

Doctoral course DSG'2019

Data Science for Geosciences

Project session

Brest, January 14-19 2019



Project Organization











General objectives: Practice of machine learning tools for a real application

- Statement of a given issue as a machine learning problem
- Implementation of a machine learning model in Python (training & testing stage)
- Evaluation and benchmarking of machine learning models
- **Deliverable: Python notebook**

Organization:

- 10 groups of 2 trainees
- 5 topics, 3 to 4 groups for each topic
- Five project sessions

 Course 

	Monday	Tuesday	Wednesday	Thursday	Friday
Morning					
Afternoon					

Topics

1. Prediction of ENSO (El Niño-Southern Oscillation)
2. Recognition of ocean-atmosphere events in satellite SAR images
3. Analysis and dimensionality reduction of Sea Surface Height
4. Automatic fish sound recognition
5. Land cover classification from hyperspectral data

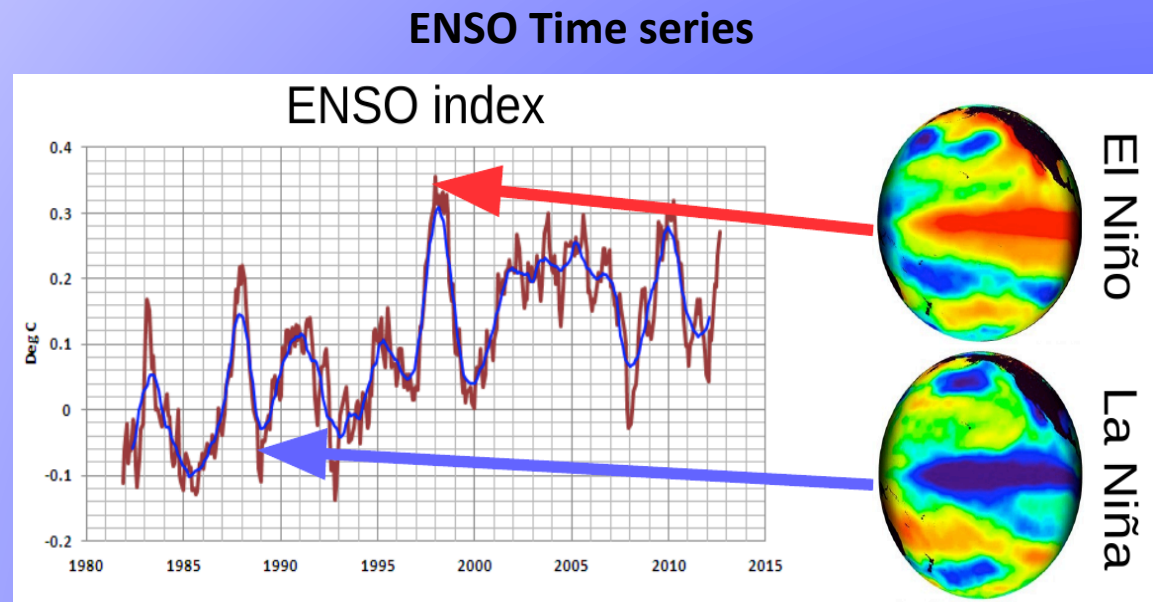
All project resources available at

<https://gricad-gitlab.univ-grenoble-alpes.fr/chatelaf/>
Data Science for Geosciences

Topics

1. **Prediction of ENSO (El Niño-Southern Oscillation) 6 months ahead**
2. Recognition of ocean-atmosphere events in satellite SAR images
3. Analysis and dimensionality reduction of Sea Surface Height
4. Automatic fish sound recognition
5. Land cover classification from hyperspectral data

Supervisor: P. Tandeo/R. Fablet
Data: multivariate time series data
Problem type: regression



