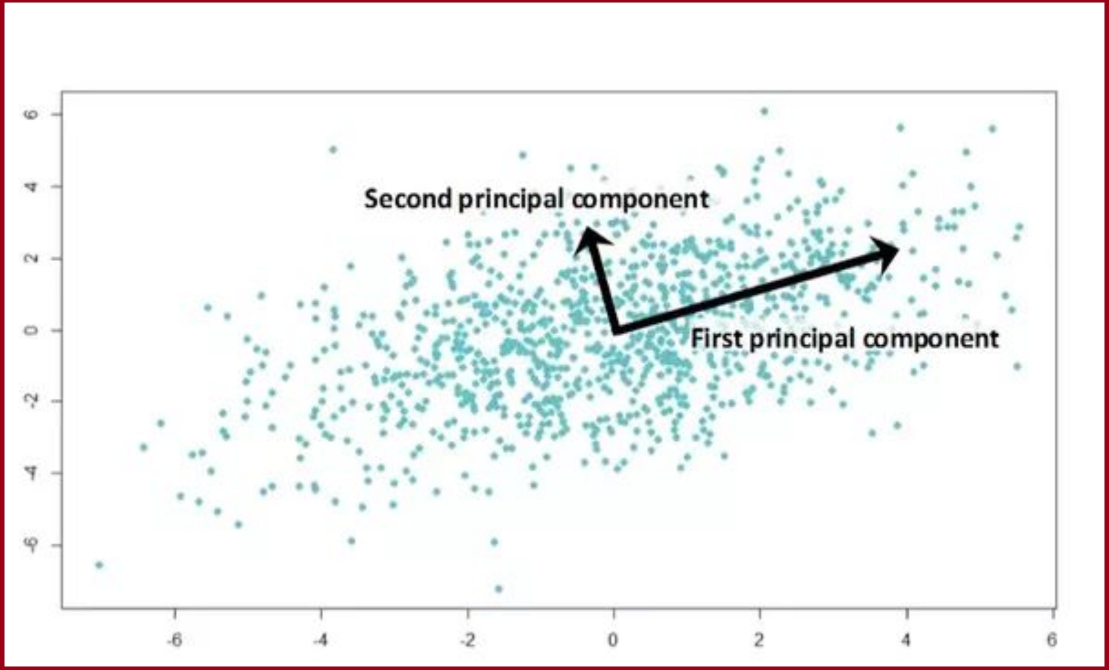
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

