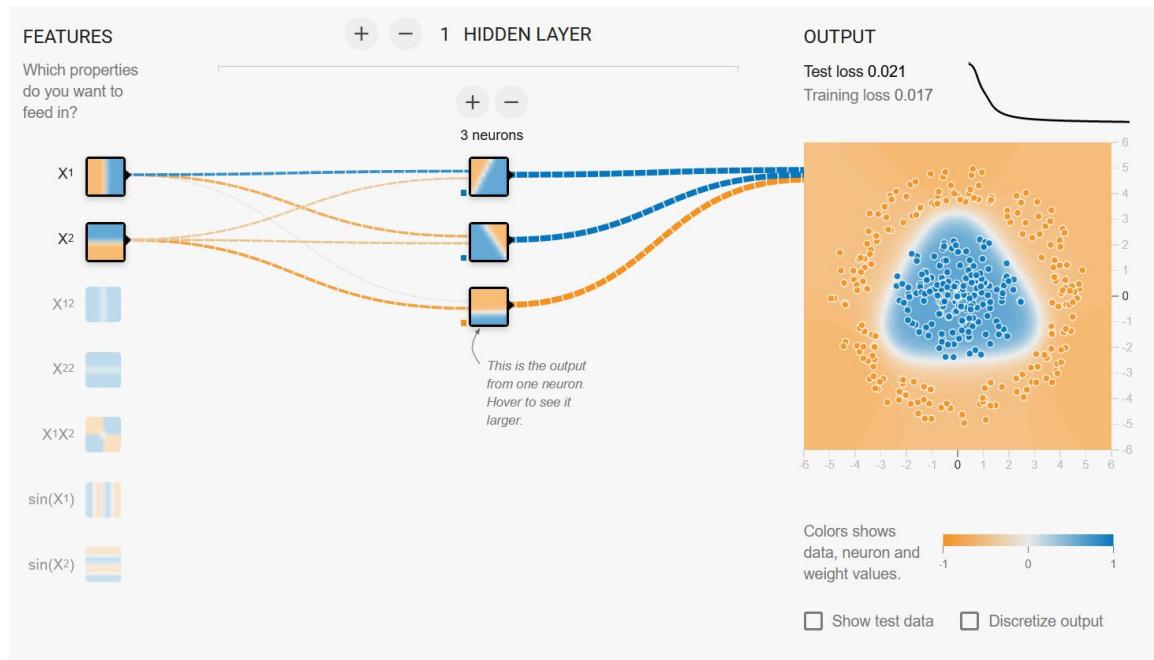
Complex relation
between x and y

(r, θ)