Topics

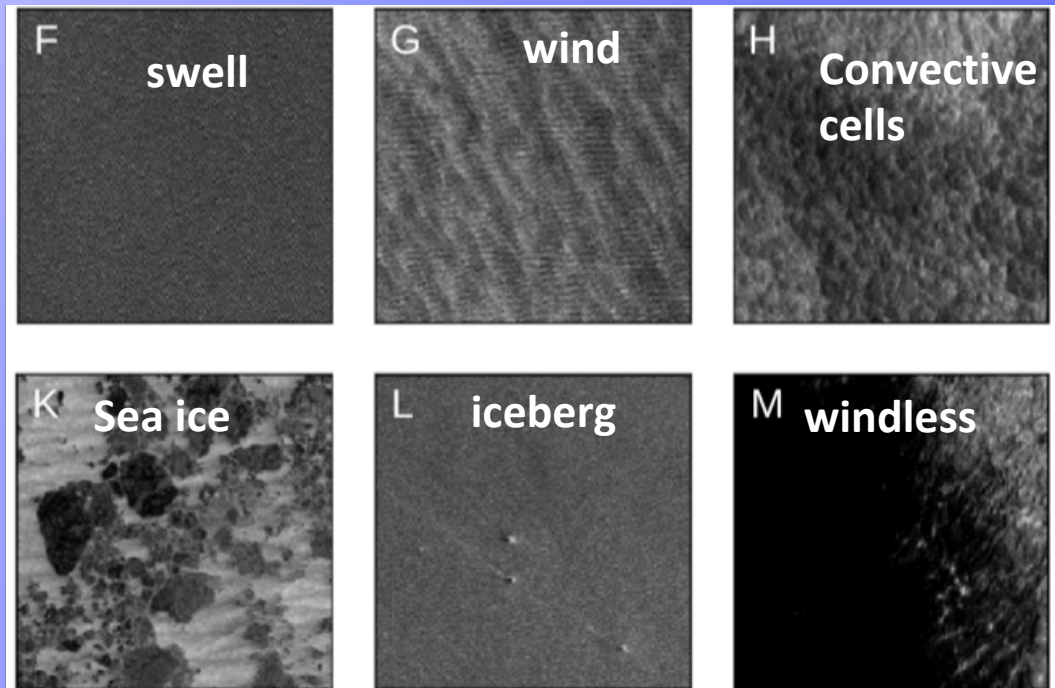
1. Prediction of ENSO (El Niño-Southern Oscillation)
2. **Recognition of ocean-atmosphere events in satellite SAR images**
3. Analysis and dimensionality reduction of Sea Surface Height
4. Automatic fish sound recognition
5. Land cover classification from hyperspectral data

Supervisor: P. Tandeo/R. Fablet

Data: SAR data

Problem type: classification

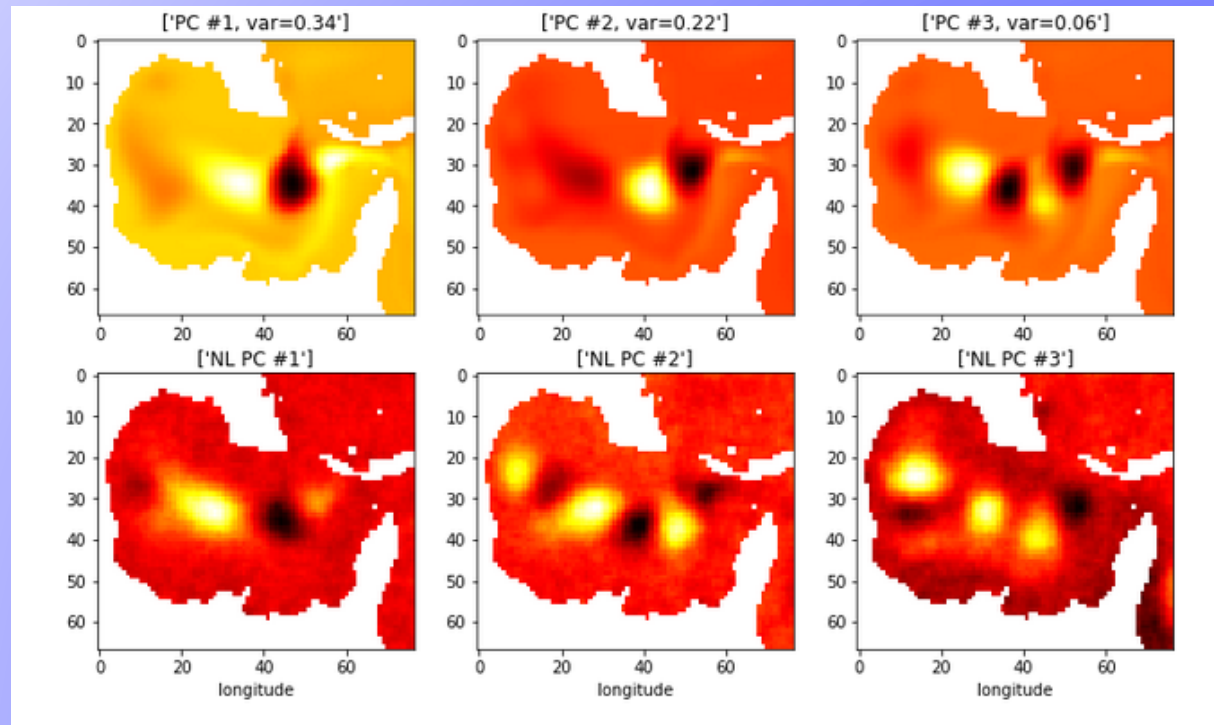
Examples of SAR sea surface observations
for different ocean/atmosphere events