UNIVERSITÀ DEGLI STUDI DI PADOVA

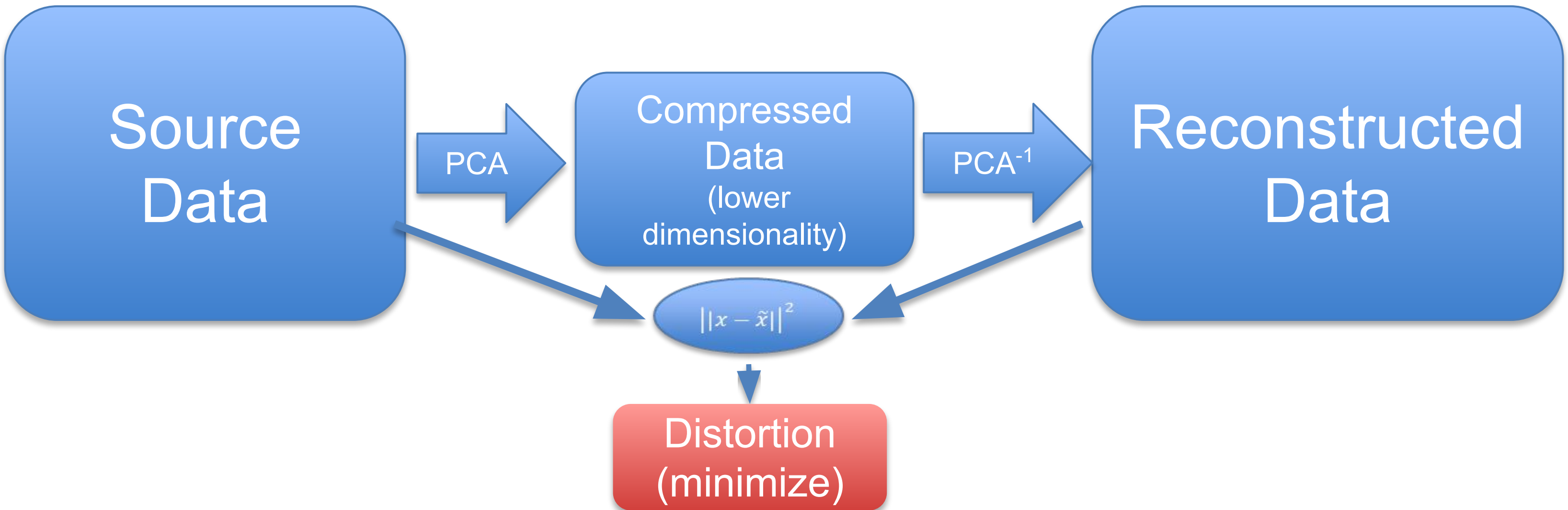


Principal Component Analysis

Machine Learning 2022-23
UML Book Chapter 23

Dimensionality Reduction

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE



- Take data from an highly dimensional space and project to a lower dimensional one
- Many applications: reduce number of features (learn with less samples, lower computation req.), capture most important aspects of the data for subsequent analysis, visualization, etc..
- Lower dimensional data should be a *good approximation* of the higher dimensional representations
- *Good approximation*: minimize error obtained by reprojecting the data back to the high dimensional space → similar to lossy data compression
- Focus on linear mapping of the data (represented by a matrix multiplication)
- *Principal Component Analysis (PCA)* : find the linear mapping that minimizes the mean squared error in the reprojection

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Principal Component Analysis (PCA)

- $\boldsymbol{x}_1, \boldsymbol{x}_2, .., \boldsymbol{x}_m \in \mathbb{R}^d$: data points

- $W \in \mathbb{R}^{n,d}$ $(n < d)$: mapping $\boldsymbol{x} \rightarrow \boldsymbol{y} = W\boldsymbol{x}$
  - where $\boldsymbol{y} = W\boldsymbol{x} \in \mathbb{R}^n$ is a lower dimensional representation of $\boldsymbol{x} \in \mathbb{R}^d$

- $U \in \mathbb{R}^{d,n}$ $(n < d)$: inverse mapping $\boldsymbol{y} \rightarrow \widetilde{\boldsymbol{x}} = U\boldsymbol{y}$
  - used to recover an approximation $\widetilde{\boldsymbol{x}} = \boldsymbol{UWx}$ of $\boldsymbol{x}$

- Target: find the lower dimensional representation that better approximates the data → that leads to minimum squared distance between $\widetilde{\boldsymbol{x}}$ and $\boldsymbol{x}$

$$\underset{W \in \mathbb{R}^{n,d}, U \in \mathbb{R}^{d,n}}{\operatorname{argmin}} \sum_{i=1}^{m} \|\boldsymbol{x}_i - \underbrace{UW\boldsymbol{x}_i}_{\widetilde{\boldsymbol{x}}_i}\|_2^2$$

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

PCA : Algorithm Idea

Target:
seek for the *n*-dimensional basis that best captures the variance in the d-dimensional data

Procedure:

1. First principal component (*p.c.*) = direction with largest projected variance
2. Second *p.c.* = orthogonal direction with largest projected variance
   - i.e., largest remaining variance after removing the first p.c.
3. ...(iterate for 3…n)….