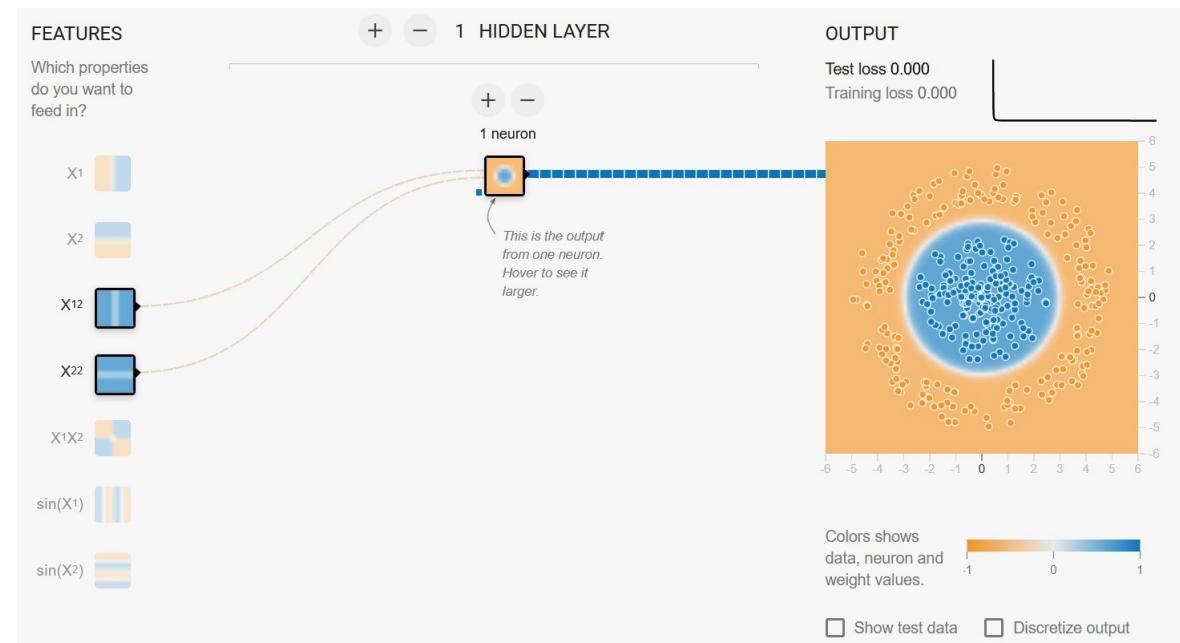
Simple relation
with r and θ

Feature transformation - Example

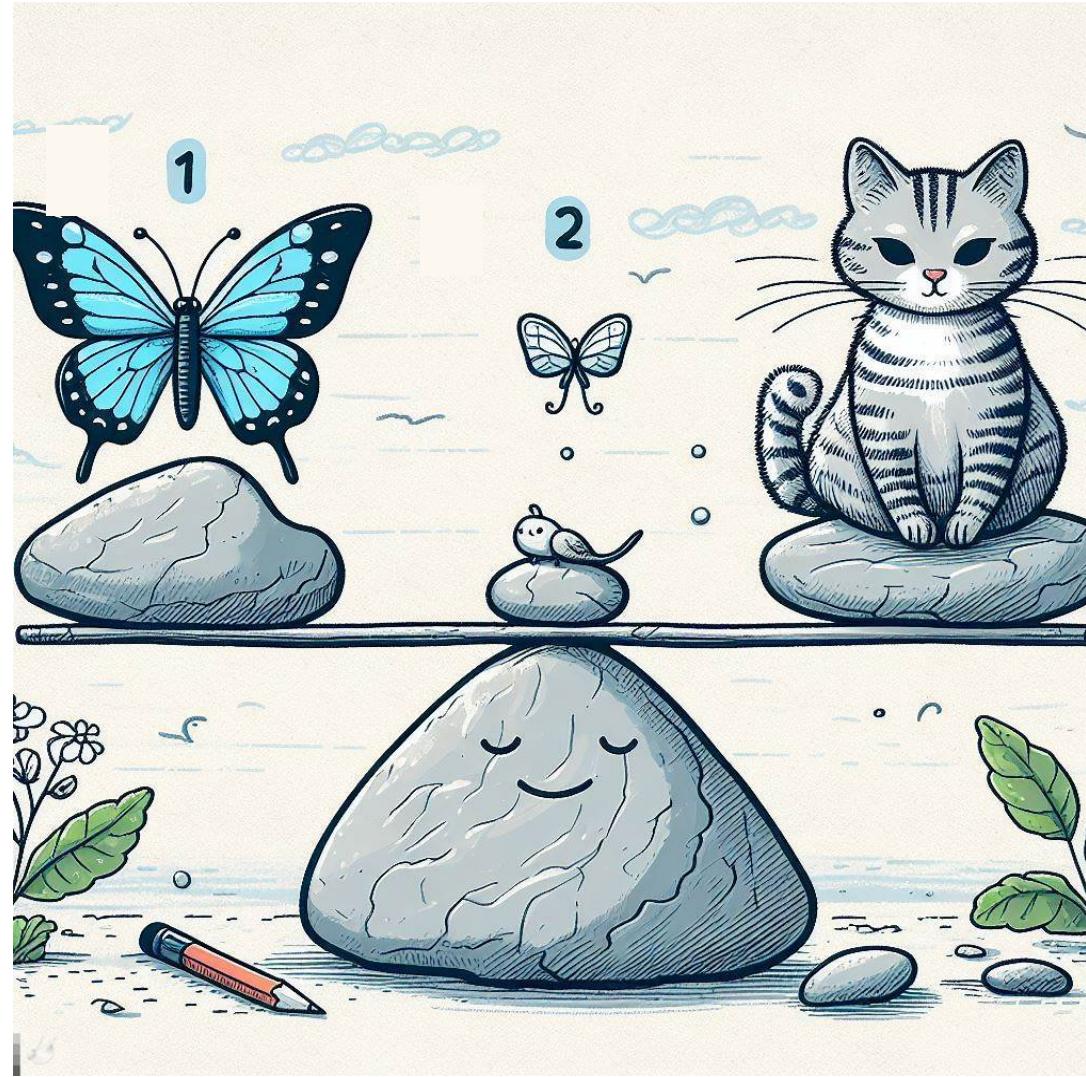


Données transformées
Réseau à 1 neurone

Données de position (x, y)
Réseau à 3 neurones

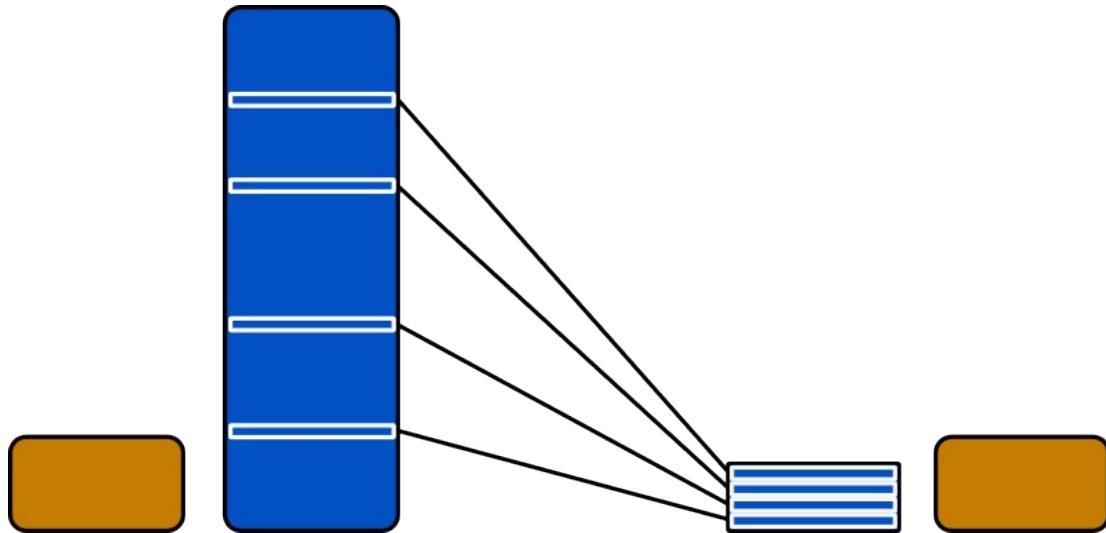


Feature transformation - Example

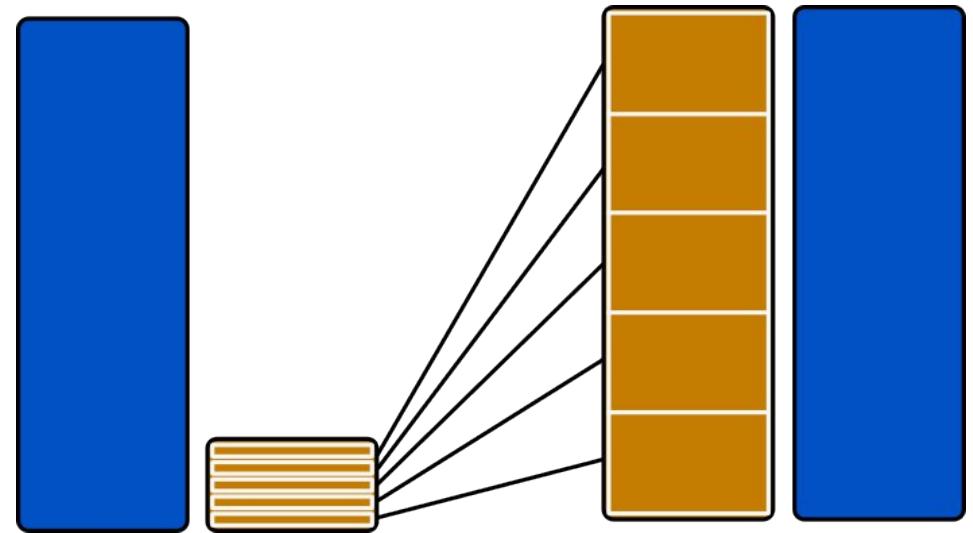


Balance the classes - Example

Undersampling

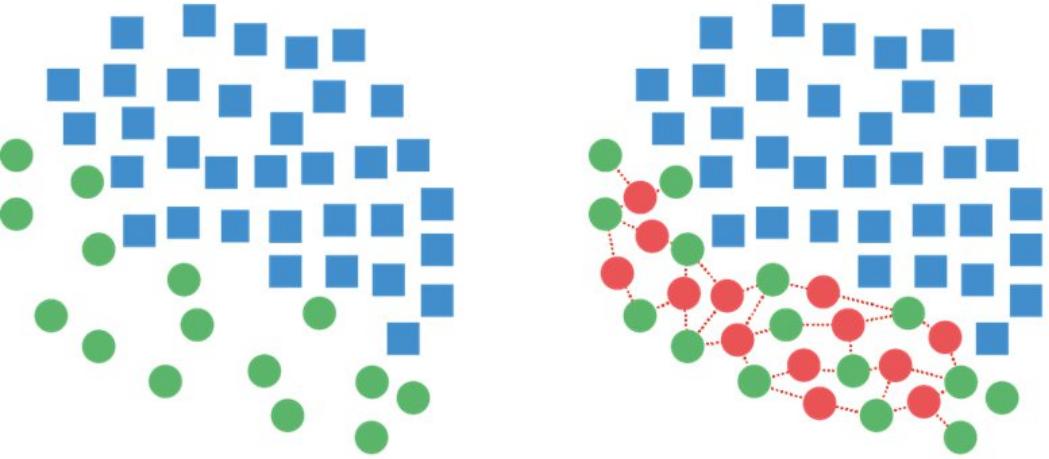


Oversampling

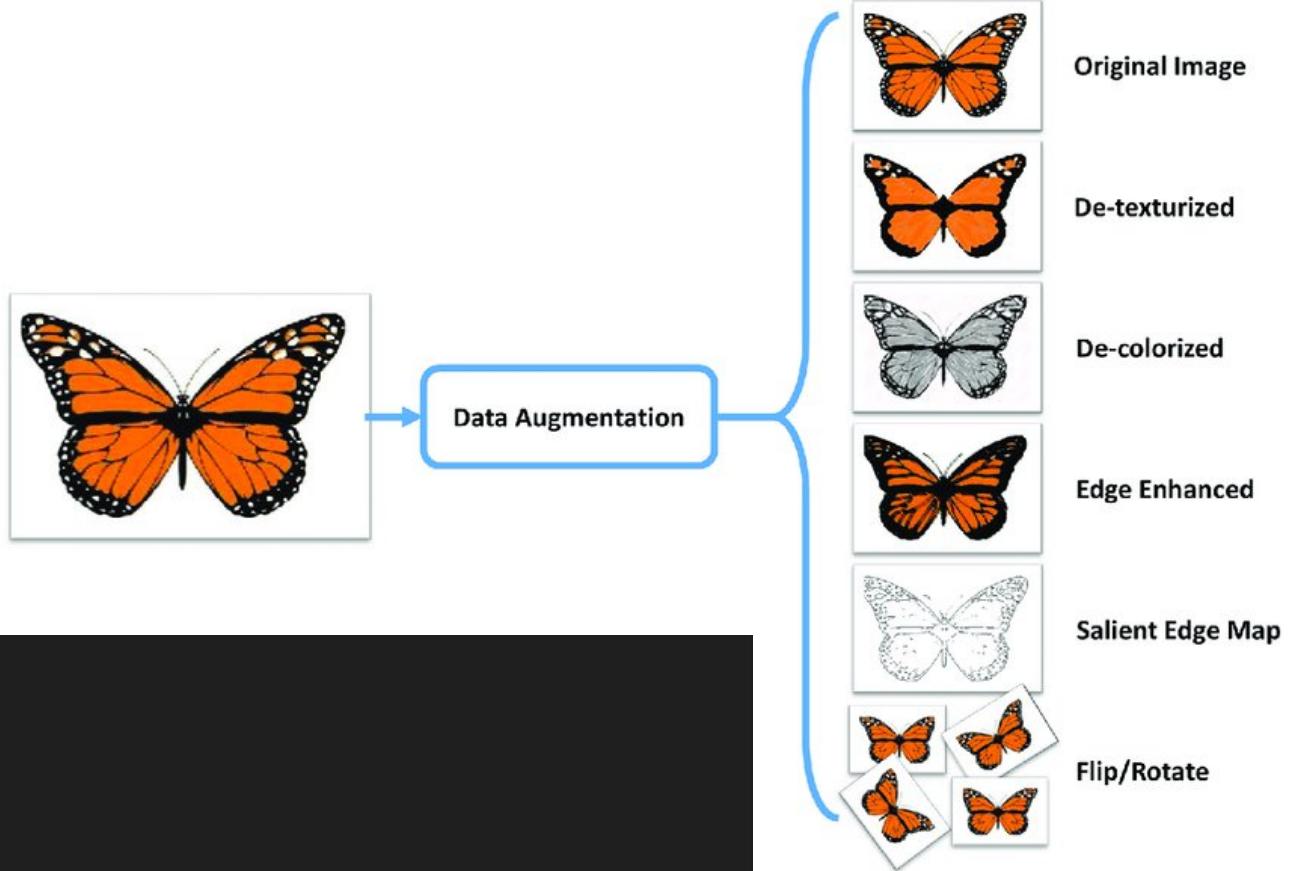


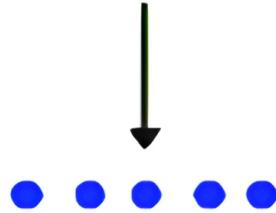
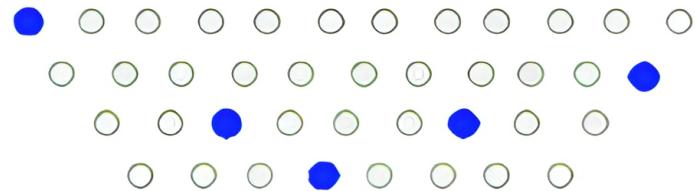
Balance the classes

Synthetic Minority Oversampling Technique

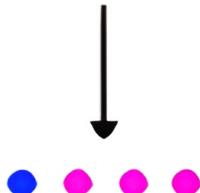
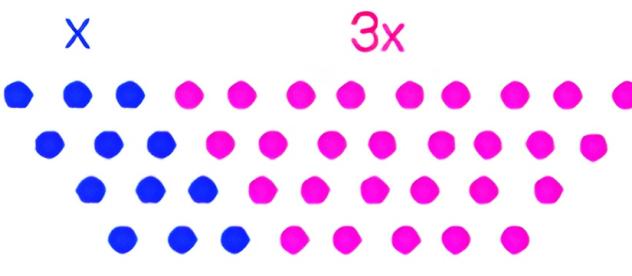


```
# Define the transformations
transform = transforms.Compose([
    transforms.RandomHorizontalFlip(),
    transforms.RandomRotation(10),
    transforms.RandomResizedCrop(224),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
])
```

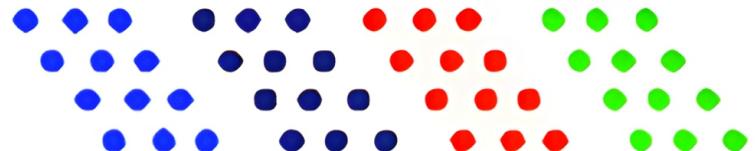




Random Sampling



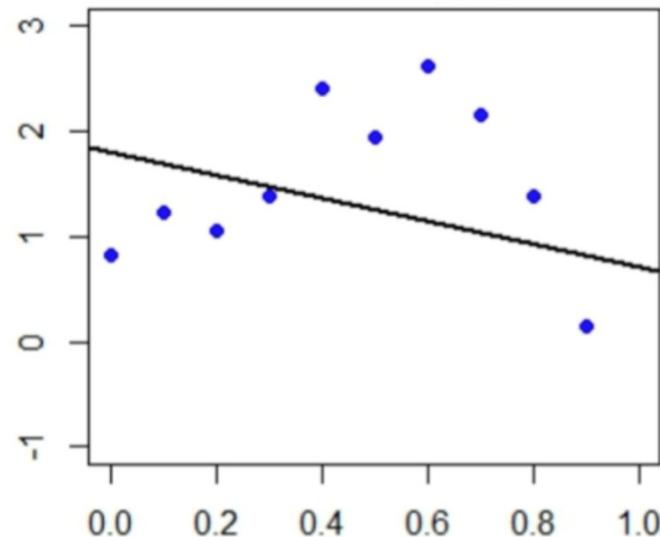
Stratified Sampling



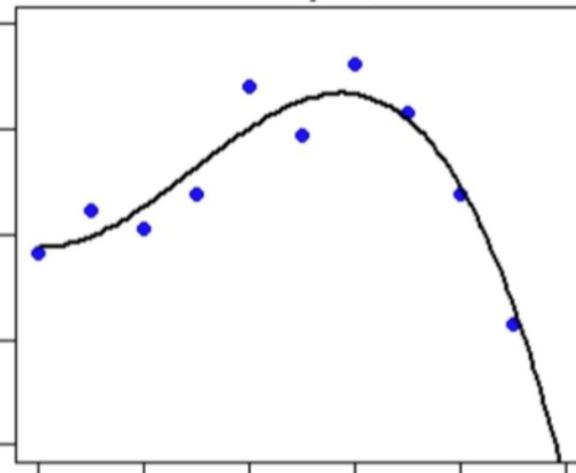
Cluster Sampling

- 1** Quel travail faire pour améliorer les données utilisées pour l'entraînement
- 2** Comment évaluer un modèle ?
- 3** Est-il possible de rendre l'entraînement plus robuste ?
- 4** Peut-on profiter d'un modèle déjà entraîné ?
- 5** Bonus : Quelques bonnes pratiques ?

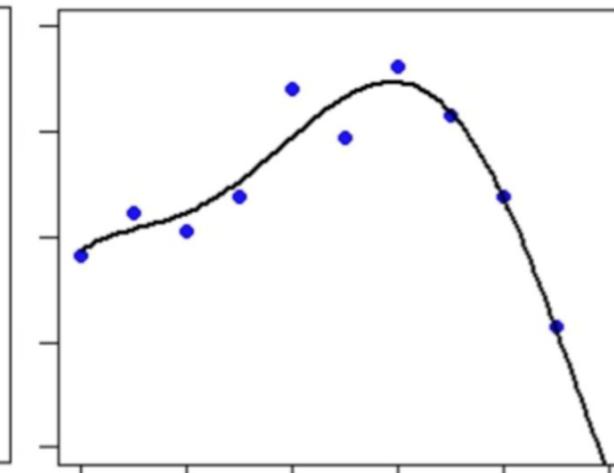
$$\hat{y}_i = a_0 + a_1 x_i$$



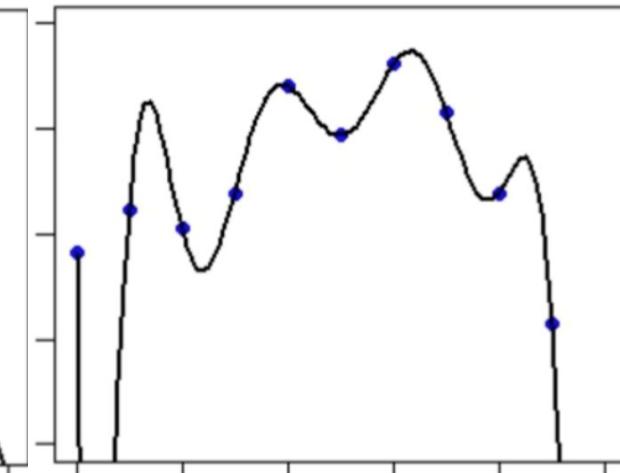
$$\hat{y}_i = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3$$



$$\hat{y}_i = a_0 + a_1 x_i + \dots + a_5 x_i^5$$



$$\hat{y}_i = a_0 + a_1 x_i + \dots + a_{10} x_i^{10}$$



Bias variance trade-off



Noise or bias ?



Bias



Noise

Noise or bias ?



Bias



Noise

- My model says it's a cat no matter the image I give him
- There is no correlation between my model predictions and the label

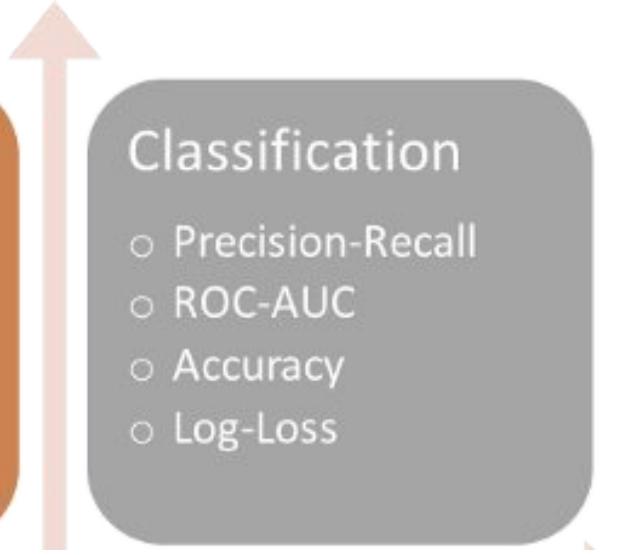
Noise or bias ?

Regression

- MSPE
- MSAE
- R Square
- Adjusted R Square

Classification

- Precision-Recall
- ROC-AUC
- Accuracy
- Log-Loss



- **What is a metric?**
- **Is it different from a loss function?**

Regression

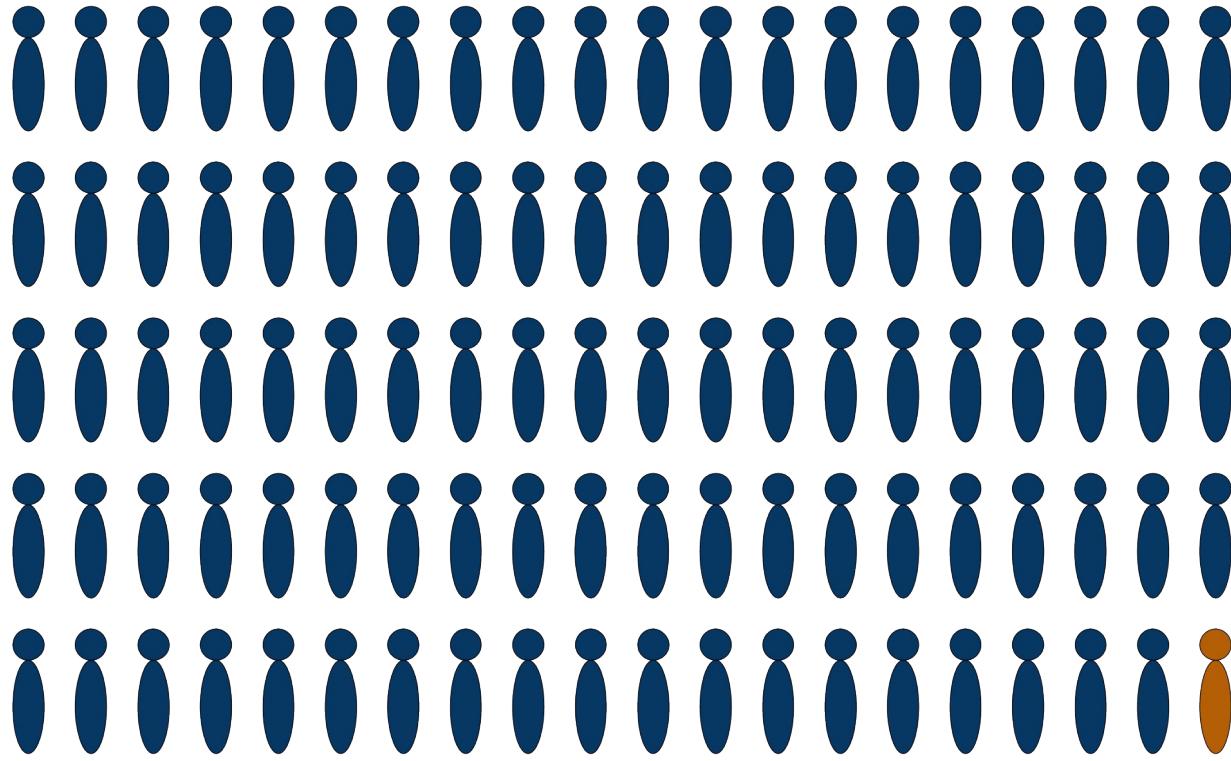
- MSPE
- MSAE
- R Square
- Adjusted R Square

Classification

- Precision-Recall
- ROC-AUC
- Accuracy
- Log-Loss



- **What is a metric?**
- **Is it different from a loss function?**
 - **Differentiability**
 - **Training vs evaluation**
 - **Number of samples**
 - **Interpretability/Meaning**



New system for illness detection – the accuracy is not a good metric for this case



All negative : accuracy = 99% ?