Topics

1. Prediction of ENSO (El Niño-Southern Oscillation)
2. Recognition of ocean-atmosphere events in satellite SAR images
3. Analysis and dimensionality reduction of Sea Surface Height
4. Automatic fish sound recognition
5. Land cover classification from hyperspectral data

Supervisor: L. Drumetz
Data: Occiput data
Problem type: regression



Using Principal Component Analysis and Autoencoders

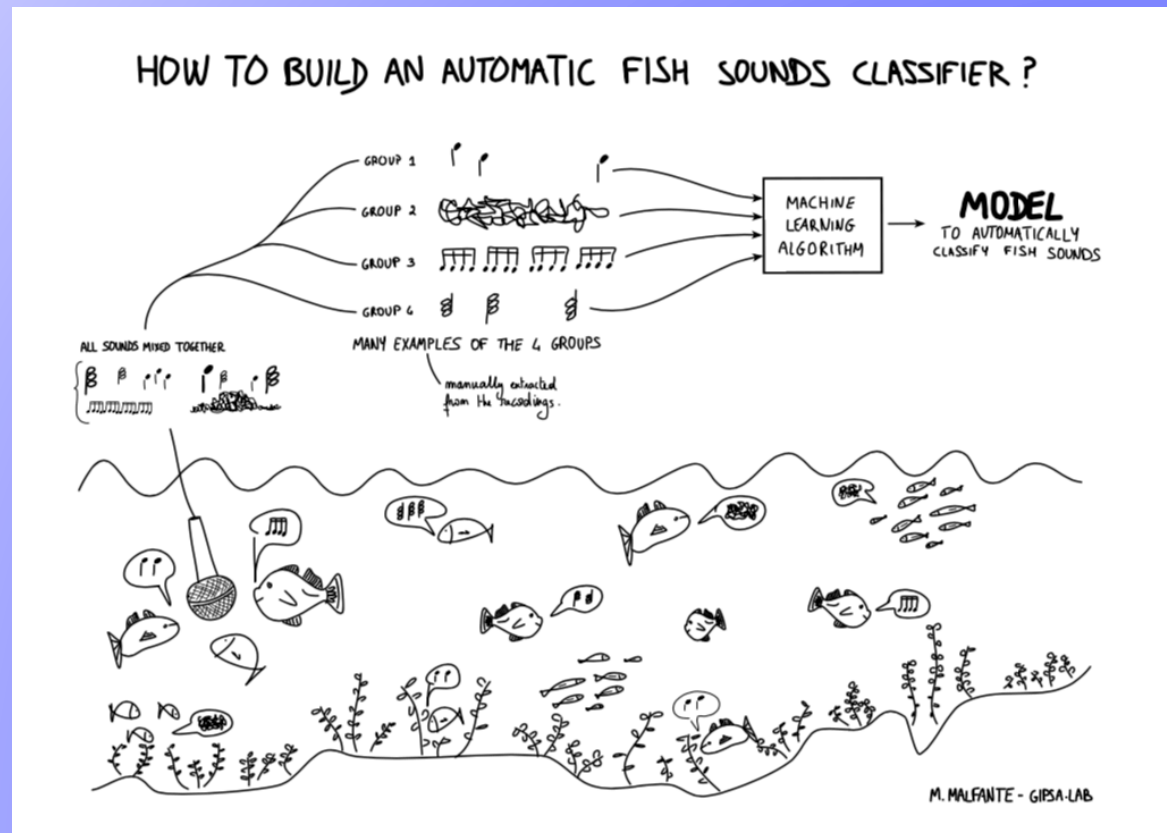
Topics

1. Prediction of ENSO (El Niño-Southern Oscillation)
2. Recognition of ocean-atmosphere events in satellite SAR images
3. Analysis and dimensionality reduction of Sea Surface Height
4. **Automatic fish sound recognition**
5. Land cover classification from hyperspectral data