- First derive the 1-dimensional subspace that maximizes the projected variance and then proceed iteratively

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Lemma

There exist an optimal solution (U^*, W^*) of $\underset{W \in \mathbb{R}^{n,d}, U \in \mathbb{R}^{d,n}}{\operatorname{argmin}} \sum_{i=1}^{m} \|\boldsymbol{x}_i - UW\boldsymbol{x}_i\|_2^2$ where:

- the columns of U^* are orthonormal (i.e., $(U^*)^T U^* = I$)
- $W^* = (U^*)^T$

Recall: orthonormal

- $\mathbf{u}_i^T \mathbf{u}_j = 0, \forall i \neq j$
- $\|\mathbf{u}_i\| = 1 = \mathbf{u}_i^T \mathbf{u}_i, \forall i$

Demonstration:

1. Fix U,W and consider the mapping $\mathbf{x} \to UW\boldsymbol{x}$
   - The range of the mapping is $R = \{UW\boldsymbol{x} \ ; \boldsymbol{x} \in \mathbb{R}^d\}$
2. $V \in \mathbb{R}^{d,n}$: matrix whose column form an orthonormal basis of R
   - Recall that $V^T V = I$ and $\forall \boldsymbol{x} \in R: V\boldsymbol{y} \ with \ y \in \mathbb{R}^n$
3. $\forall \boldsymbol{x} \in \mathbb{R}^d, \forall \boldsymbol{y} \in \mathbb{R}^n$:
   - $\|\boldsymbol{x} - V\boldsymbol{y}\|_2^2 = \|\boldsymbol{x}\|^2 + \boldsymbol{y}^T V^T V \boldsymbol{y} - 2\boldsymbol{y}^T V^T \boldsymbol{x} = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\boldsymbol{y}^T (V^T \boldsymbol{x})$
4. Minimize $\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\boldsymbol{y}^T (V^T \boldsymbol{x})$ w.r.t y: set $\nabla = 0 \to 2\boldsymbol{y} - 2(V^T \boldsymbol{x}) = 0 \to \boldsymbol{y}_{opt} = V^T \boldsymbol{x}$
5. $\forall \boldsymbol{x}$: $\underset{\tilde{\boldsymbol{x}} \in R}{\operatorname{argmin}} \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2^2 = V\boldsymbol{y}_{opt} = V(V^T \boldsymbol{x})$: it is the best approximation in subspace R
6. $\forall \boldsymbol{x}$: includes also $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ (data vectors): $\sum_{i=1}^{m} \|\boldsymbol{x}_i - UW\boldsymbol{x}_i\|^2 \geq \sum_{i=1}^{m} \|\boldsymbol{x}_i - VV^T \boldsymbol{x}_i\|^2$, so we can replace U,W with VV^T without increasing the objective
7. Holds for $\forall U, W$: there exist a solution that minimize $\sum_{i=1}^{m} \|\boldsymbol{x}_i - UW\boldsymbol{x}_i\|^2$ with V orthonormal columns and $W = \mathrm{U}^T$

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Optimization Problem

There exist an optimal solution (U^*,W^*) of $\underset{W\in\mathbb{R}^{n,d},U\in\mathbb{R}^{d,n}}{\operatorname{argmin}} \sum_{i=1}^{m}\|\boldsymbol{x}_i - UW\boldsymbol{x}_i\|_2^2$ where:

- the columns of U^* are orthonormal (i.e., $(U^*)^T U^* = I$)
- $W^* = (U^*)^T$

The optimization problem can be rewritten as:

$$\underset{U\in\mathbb{R}^{d,n}:U^TU=I}{\operatorname{argmin}} \sum_{i=1}^{m} \|\boldsymbol{x}_i - UU^T\boldsymbol{x}_i\|_2^2$$

- Trace: Σ elements on diagonal
- It is a scalar
- $trace(A^TB) = trace(AB^T) = = trace(B^TA) = trace(BA^T)$