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

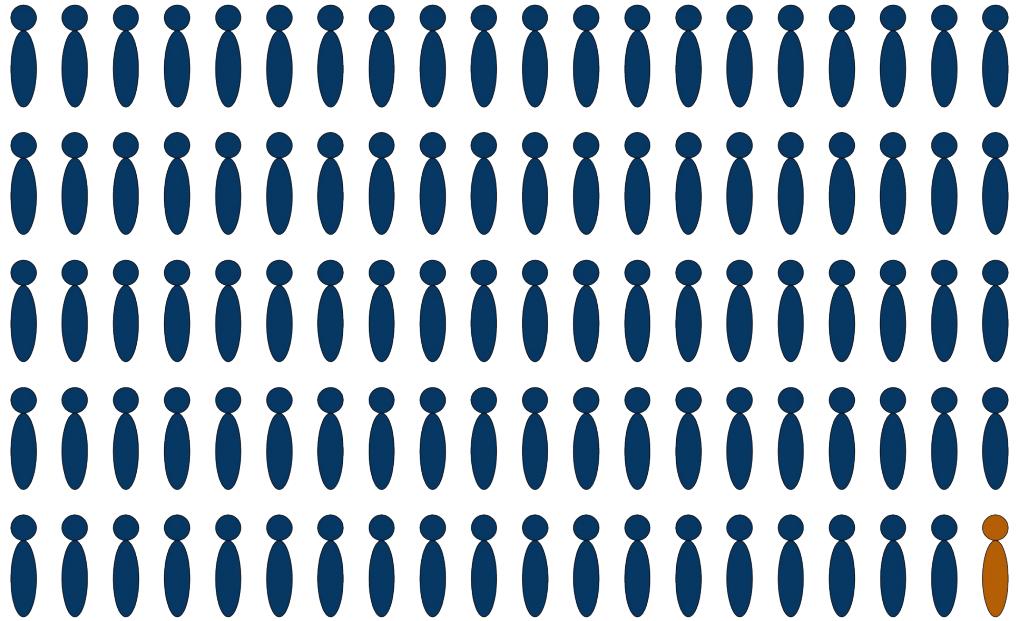
$$\text{precision} = \frac{TP}{TP + FP}$$

Above all positive prediction, how many are positive data

$$\text{recall} = \frac{TP}{TP + FN}$$

Above all positive data, how many have been predicted positive

Metrics - Precision/Recall



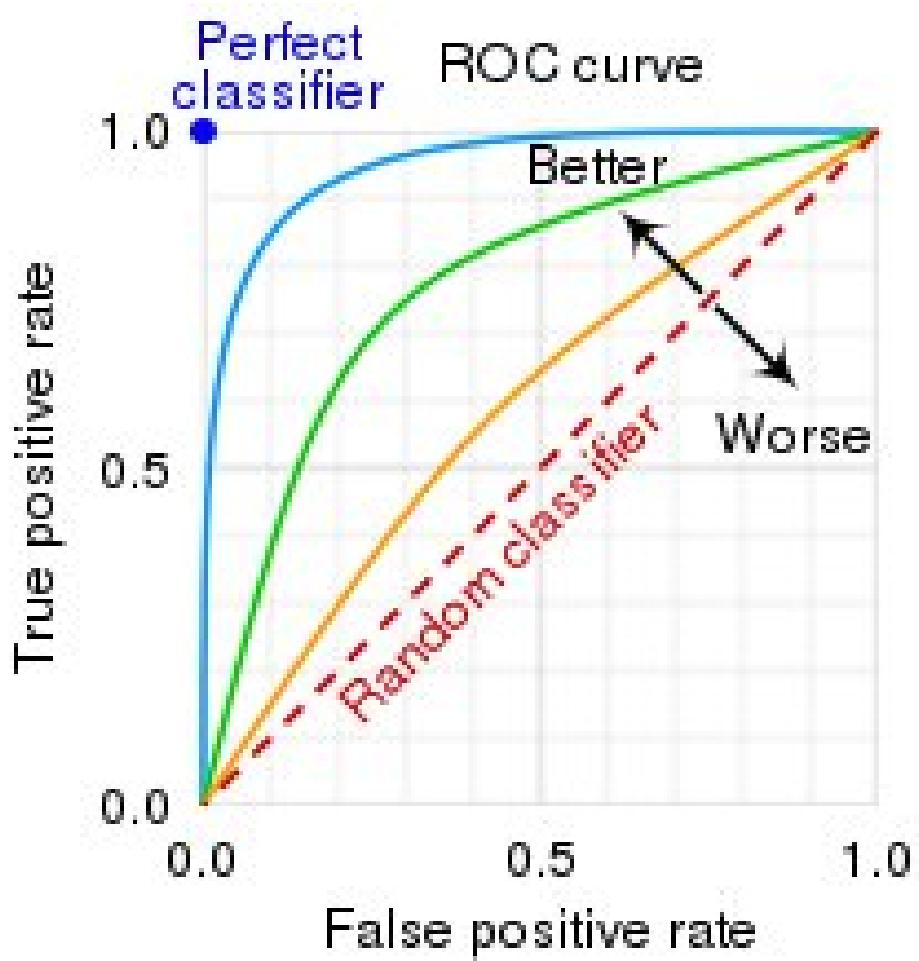
True class	
Predicted class	0
0	0
1	99

All negative :
accuracy = 99%

precision = nan
recall = 0



Metrics - Precision/Recall



Variation of the acceptance threshold of a class
to obtain the curve

$$TPR = \frac{TP}{TP + FN}$$

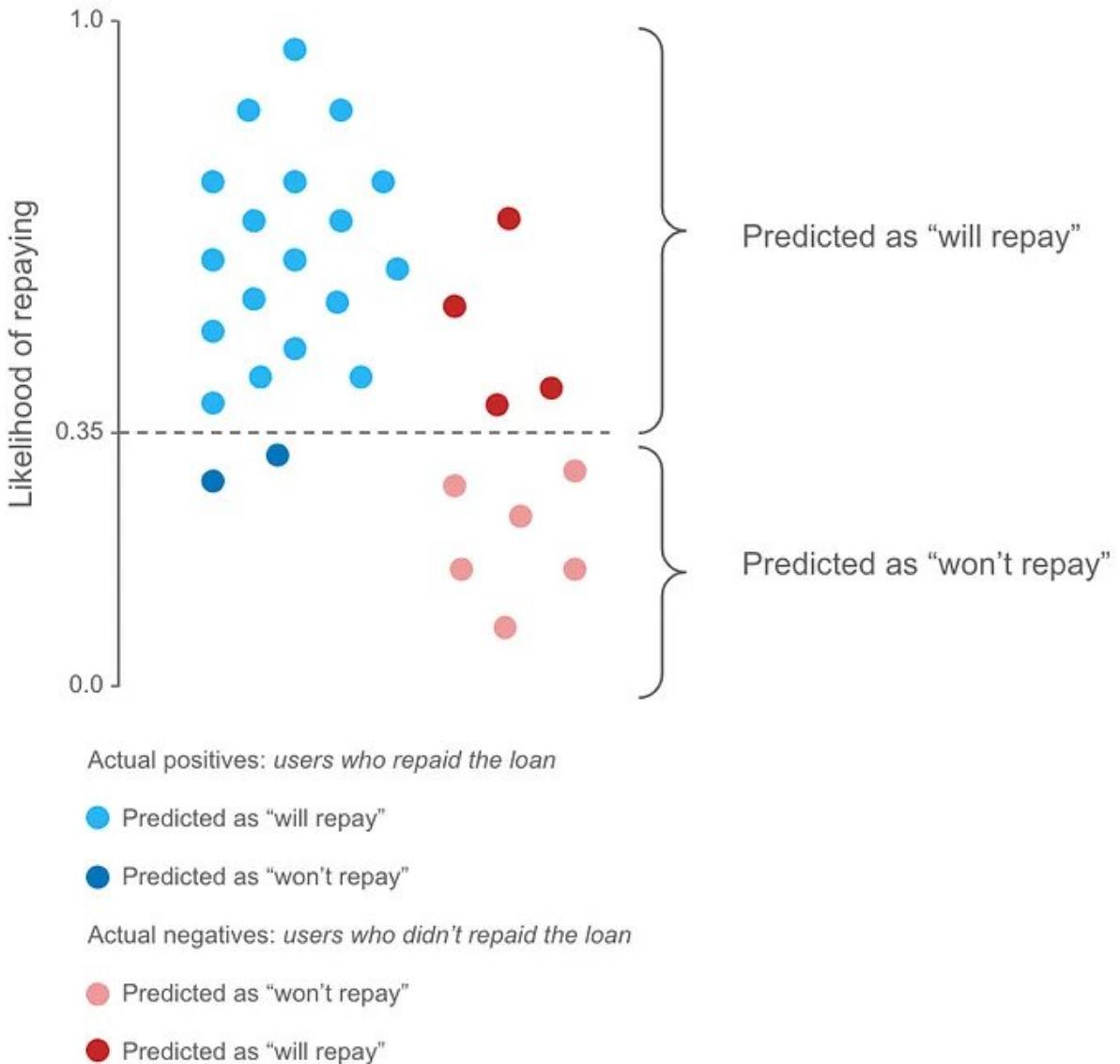
Above all positive data,
how many have been
predicted positive

$$FPR = \frac{FP}{FP + TN}$$

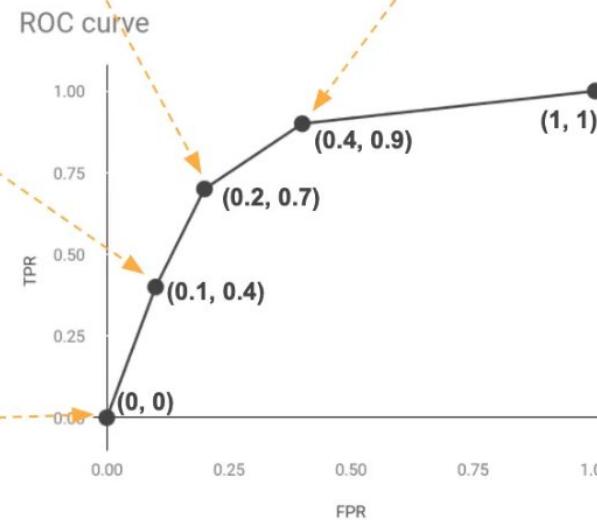
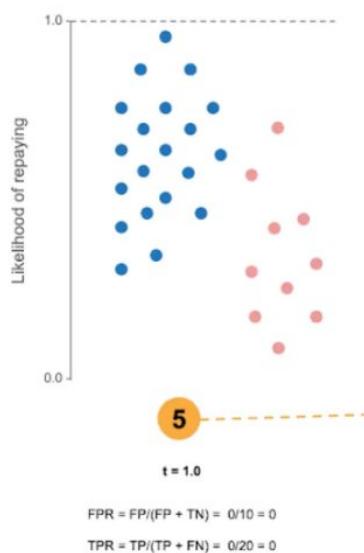
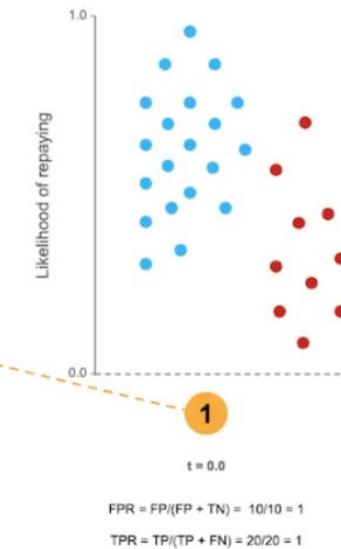
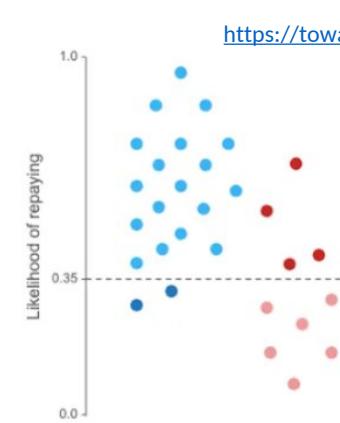
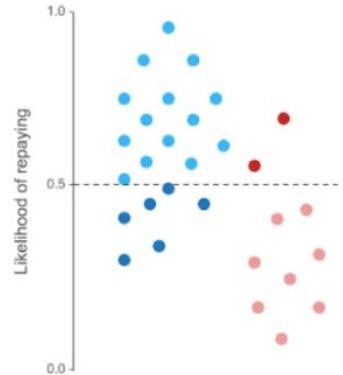
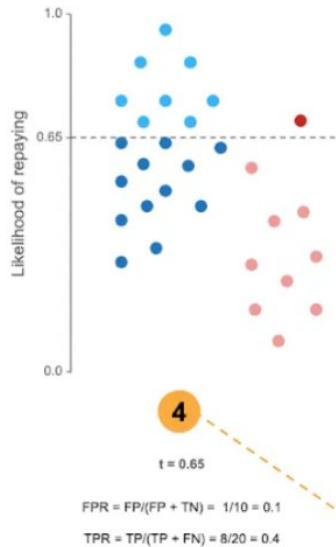
Above all negative data,
how many have been
predicted positive

NB : ROC = Receiver Operating Characteristics

Metrics - ROC curve



Metrics - ROC curve

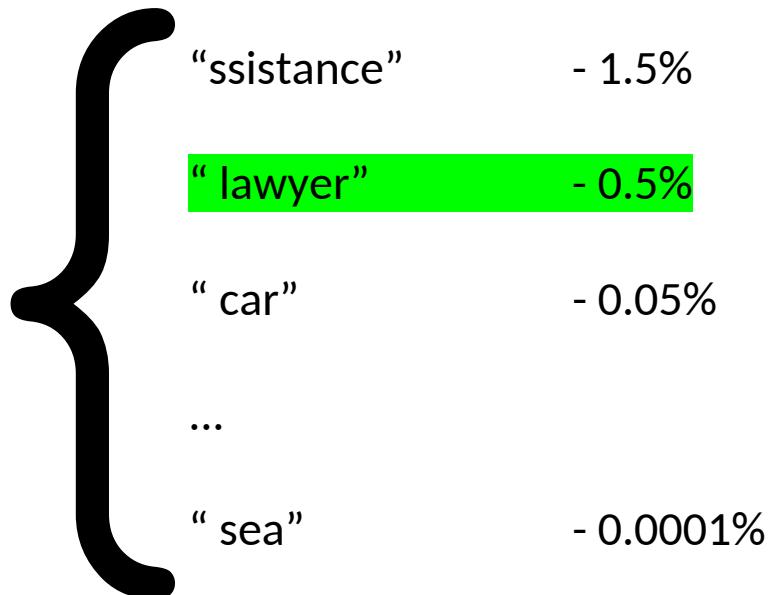


Metrics - ROC curve

Perplexity :
Is model surprised to see this text?

$$\text{PPL}(X) = \exp \left\{ -\frac{1}{t} \sum_i^t \log p_{\theta}(x_i | x_{<i}) \right\}$$

This person is innocent, she needs a **lawyer**



Perplexity is not All you Need !!