Supervisor: M. Dalla Mura

Data: acoustic data

Problem type: classification



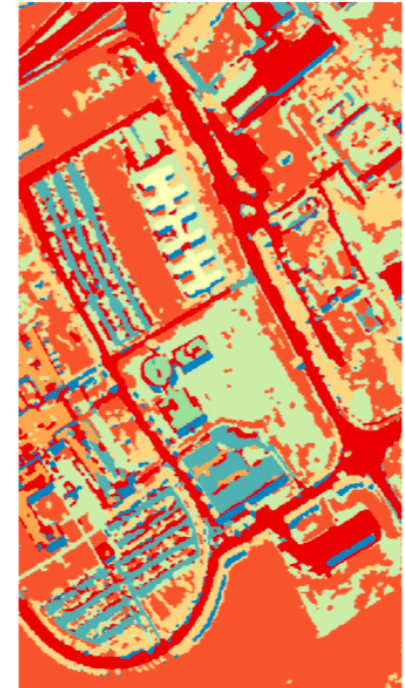
Topics

1. Prediction of ENSO (El Niño-Southern Oscillation)
2. Recognition of ocean-atmosphere events in satellite SAR images
3. Analysis and dimensionality reduction of Sea Surface Height
4. Automatic fish sound recognition
5. **Land cover classification from hyperspectral data**

Supervisor: M. Fauvel

Data: hyperspectral data

Problem type: classification



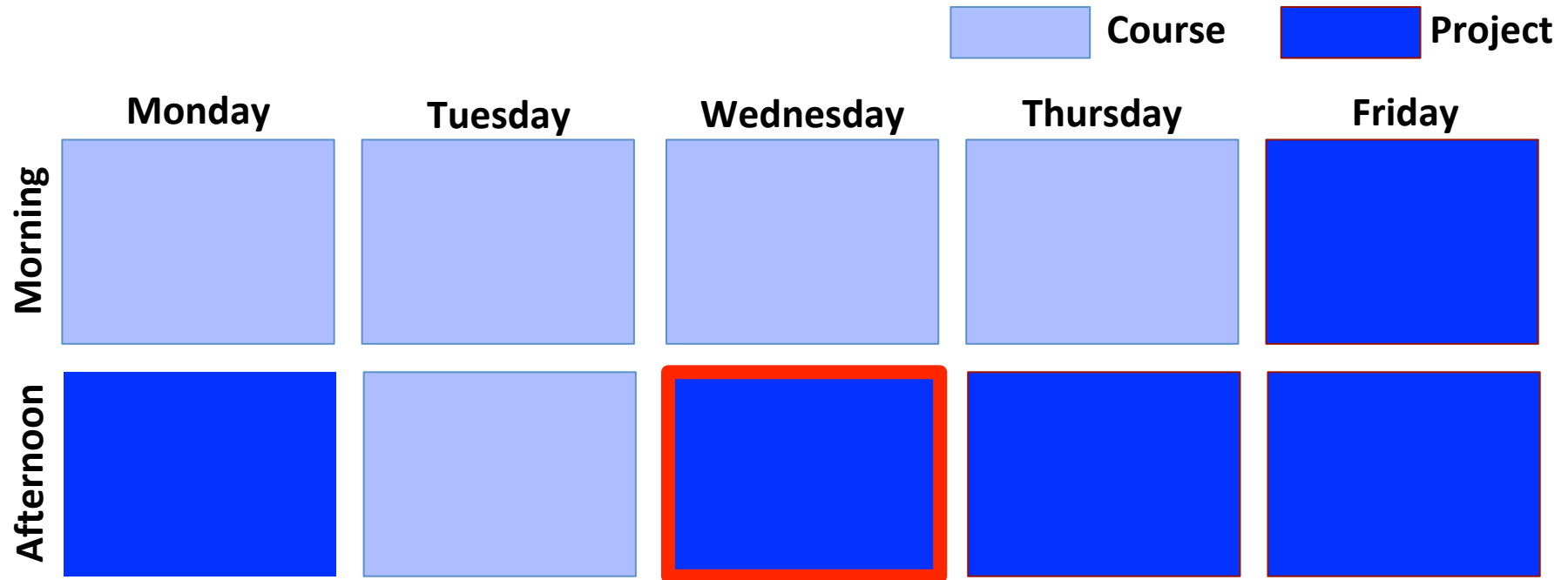
Specific objectives

		<div></div> Course		<div></div> Project		
		Monday	Tuesday	Wednesday	Thursday	Friday
Morning	Afternoon	<div></div>	<div></div>	<div></div>	<div></div>	<div></div>
		<div></div>	<div></div>	<div></div>	<div></div>	<div></div>