With some manipulations: $\|\boldsymbol{x} - UU^T\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|^2 - 2x^TUU^Tx + x^TUU^TUU^Tx =$

$$= \|\boldsymbol{x}\|^2 - \boldsymbol{x}^TUU^T\boldsymbol{x} = \|\boldsymbol{x}\|^2 - trace(\boldsymbol{x}UU^T\boldsymbol{x}) = \|\boldsymbol{x}\|^2 - trace\left(U^T\boldsymbol{x}\boldsymbol{x}^TU\right)$$

$$\underset{U\in\mathbb{R}^{d,n}:U^TU=I}{\operatorname{argmax}} \; trace\left(U^T \sum_{i=1}^{m} \boldsymbol{x}_i\boldsymbol{x}_i^T \, U\right) = \underset{U\in\mathbb{R}^{d,n}:U^TU=I}{\operatorname{argmax}} \; trace(U^TAU)$$

Notice: $A = \sum_{i=1}^{m} \boldsymbol{x}_i\boldsymbol{x}_i^T$ is symmetric and positive semidefinite. It can be rewritten as $A = VDV^T$ where D is diagonal (with eigenvalues $D_{d,d} \geq 0$) and $V^TV = VV^T = I$ (the columns of V are the eigenvectors of A)

Theorem (PCA)

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, .., \boldsymbol{x}_m$ be arbitrary vectors in $\mathbb{R}^d$
let $A = \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^T$
let $\boldsymbol{u}_1, .., \boldsymbol{u}_n$ be *n* eigenvectors of A corresponding to the largest *n* eigenvalues of A

Then a solution of the PCA optimization $\underset{W \in \mathbb{R}^{n,d}, U \in \mathbb{R}^{d,n}}{\operatorname{argmin}} \sum_{i=1}^{m} \|\boldsymbol{x}_i - UW\boldsymbol{x}_i\|_2^2 = \underset{U \in \mathbb{R}^{d,n} : U^T U = I}{\operatorname{argmax}} \; trace(U^T A U)$
is to set U to be the matrix whose columns are $\boldsymbol{u}_1, .., \boldsymbol{u}_n$ and to set $W = U^T$

Notes:

- Recall: Decompose A as VDV^T (SVD decomposition, D diag. and $V^T V = VV^T = I$)
- It is a common practice to *"center"* the examples before applying PCA (i.e., subtract the mean)
- Computation time is $O(d^3) + O(md^2)$ (the first term for calculating eigenvalues and the second for constructing A)
- Trick for faster solution in case *d>>m* *(not part of the course)*

Pseudocode

grayed part: trick for *d>>m*
(not part of the course)

PCA

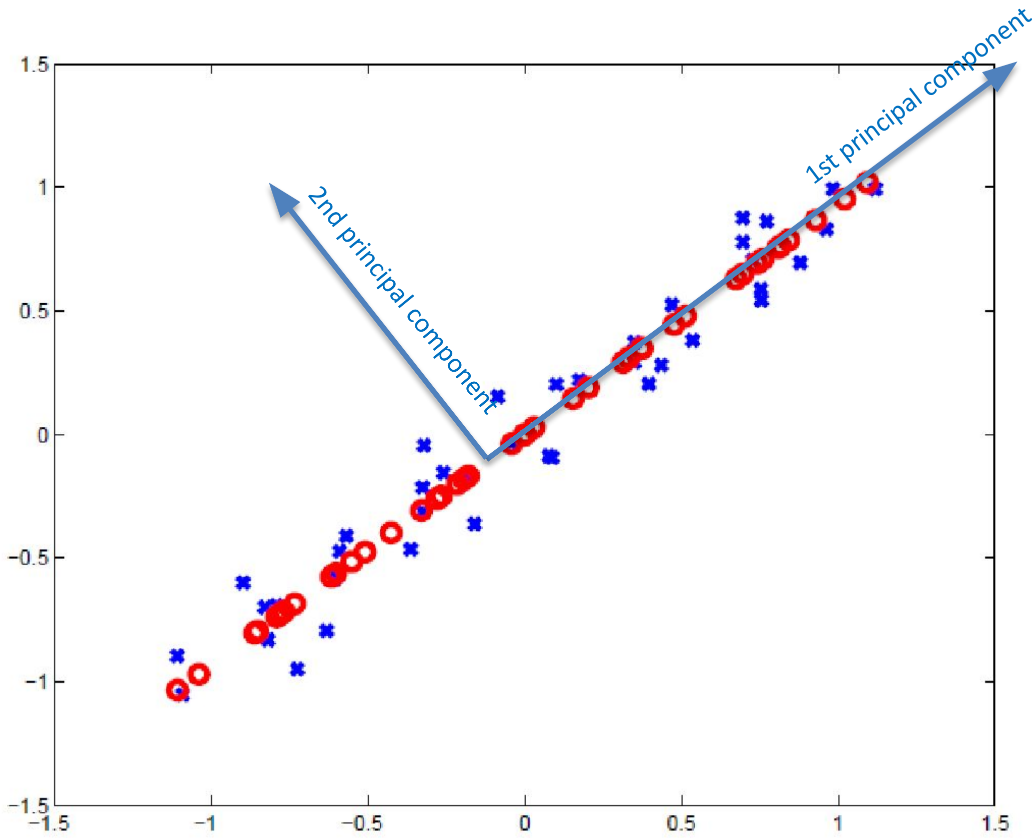
input
A matrix of m examples $X \in \mathbb{R}^{m,d}$
number of components n