T	Model	Average	ARC	HellaSwag	MMLU	TruthfulQA	Winogrande	GSM8K	Architecture	#Params (B)
■	moreh/MoMo-70B-lora-1.8.6-DPO	77.29	70.14	86.03	77.4	69	84.37	76.8	LlamaForCausalLM	72.29
■	moreh/MoMo-70B-lora-1.8.4-DPO	76.23	69.62	85.35	77.33	64.64	84.14	76.27	LlamaForCausalLM	72.29
◆	TomGrc/FusionNet_7Bx2_MoE_14B	75.91	73.55	88.84	64.68	69.6	88.16	70.66	MixtralForCausalLM	12.88
◆	Weyaxi/Helion-4x34B	75.48	69.71	85.28	77.33	63.91	84.37	72.25	MixtralForCausalLM	113.66
◆	one-man-army/UNA-32Beagles-32K-bf16-v1	75.41	73.55	85.93	76.45	73.55	82.95	60.05	LlamaForCausalLM	34.39
◆	Weyaxi/Cosmosis-3x34B	75.39	69.71	85.18	77.25	63.82	84.14	72.25	MixtralForCausalLM	87.24
◆	Weyaxi/Bagel-Hermes-2x34b	75.1	69.8	85.26	77.24	64.82	84.77	68.69	MixtralForCausalLM	60.81
○	jondurbin/bagel-dpo-34b-v0.2	74.69	71.93	85.25	76.58	70.05	83.35	60.96	LlamaForCausalLM	34.39
◆	jondurbin/nontoxic-bagel-34b-v0.2	74.69	72.44	85.64	76.41	72.7	82.48	58.45	LlamaForCausalLM	34.39
◆	moreh/MoMo-70B-LoRA-V1.4	74.67	69.2	85.07	77.12	62.66	83.74	70.2	LlamaForCausalLM	72.29
■	udkai/Turdus	74.66	73.38	88.56	64.52	67.11	86.66	67.7	MistralForCausalLM	7.24

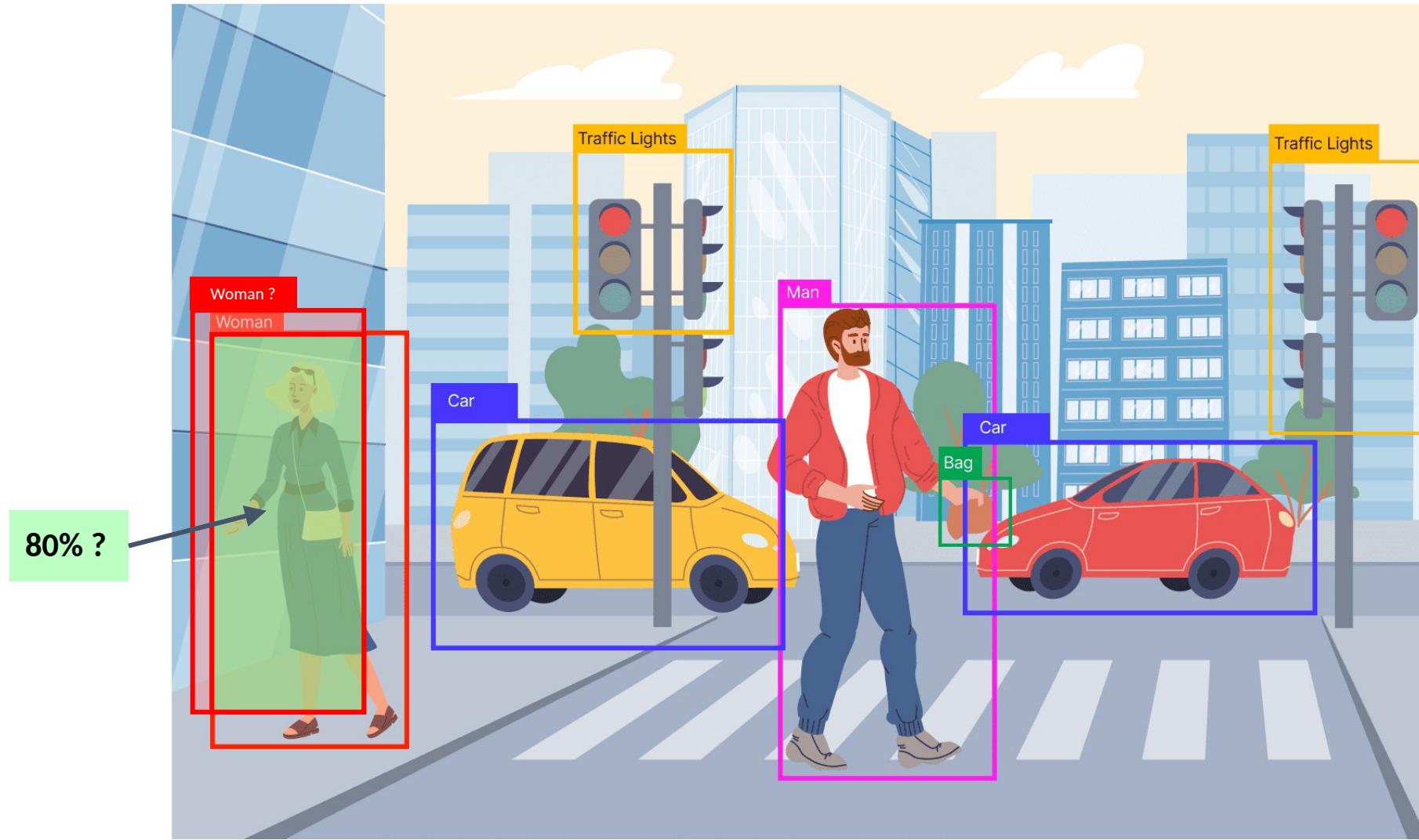
Language models evaluation benchmark

The evaluation is complicated, there are many different tasks

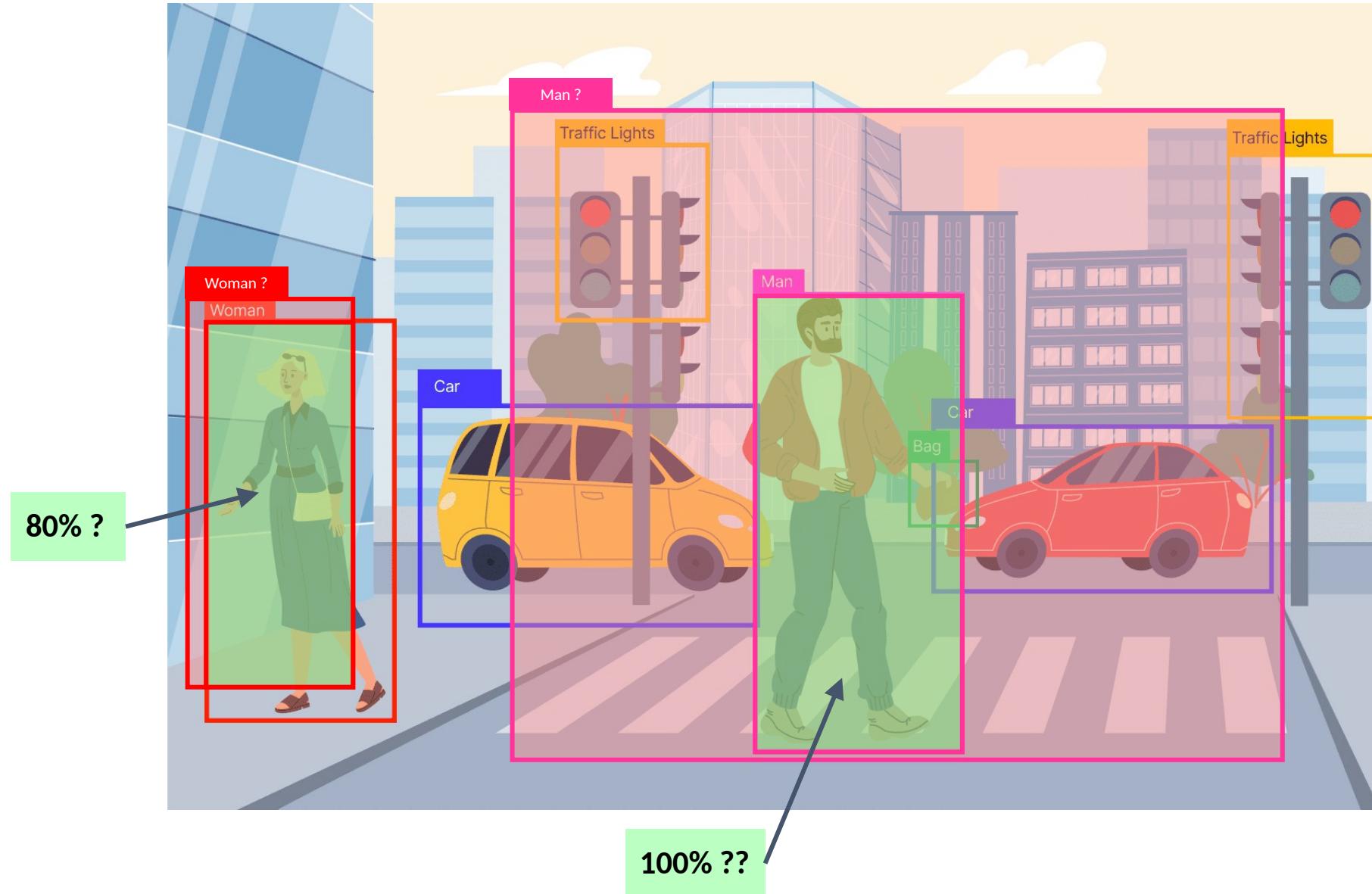
It can depends on the targeted application

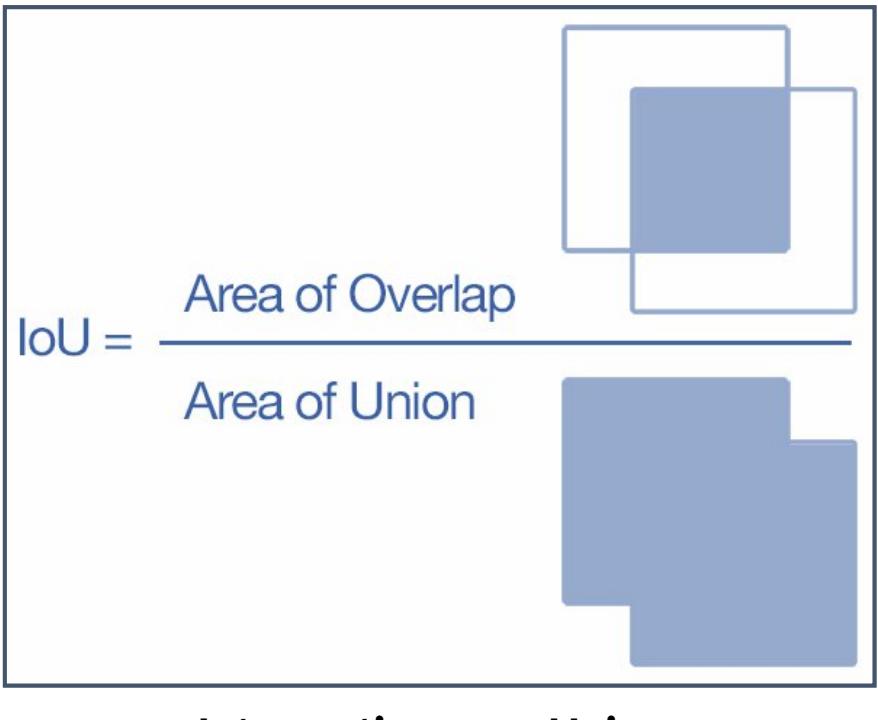
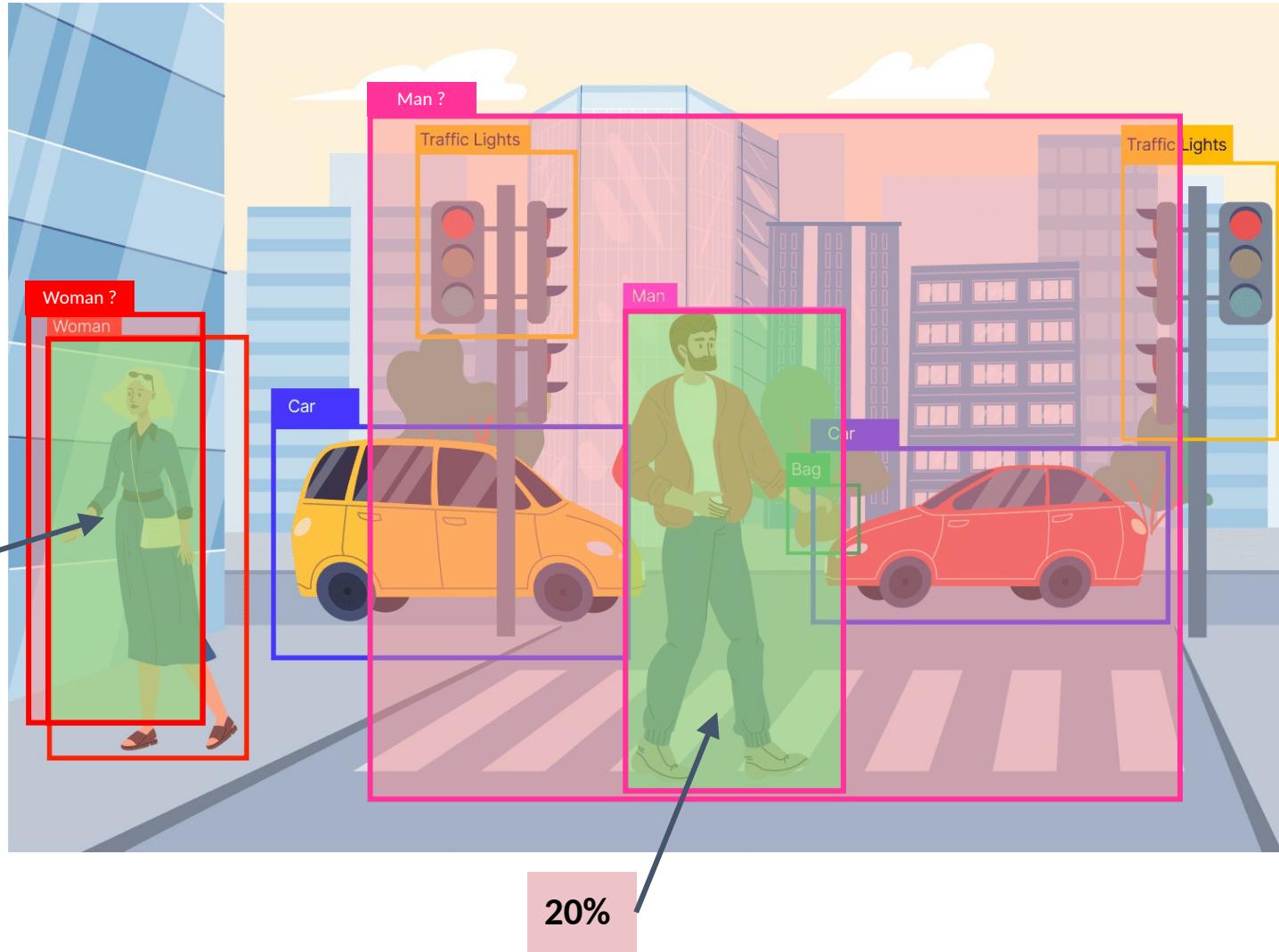
It is hard to be fair/objective without knowing the training data