Session 1: Monday, 2pm-5pm

- Understanding of the targeted problem and of the associated data
- Exploratory analysis and visualisation of the dataset
- Statement of the targeted problem as a machine learning issue

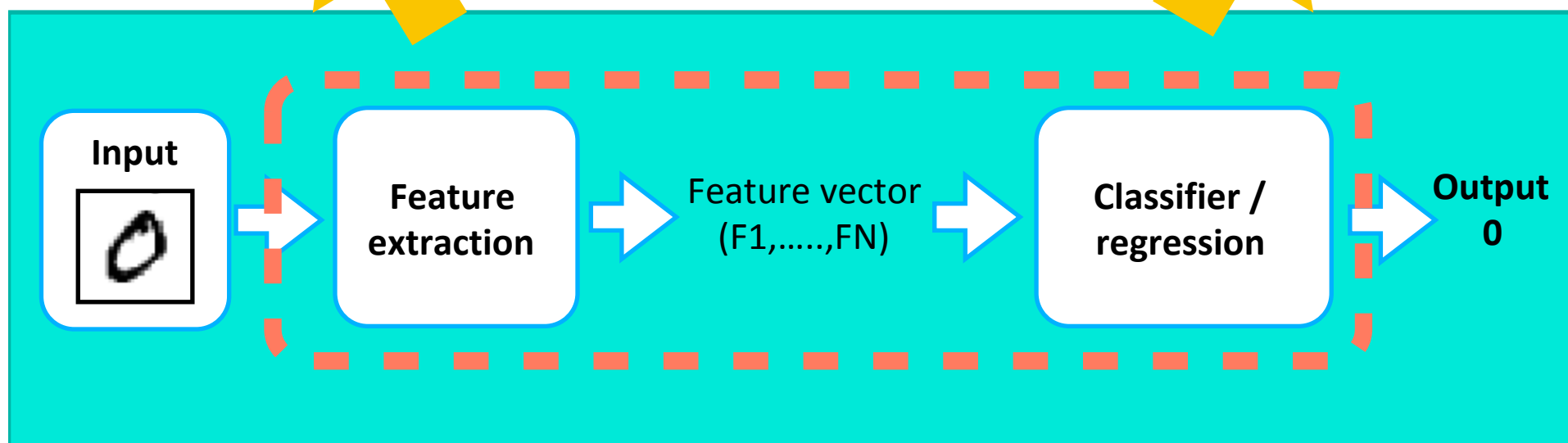
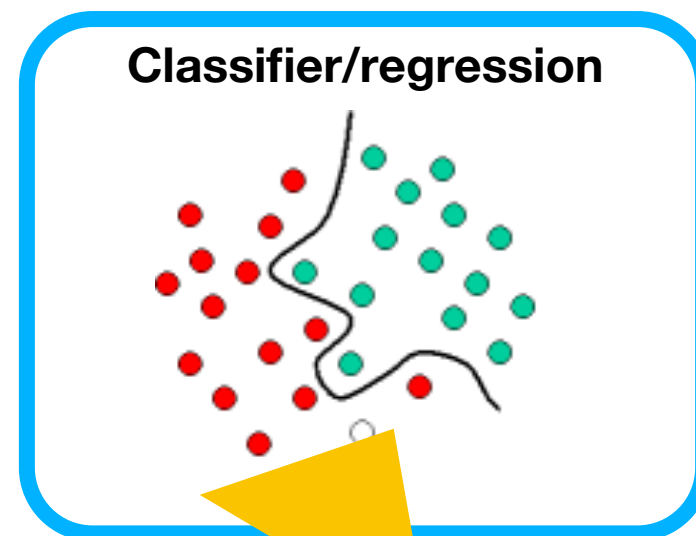
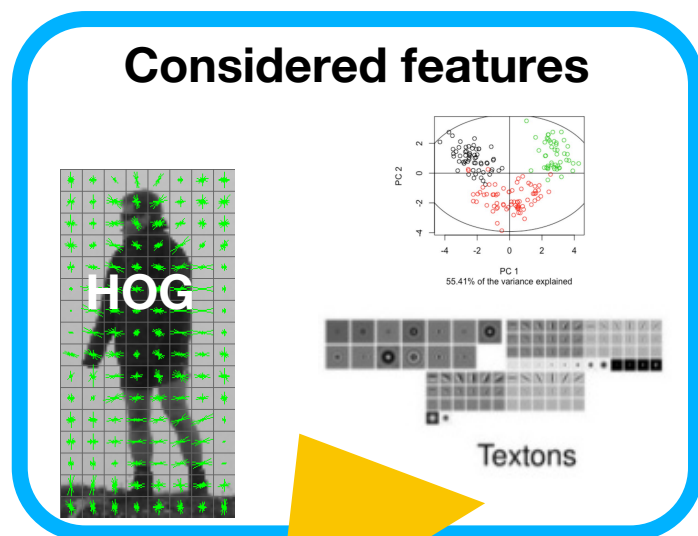
Specific objectives