if $(m > d)$
$A = X^\top X$
Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be the eigenvectors of A with largest eigenvalues

else
$B = XX^\top$
Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the eigenvectors of B with largest eigenvalues
for $i = 1, \ldots, n$ set $\mathbf{u}_i = \frac{1}{\|X^\top \mathbf{v}_i\|} X^\top \mathbf{v}_i$

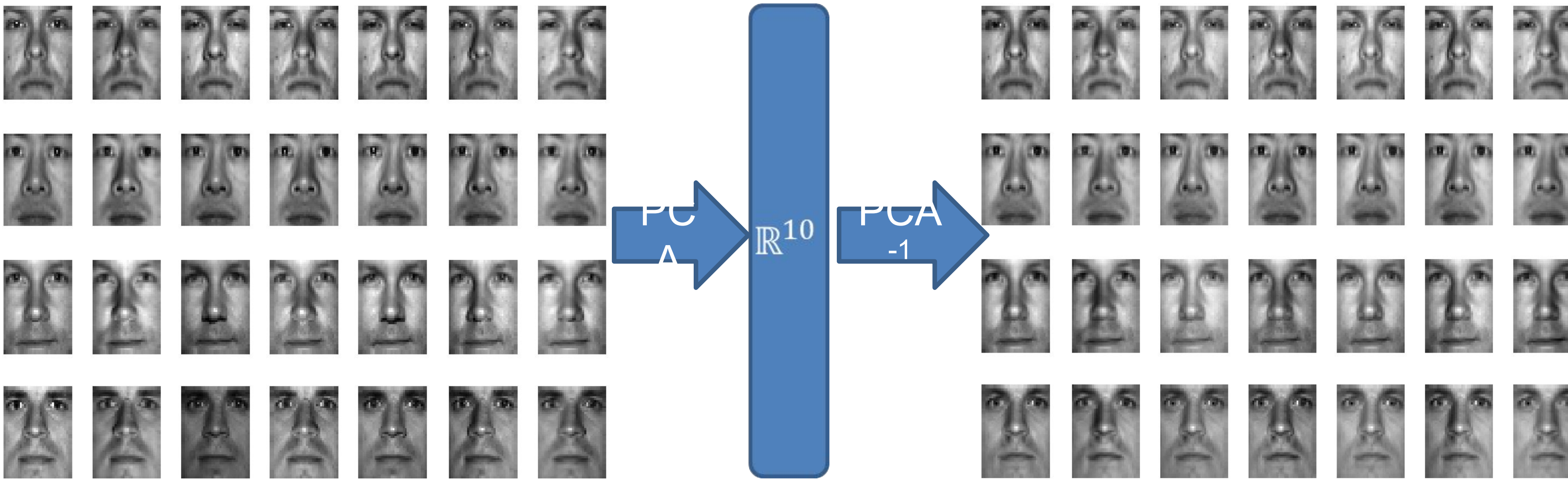
output: $\mathbf{u}_1, \ldots, \mathbf{u}_n$