Metrics - Language models evaluation



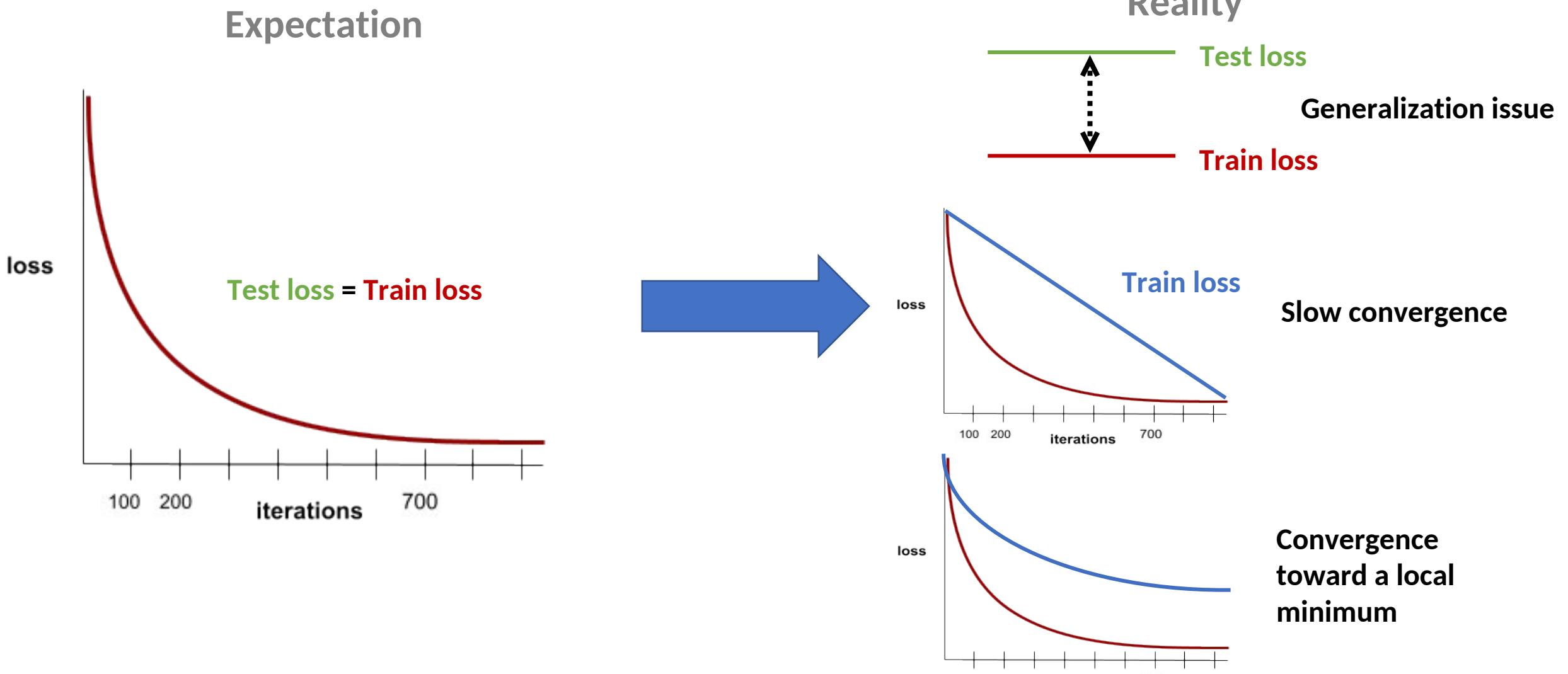
Metrics - Intersection over Union





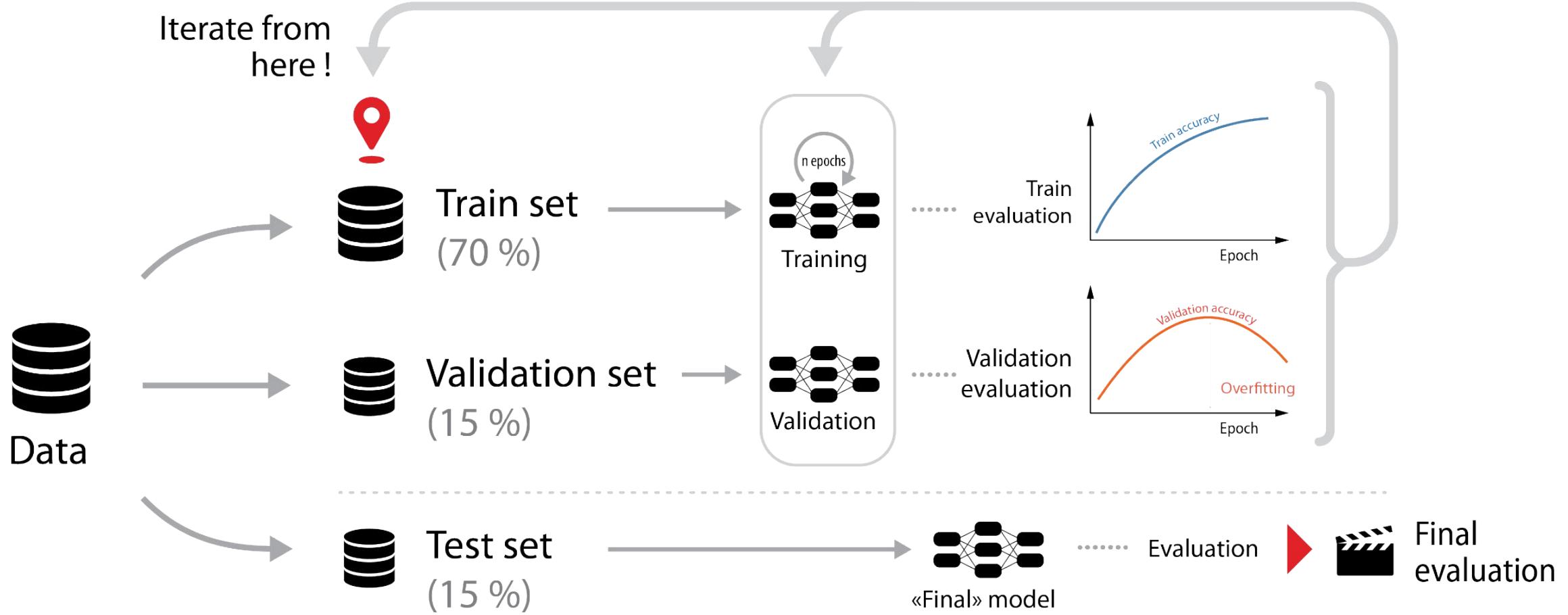
Metrics - Intersection over Union

- 1** Quel travail faire pour améliorer les données utilisées pour l'entraînement
- 2** Comment évaluer un modèle ?
- 3** Est-il possible de rendre l'entraînement plus robuste ?
- 4** Peut-on profiter d'un modèle déjà entraîné ?
- 5** Bonus : Quelques bonnes pratiques ?



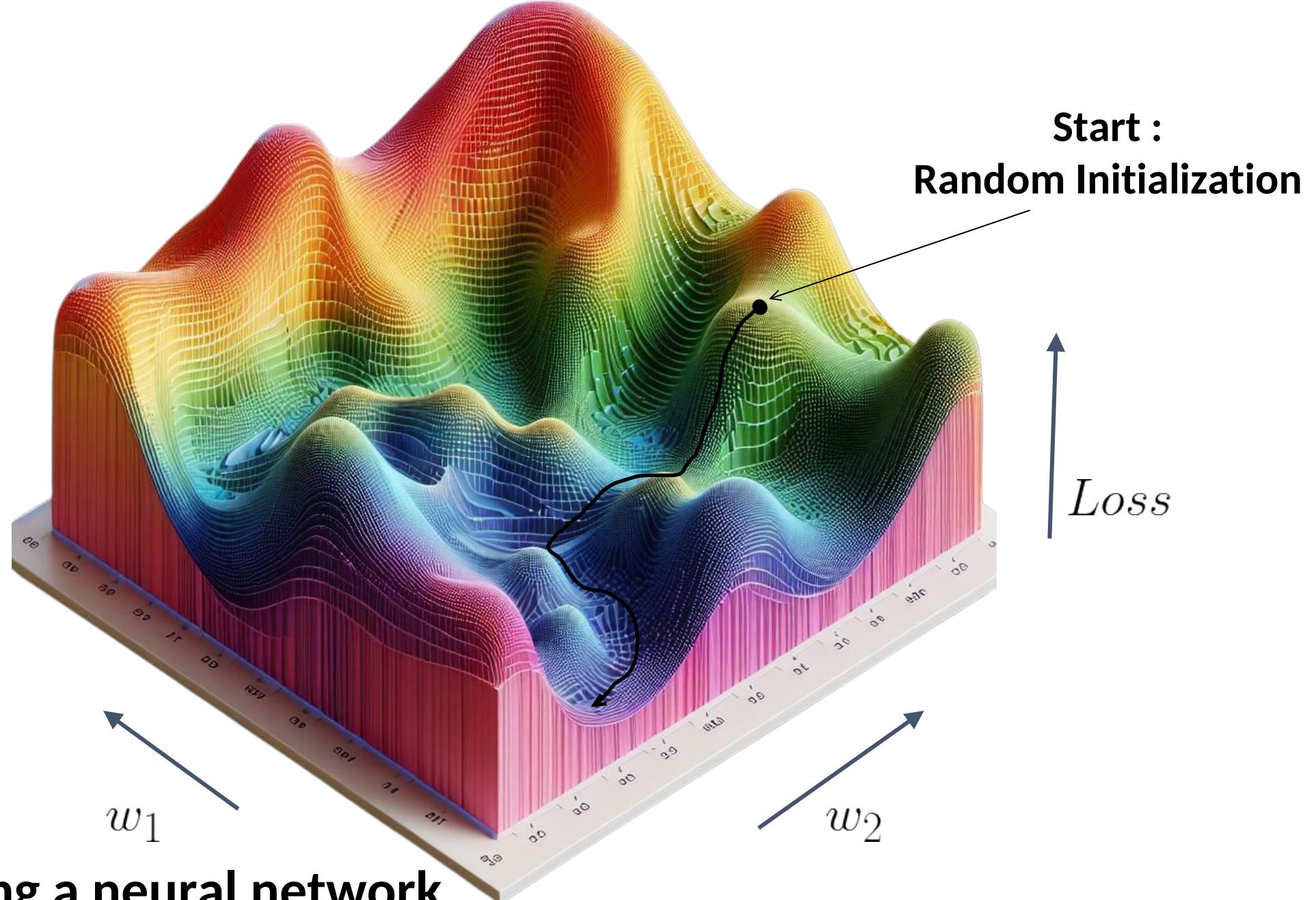
What happens during training ?