Session 2: Wednesday, 2pm-5pm

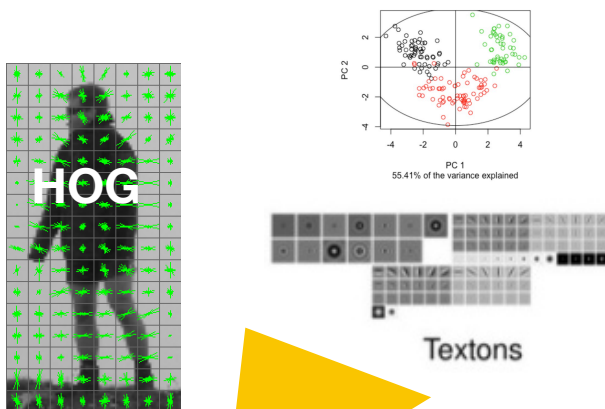
- Design of a machine learning framework for the targeted problem
- Selection of ML models and libraries
- Definition of evaluation criteria
- **5' pitch by each project member to other trainees**
- Implementation of a first solution

Session2: Designing your ML pipeline

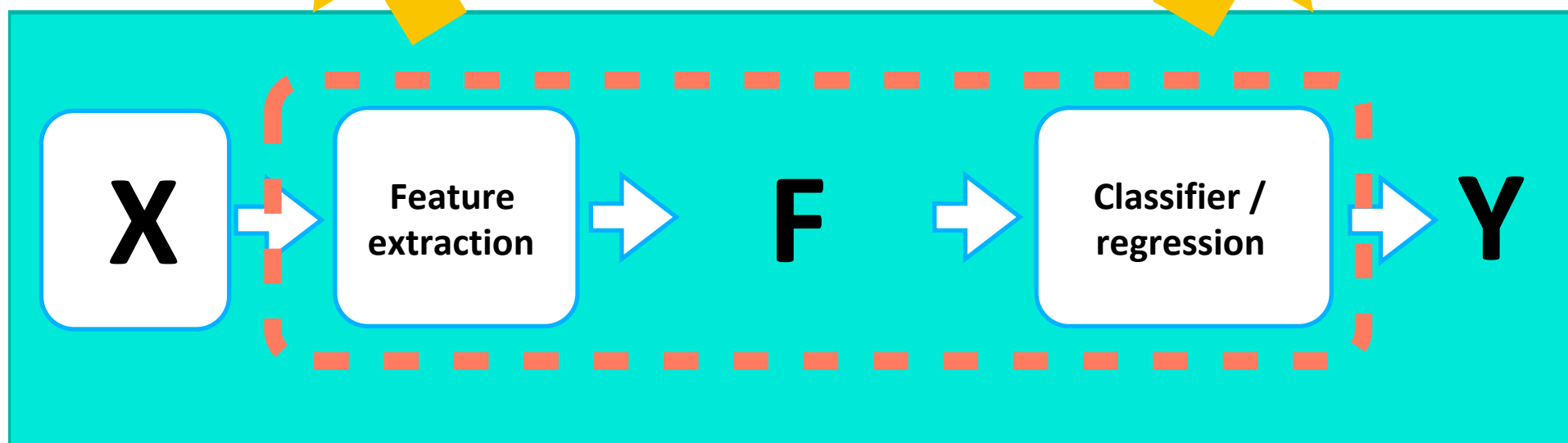
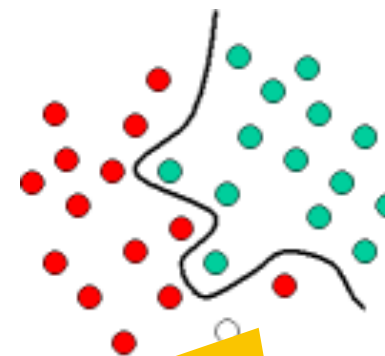