Example: From 2D to 1D



Set of 2D vectors (blue) and their reconstruction (red) after dimensionality reduction to 1D with PCA

Example: Face Compression



$\mathbb{R}^{50x50} = \mathbb{R}^{2500}$

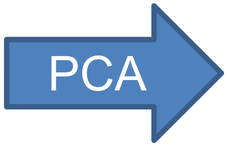
$\mathbb{R}^{50x50} = \mathbb{R}^{2500}$

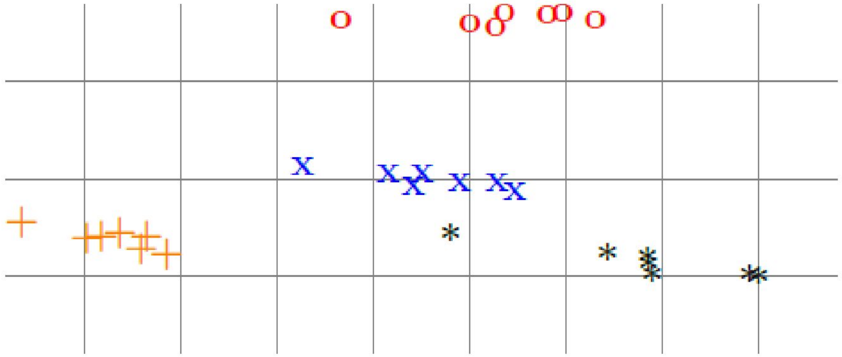
enlarged example

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Example: Face Recognition

$\mathbb{R}^{50x50} = \mathbb{R}^{2500}$





$\mathbb{R}^2$

- Faces with the same type of mark belong to the same individual
- PCA can be used for face recognition ! (*eigenfaces* algorithm)

Logistic info: January

- Wednesday, January 11: exercises + sample exam solution
- Friday, January 13: Lab 4 (Keras tutorial) - more information coming soon on Elearning
- Friday, January 20: exercises + Q&A

You can always ask for help or a meeting, but response times might be long during the holidays