C Update model

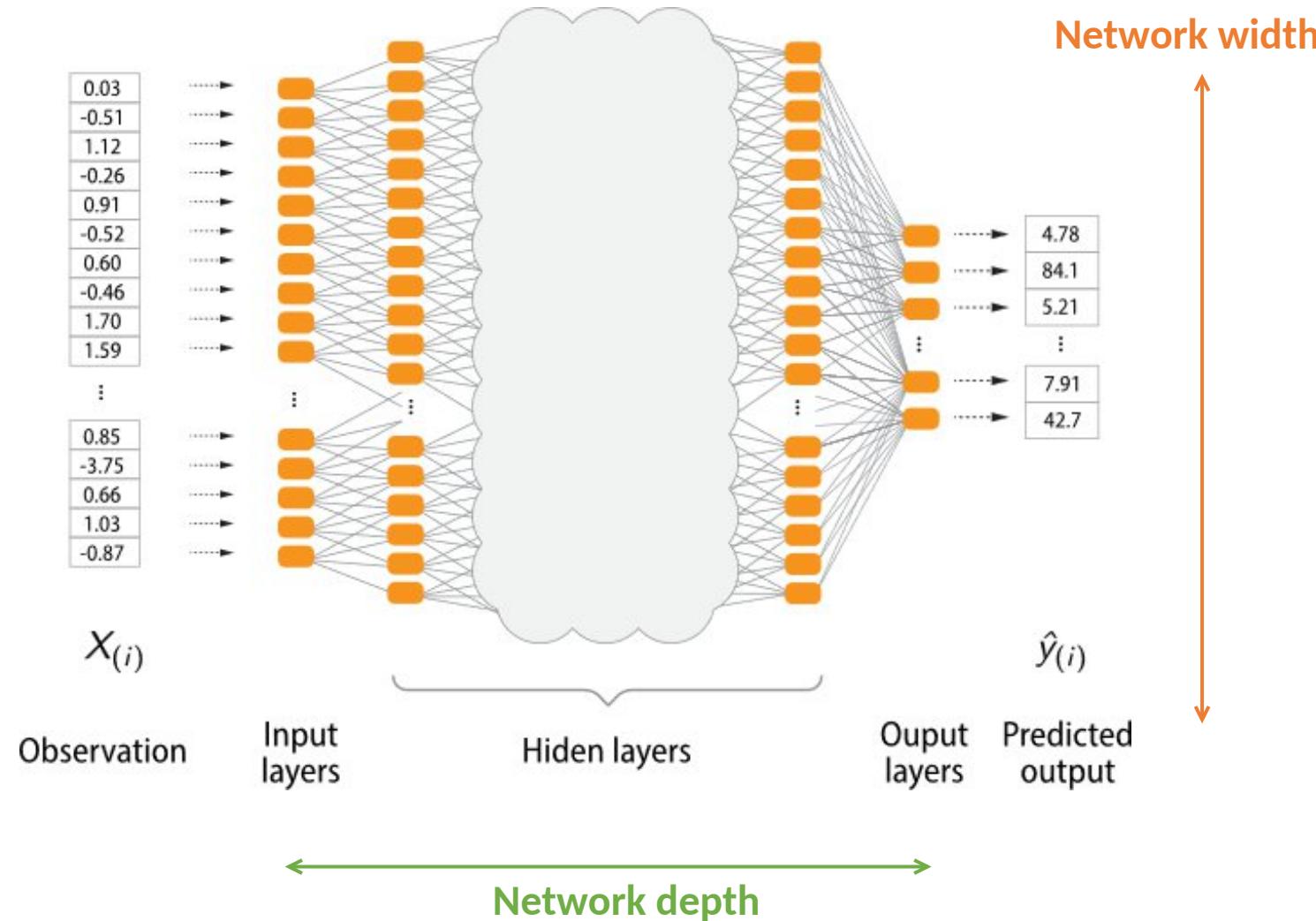


Data splitting

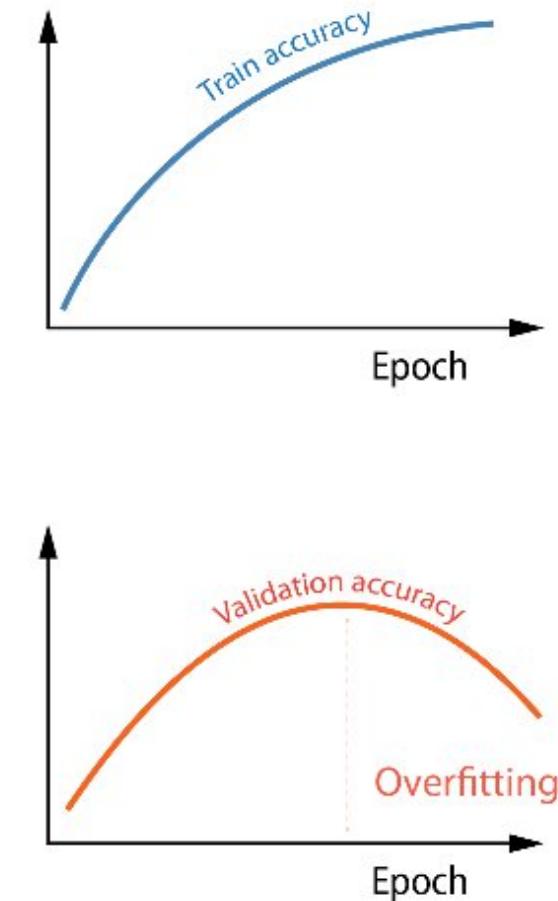
Gradient Descent



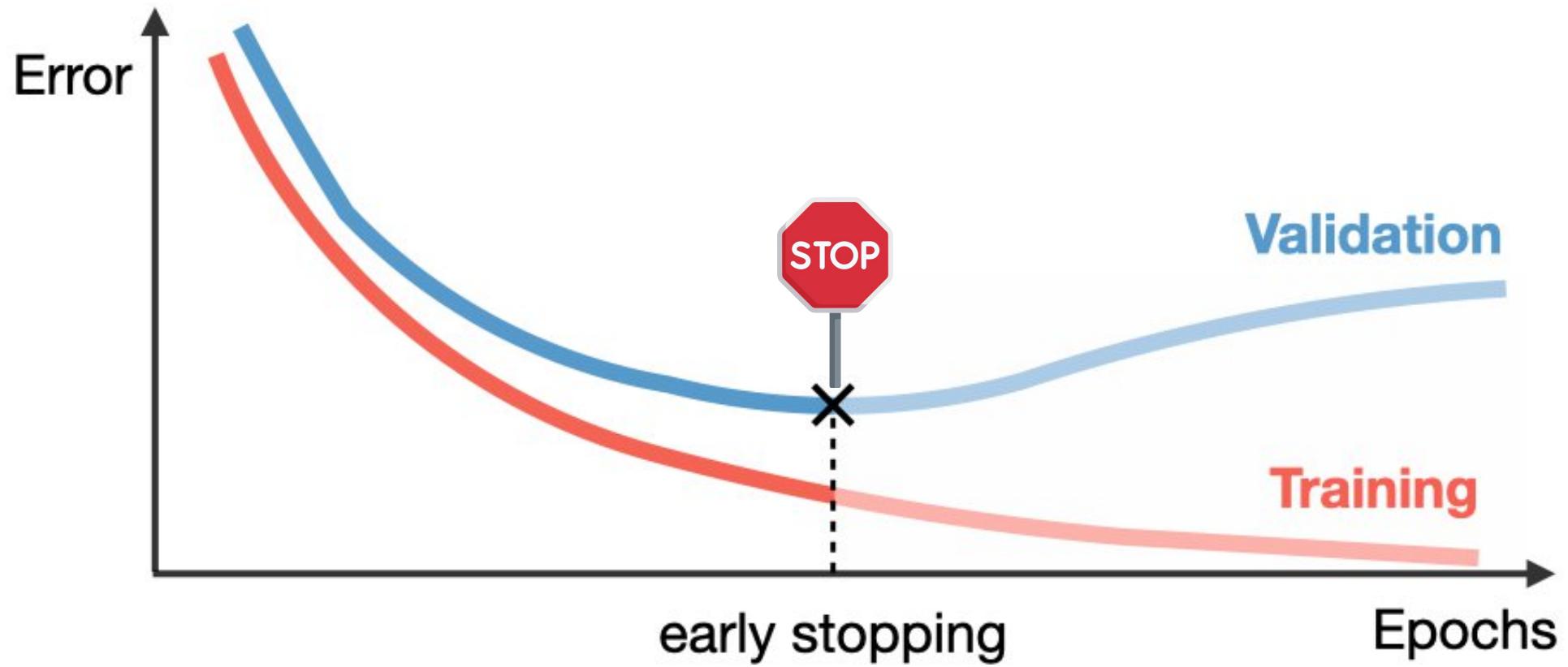
Training a neural network



Network width



Model architecture and overfitting



Basic way : Early stopping

Hyper-parameters

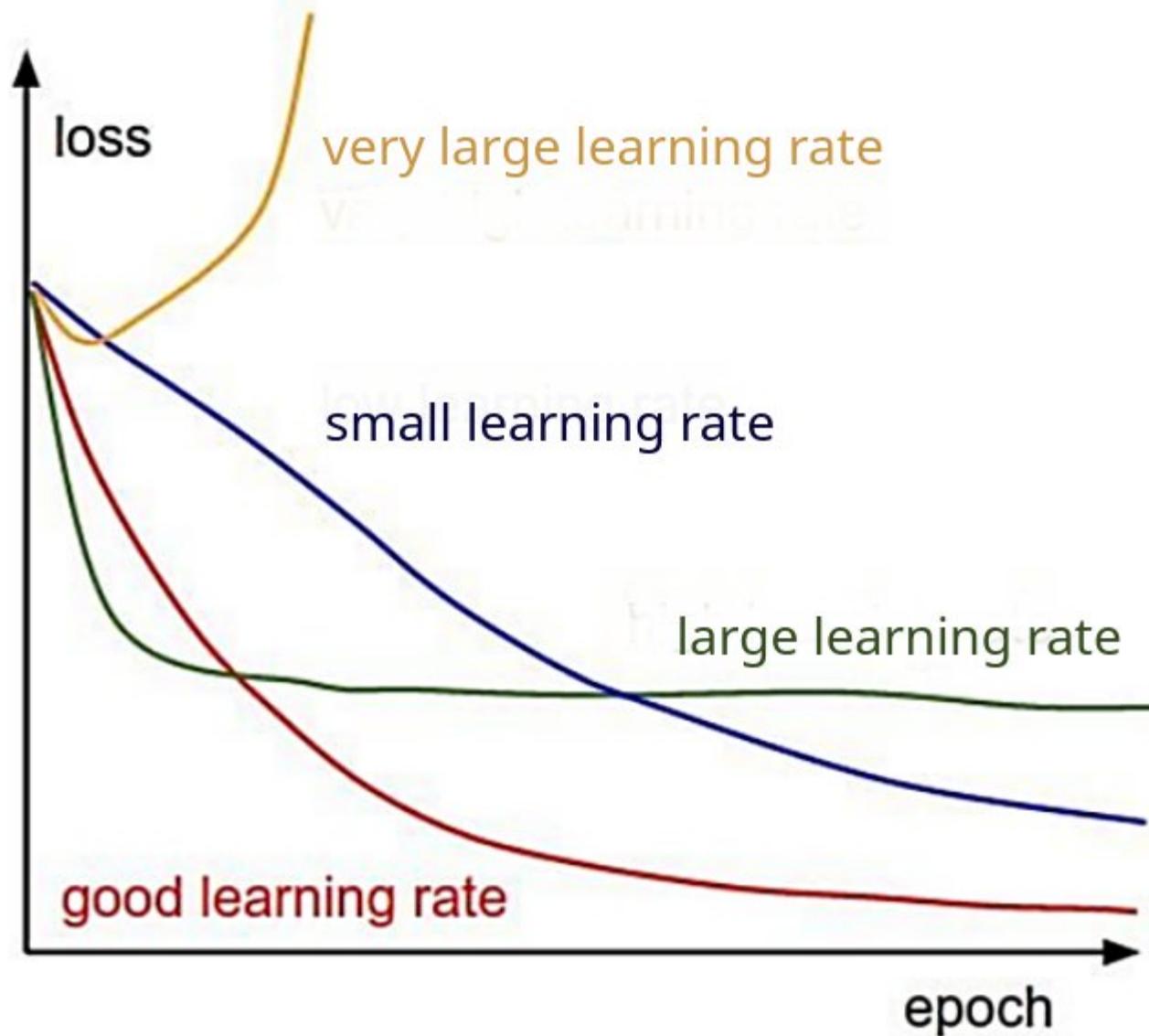
- Learning rate
- Regularization
- Optimizer
- Model architecture
- Batch size
- ...

$$\Theta_{t+1} = \Theta_t - \eta \nabla_{\Theta} [\mathcal{L}(\hat{y}_i, y_i)]$$

Updated weights = Weights before update — Learning rate * Gradient [Cost function (Prediction, Label)]

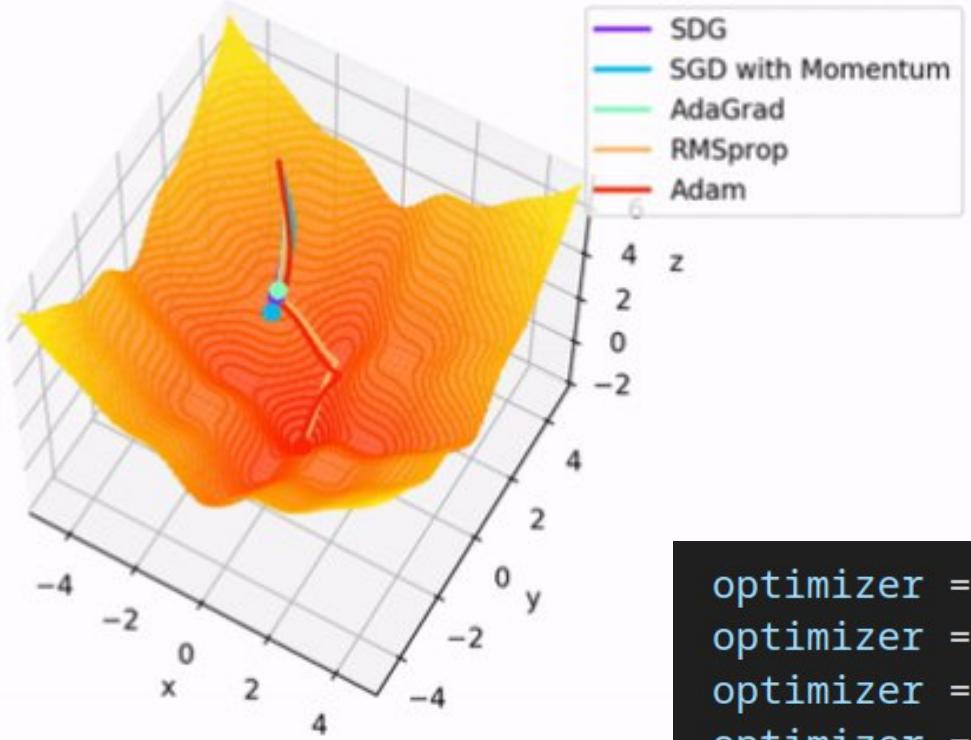
Weight update equation

Regularization - Loss function and weight update



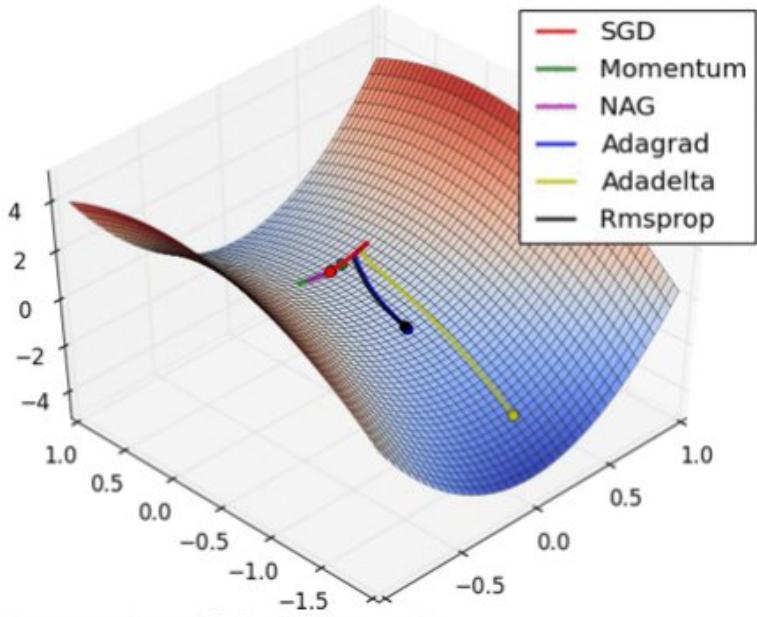
Learning rate

Optimizers

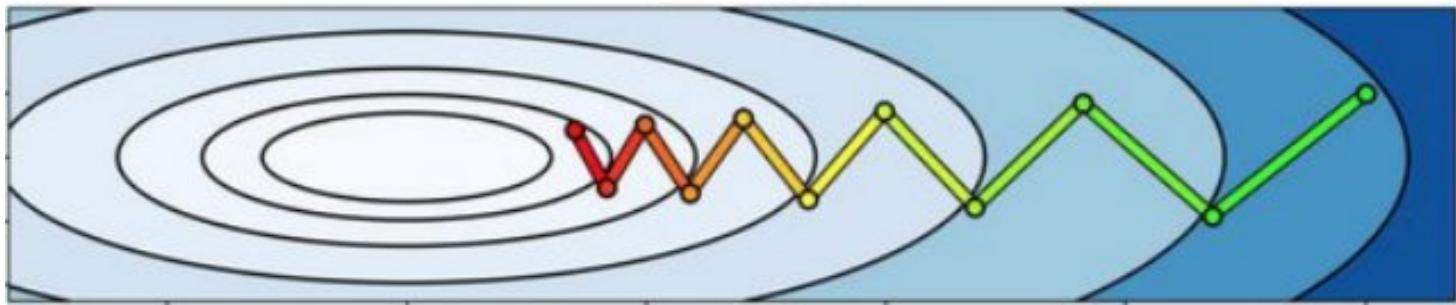


```
optimizer = torch.optim.SGD  
optimizer = torch.optim.SGD with Momentum  
optimizer = torch.optim.Adagrad  
optimizer = torch.optim.RMSprop  
optimizer = torch.optim.Adam
```

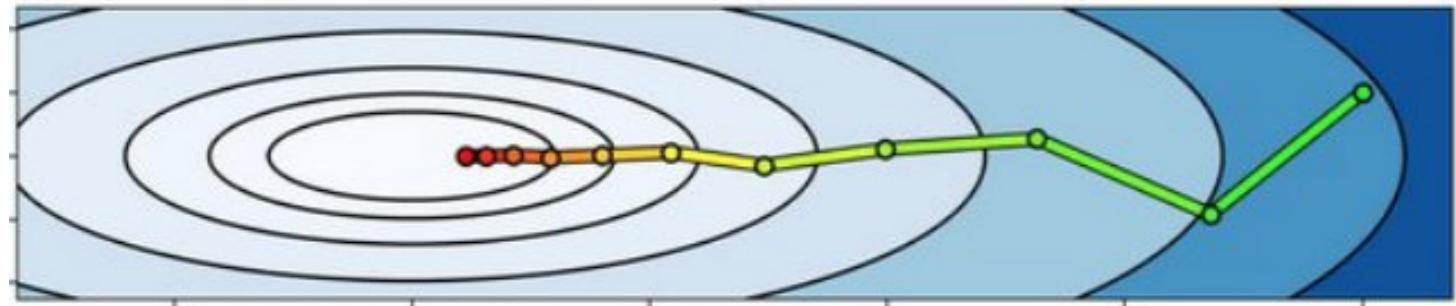
https://www.makerluis.com/content/images/2023/12/Advanced_Optimizations_Saddle_optim-1.gif
<https://medium.com/@LayanSA/complete-guide-to-adam-optimization-1e5f29532c3d>



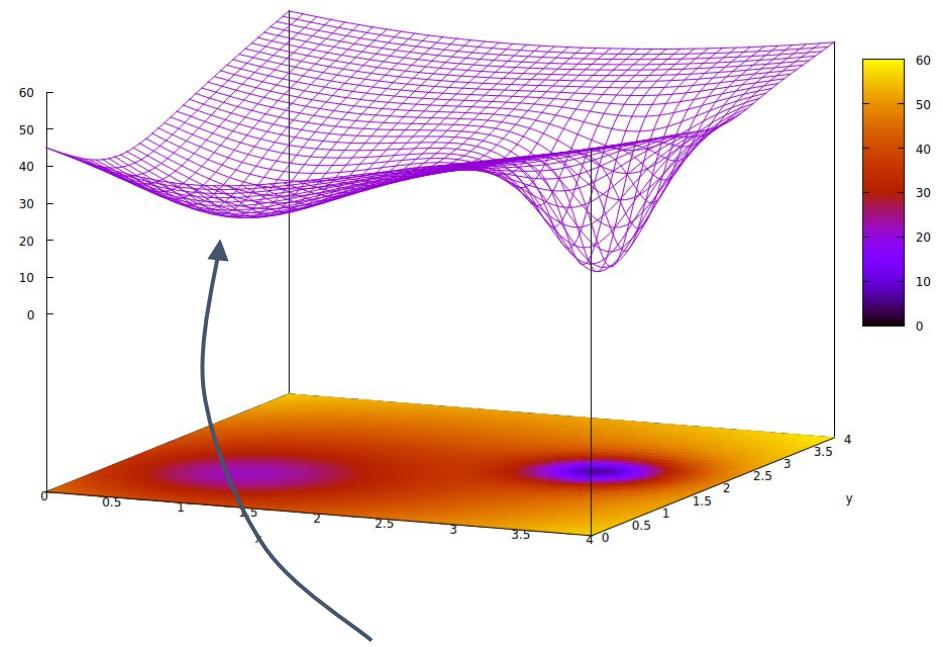
Convergence



Without momentum



With momentum



Local minimum

$$\Theta_{t+1} = \Theta_t - \eta \nabla_{\Theta} [\mathcal{L}(\hat{y}_i, y_i) + \lambda R(\Theta_t)]$$

Updated weights = Weights before update - Learning rate * Gradient [Cost function (Prediction, Label) + Regularization rate * Regularization function (Weights before update)]