Session2: Designing your ML pipeline

Considered features



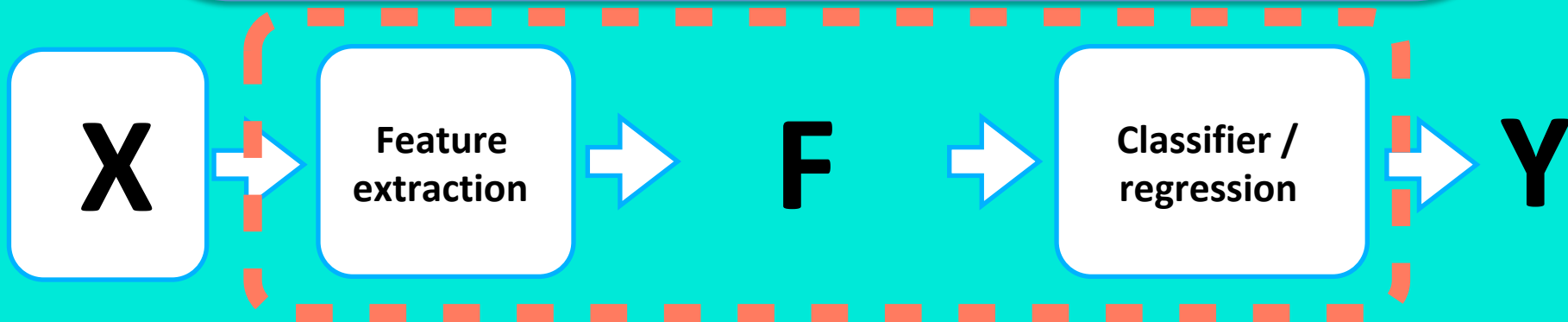
Classifier/regression



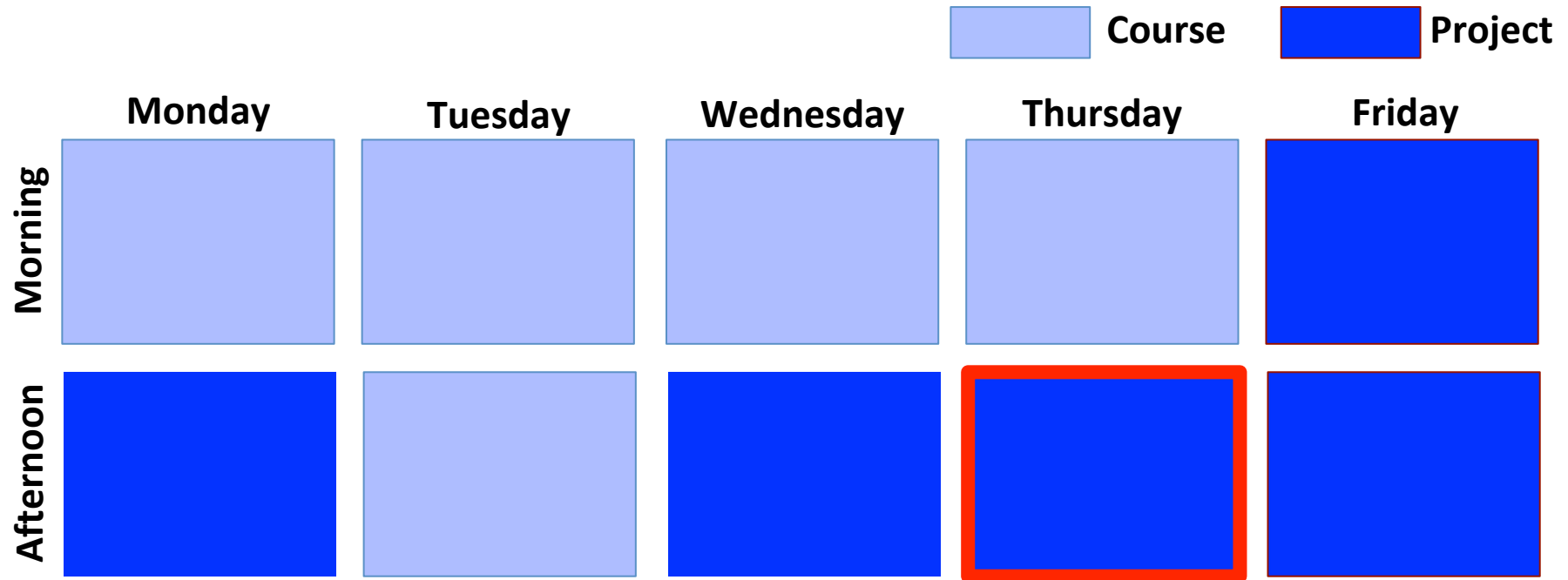
Session2: Designing your ML pipeline

Key steps to be presented in the 5' pitch

1. Define inputs and outputs X and Y
2. Build a groundtruthed dataset of pairs (X_i, Y_i)
3. Define the feature vector F
4. Select classifiers/regression models
5. Define a training/evaluation criterion



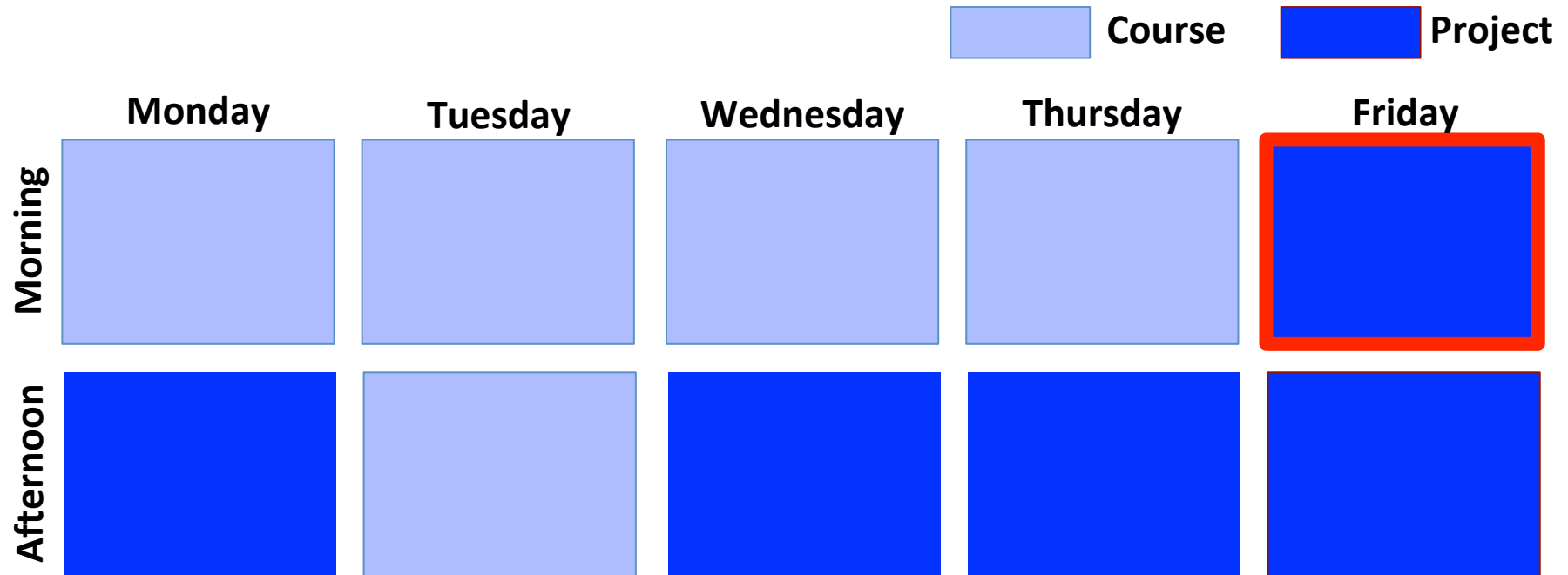
Specific objectives



Session 3: Thursday, 2pm-5pm

- Implementation of the proposed solutions
- In-depth evaluation of at least two different models (sensitivity analysis)
- Possible update/improvement of the proposed solutions and/or evaluation protocol













Specific objectives



Session 4: Friday, 9am-12am

- Project synthesis (Python notebook)
- Preparation of a joint presentation between the two groups for each project (initially, 1 member from each group then all members)

Specific objectives

		 Course		 Project		
		Monday	Tuesday	Wednesday	Thursday	Friday
Morning						
						
Afternoon						

Session 5: Friday, 1.30pm-4.30pm

- Presentation of each project (« poster » session)
- Synthesis: question/response session
- Feedbacks on the content and organization of the course

And now, it's up to you

Choose your group and topic