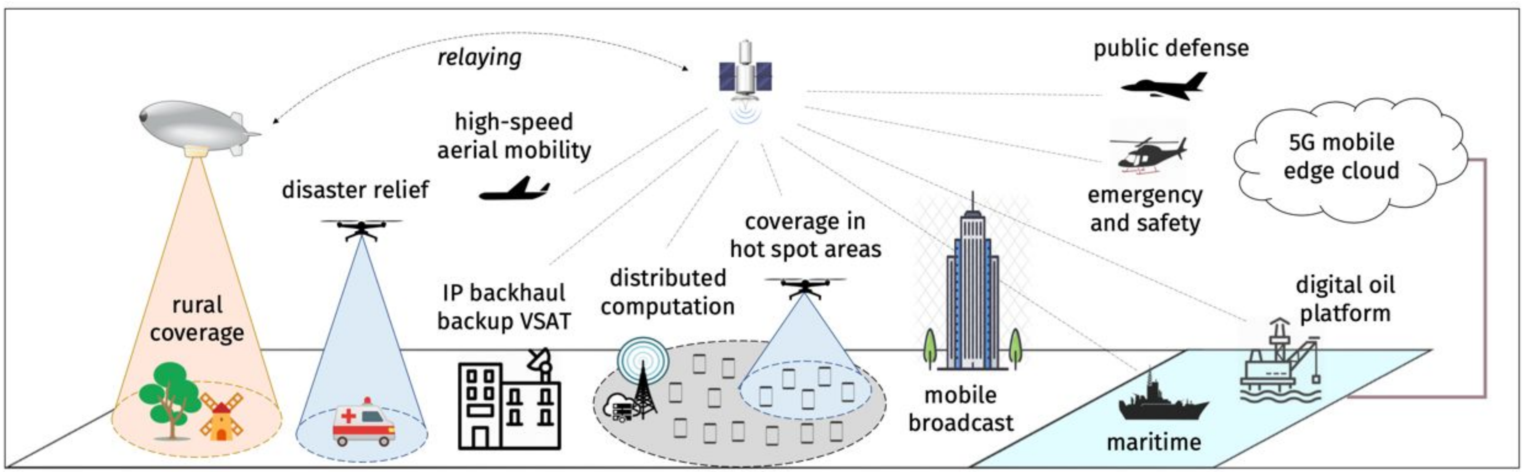
SIGNET thesis projects

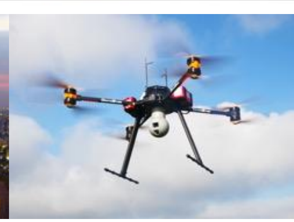
- Could be useful for thesis or course projects
  - The scope of the project can be tuned
- You can use ML in a real research context
- You can also propose your own ideas!

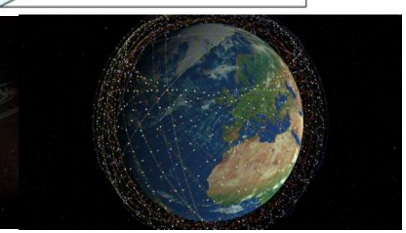
DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Non-Terrestrial Networks

- Complementing terrestrial infrastructures with aerial nodes (drones, satellites, high altitude platforms, etc.)
- ML optimization: how do we distribute comm/computation tasks?