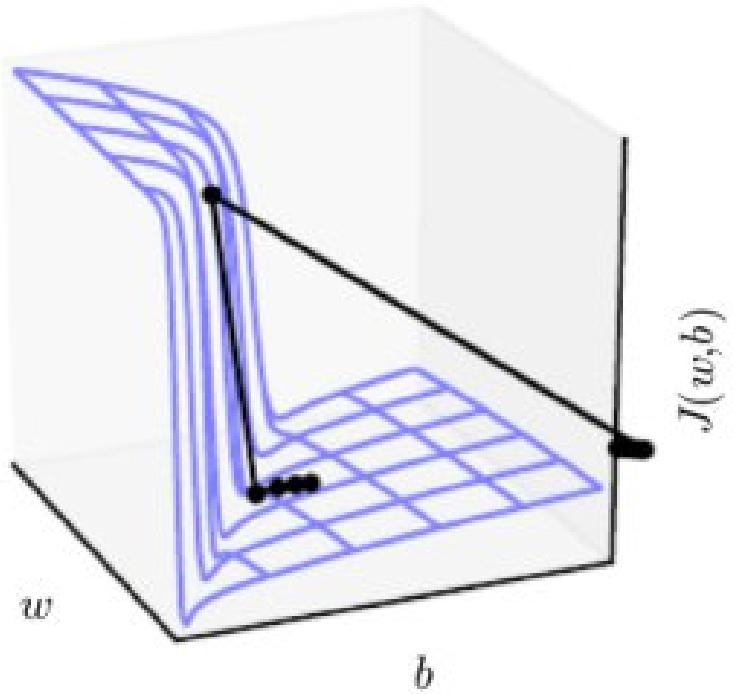
Weight update equation

- L1 Regularization
- L2 Regularization
- Max norm Regularization
- Regularization with the cost function
- Dropout

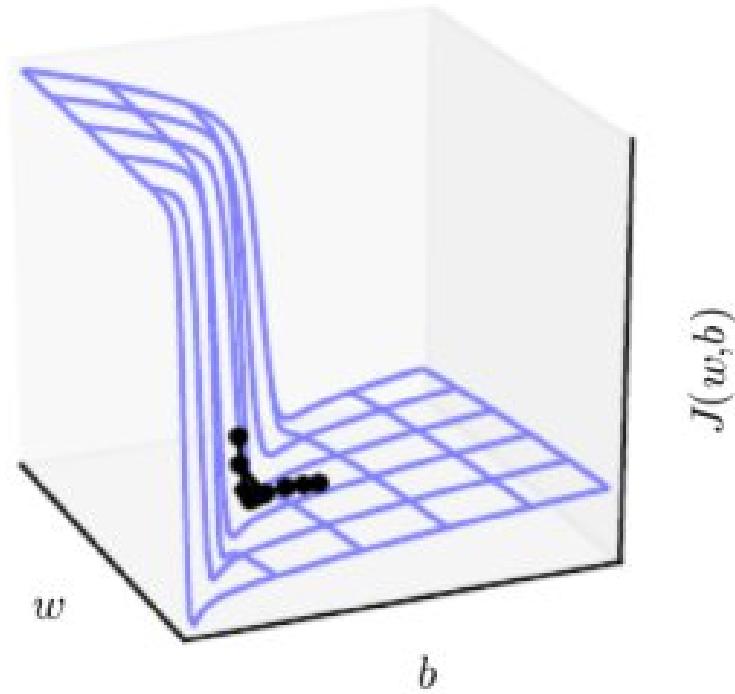
L1 : LASSO	L2 : Ridge
$ \Theta $	Θ^2

Regularization : L1 and L2

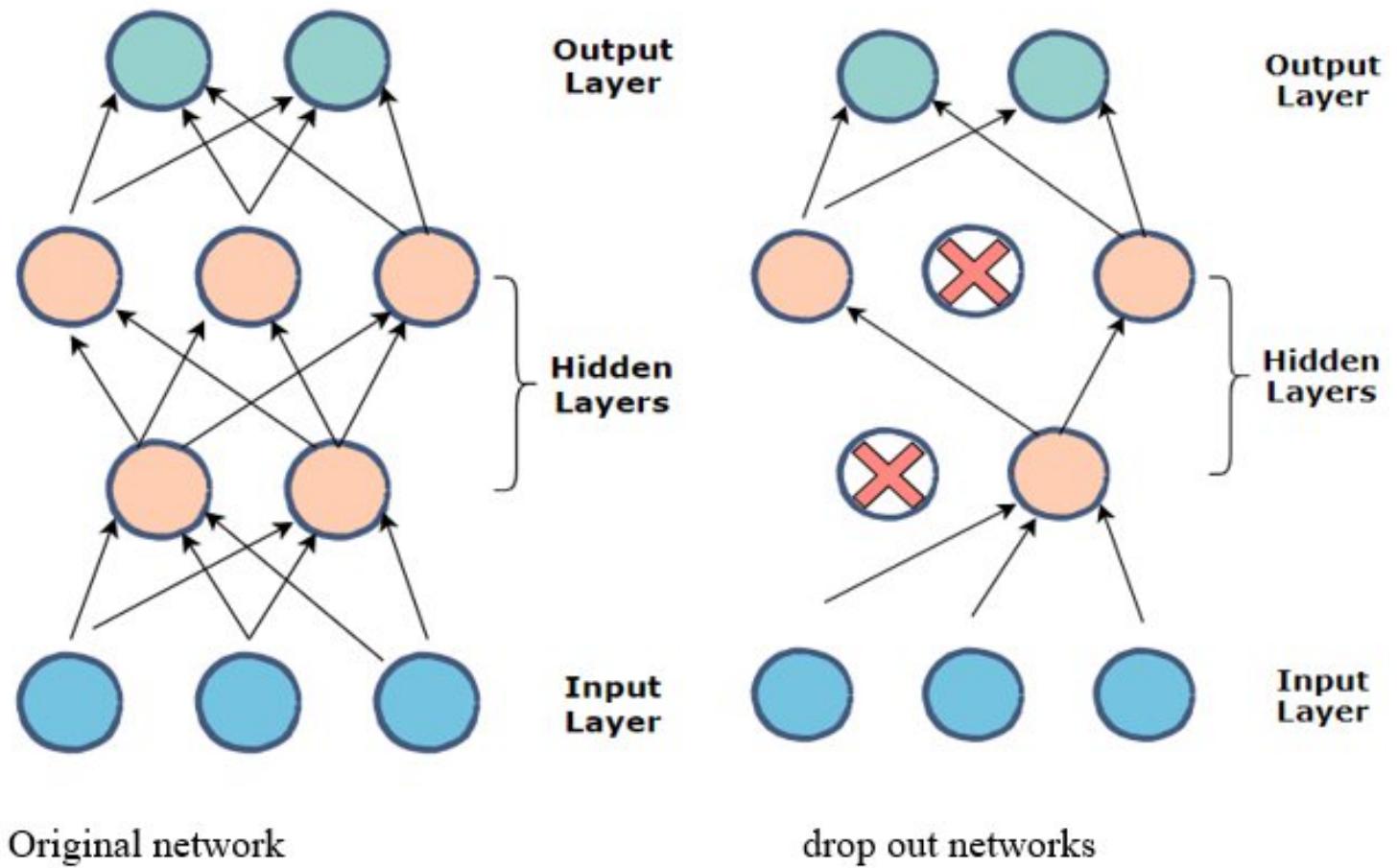
Without clipping



With clipping



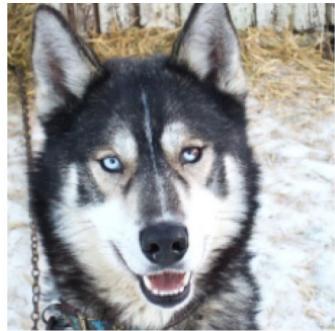
Regularization : Gradient clipping



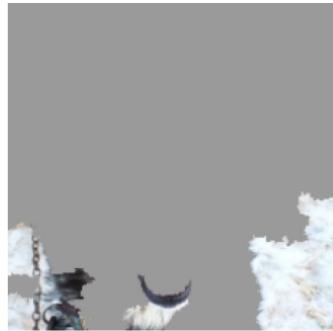
Regularization : Dropout



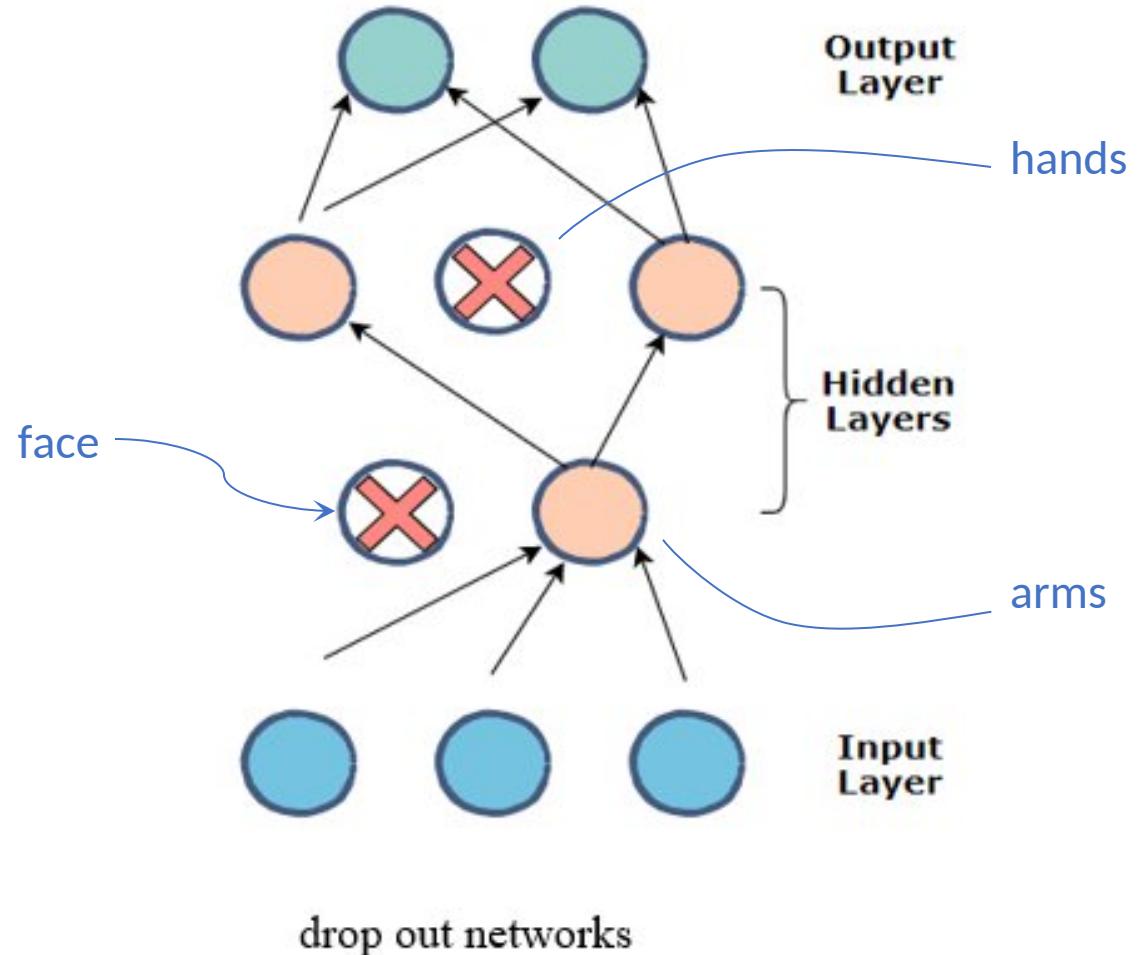
Human = face ?



(a) Husky classified as wolf



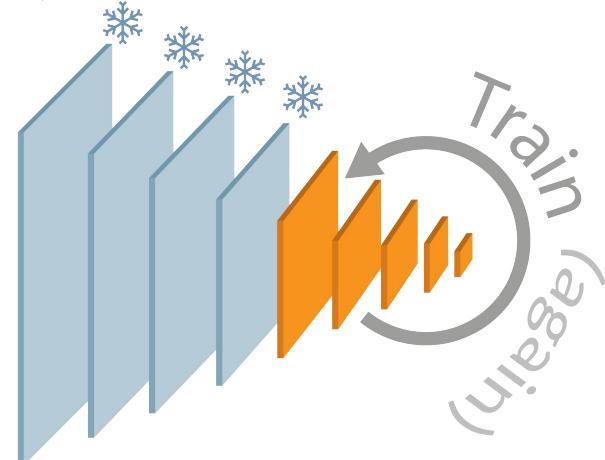
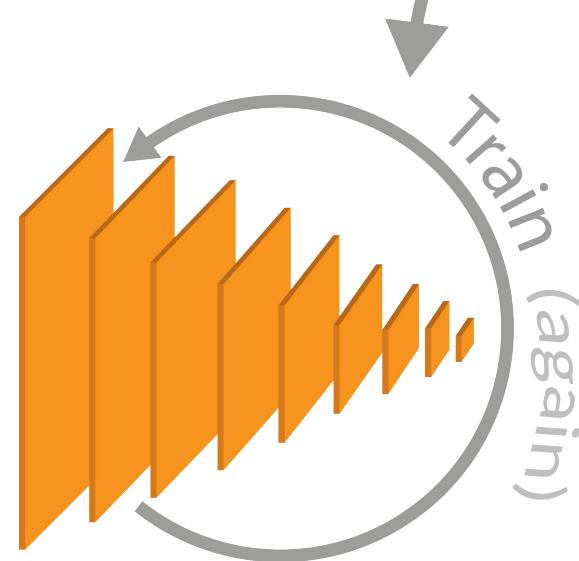
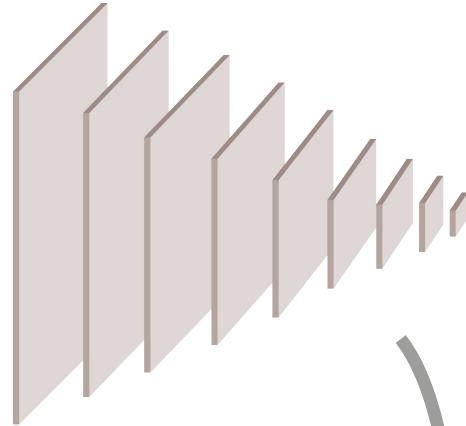
(b) Explanation



Regularization : Dropout

- 1** Quel travail faire pour améliorer les données utilisées pour l'entraînement ?
- 2** Comment évaluer un modèle ?
- 3** Est-il possible de rendre l'entraînement plus robuste ?
- 4** Peut-on profiter d'un modèle déjà entraîné ?
- 5** Bonus : Quelques bonnes pratiques ?

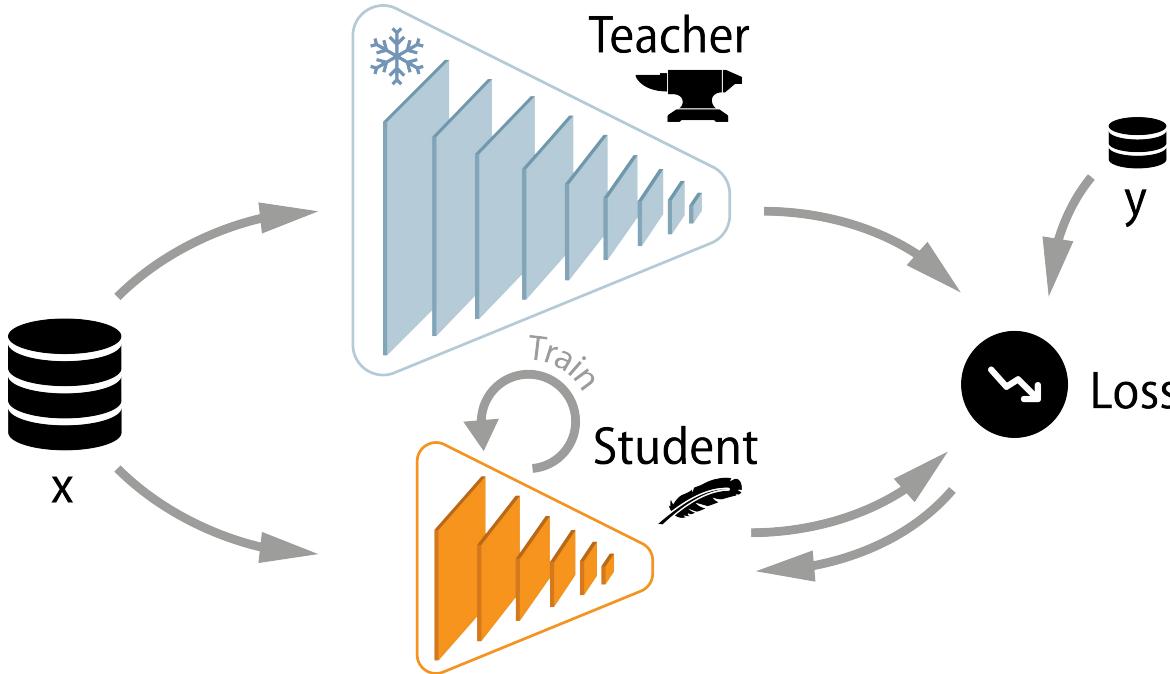
Pretrained
model



Hugging Face

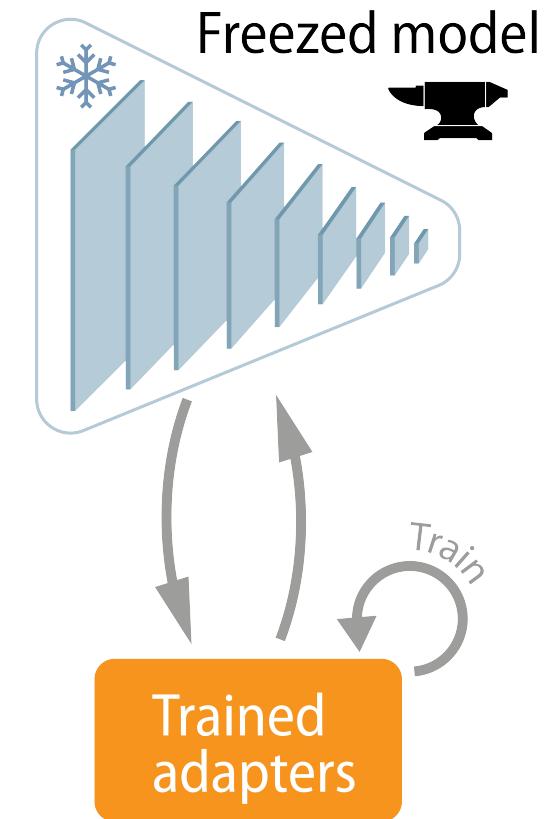
Transformers
Timm





Knowledge distillation

Transfer learning



Adapters



PEFT

Parameter-Efficient
Fine-Tuning

- 1** Quel travail faire pour améliorer les données utilisées pour l'entraînement ?
- 2** Comment évaluer un modèle ?
- 3** Est-il possible de rendre l'entraînement plus robuste ?
- 4** Peut-on profiter d'un modèle déjà entraîné ?
- 5** Bonus : Quelques bonnes pratiques ?

Hyper-parameters

- Learning rate
- Regularization
- Optimizer
- Model architecture
- Batch size
- ...

Methods

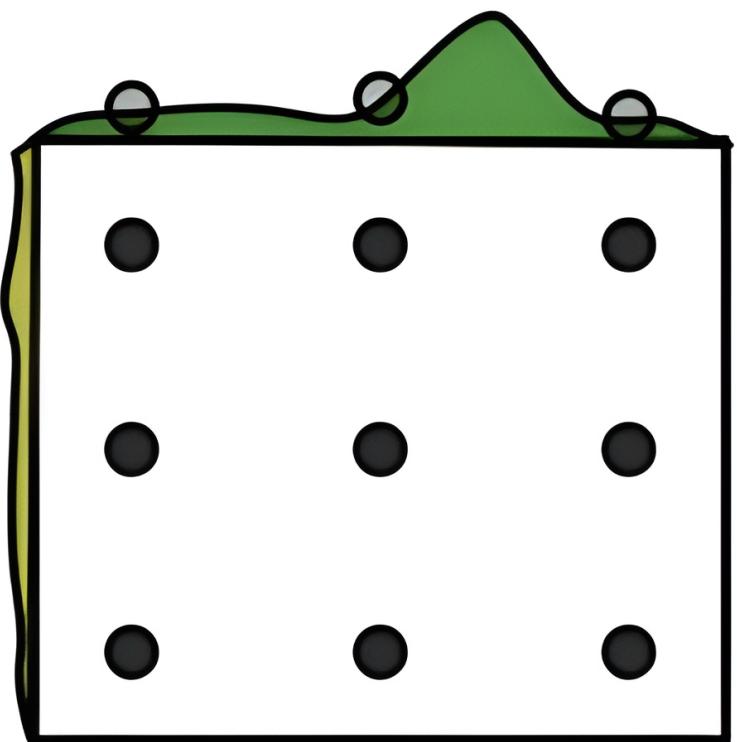
- Manual research
- Grid search
- Random research
- Gradient
- Evolutionary algorithms
- ...

HPO (Hyperparameter Optimization)

Find the good hyper-parameters

Grid Layout

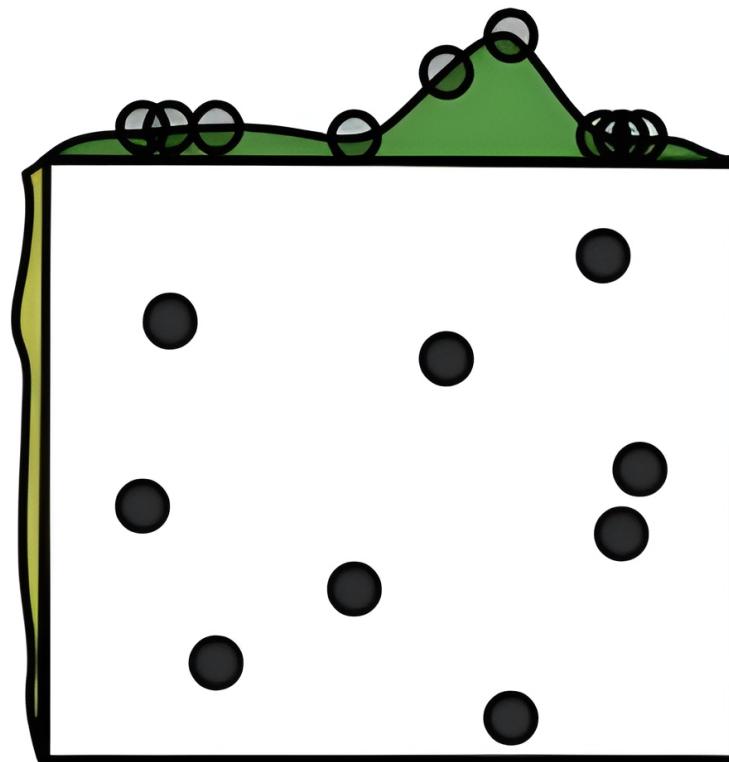
Unimportant parameter



Important parameter

Random Layout

Unimportant parameter



Important parameter

HPO : Grid & Random

Bayesian Optimization in Action

<https://maelfabien.github.io/assets/images/bo.gif>

<https://maelfabien.github.io/machinelearning/Explorium4/#acquisition-function>

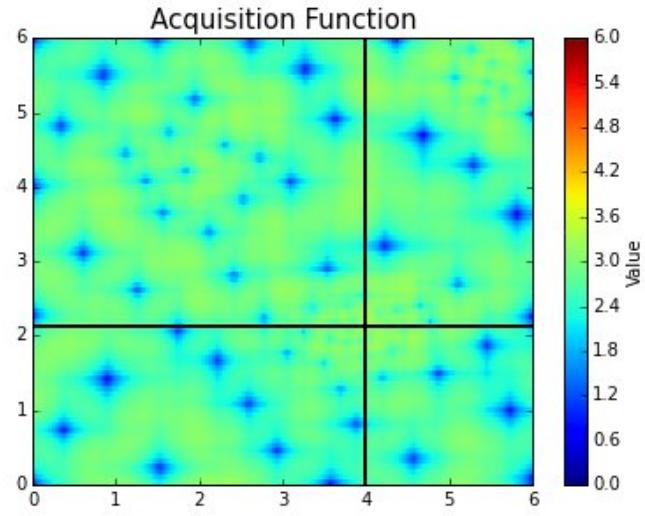
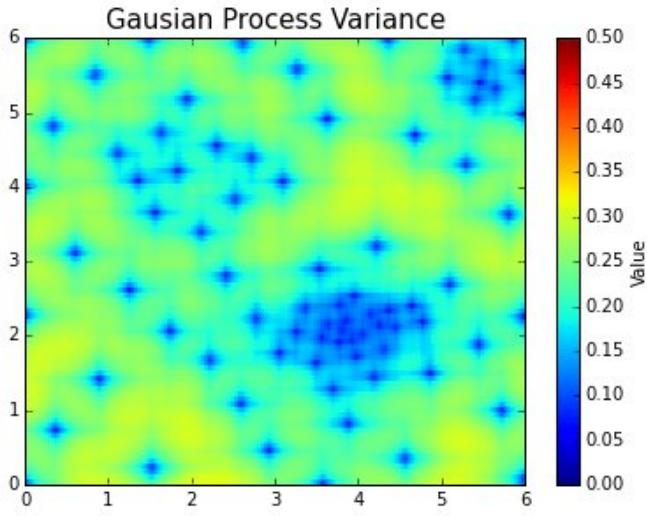
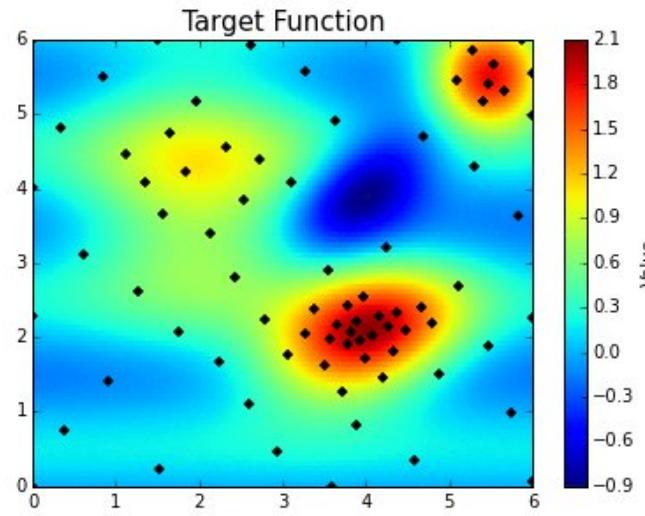
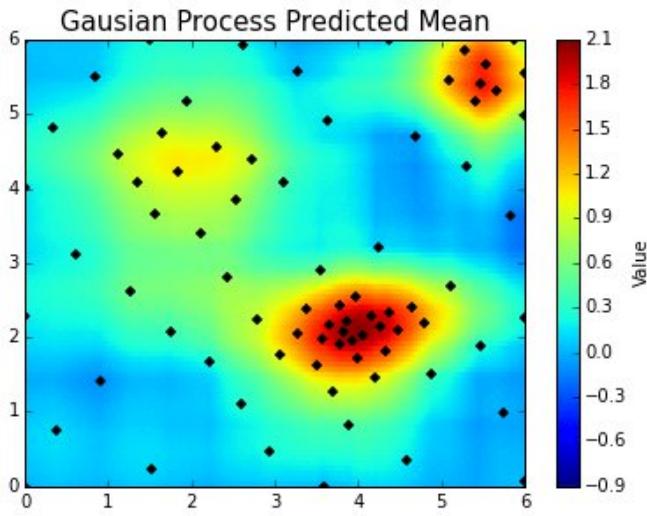


Table view

Chart view

metrics.rmse < 1 and params.model = "tree"

i

Sort: Created

Columns

⋮

Refresh

Time created: All time

State: Active

		Metrics			Parameters							
	Run Name	Created	🕒	Duration	test_accuracy	training_loss	batch_size	epochs	learning_rate	momentum	weight_clampir	world_size
☐	🕒 adorable-flea-594	🕒 1 day ago		1.5min	58.73	0.935	32	10	0.01	0.9	False	2
☐	🕒 righteous-calf-205	🕒 1 day ago		1.5min	61.84	0.738	32	10	0.01	0.9	True	2
☐	🕒 loud-dog-130	🕒 1 day ago		2.8min	55.87	1.303	32	10	0.01	0.9	True	1
☐	🕒 industrious-yak-917	🕒 1 day ago		2.9min	57.57	0.915	32	10	0.01	0.9	False	1



Collaboration & Reproducibility

Gitlab Org > Issue Boards

Development < Search or filter results... Show labels Group by None Edit board Create list

> Open □ 56 □ 10 +

Milestones swimlanes Doing dev plan gitlab-org/gitlab-test#80 Jun 17 □ 4

Assign issue to epic To Do backend gitlab-org/gitlab-test#45

Create group To Do gitlab-org/gitlab-test#78 □ 3

Remove issue from board To Do gitlab-org/gitlab-shell#38

Add lists for assignees and milestones To Do gitlab-org/gitlab-shell#2

> Deliverable □ 32 □ 8 +

Update issue due date from sidebar To Do frontend gitlab-org/gitlab-test#54

Update issue's labels from sidebar To Do frontend gitlab-org/gitlab-test#55

Update issue labels Doing gitlab-org/gitlab-test#75

Drag and drop issue between epics To Do frontend gitlab-org/gitlab-test#33

Paginate issues in Swimlanes To Do gitlab-org/gitlab-test#39

> Closed □ 59 □ 2

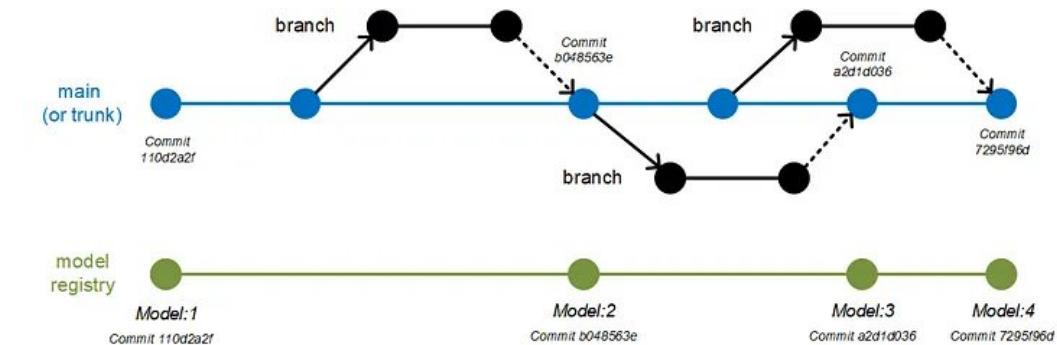
Persist collapsed state of Swimlanes Deliverable test test gitlab-org/gitlab-test#32 Dec 30, 2020 □ 1w □ 2

Remove list from board Caliber Colorado Doing feature proposal gitlab-org/gitlab-test#44

Remove issue from Swimlane To Do frontend gitlab-org/gitlab-test#36

Expand diff to entire file Premium To Do dev manage frontend gitlab-org/gitlab-test#25

Laboriosam commodi ab in eum qui suscipit necessitatibus modi fuga. Deliverable frontend gitlab-org/gitlab-test#2



A structured project