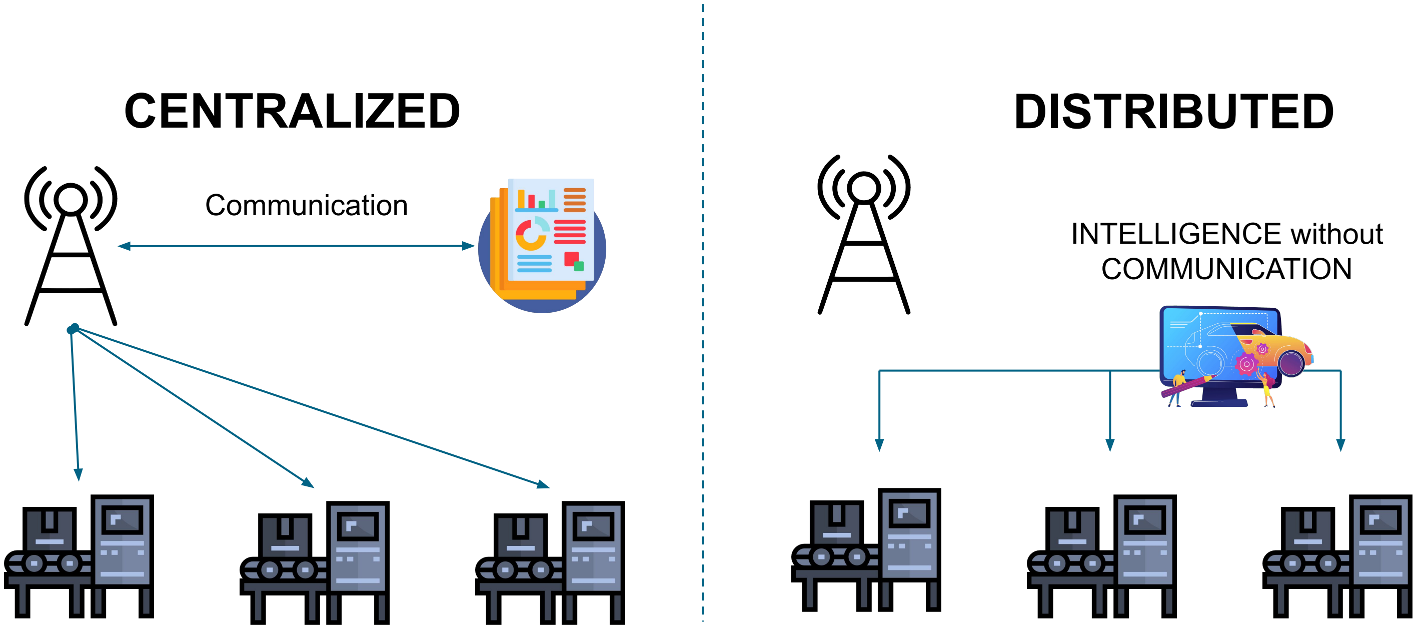



Industrial IoT

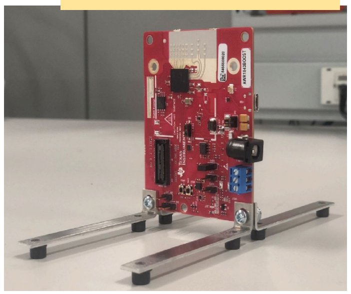
URLLC: extremely low latency services

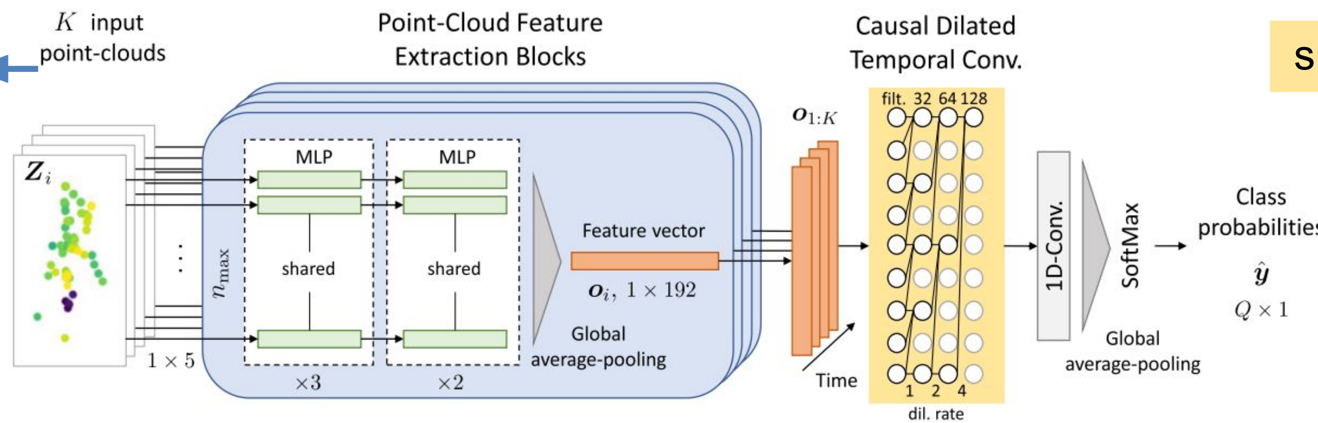
How do we distribute intelligence to meet the deadlines?



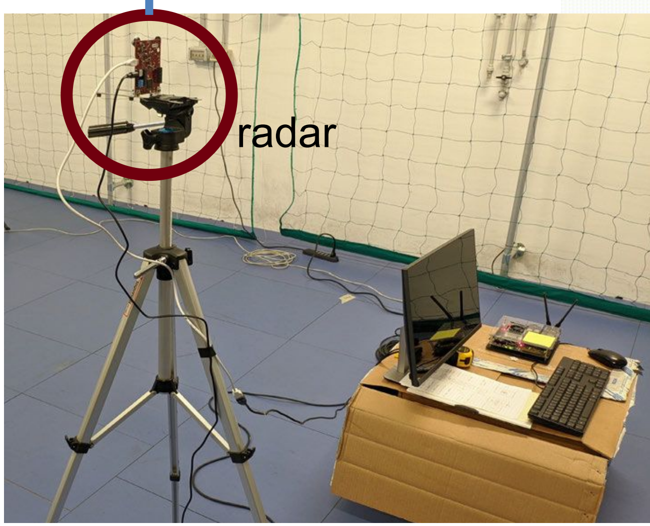
Radar identification

radar raw data



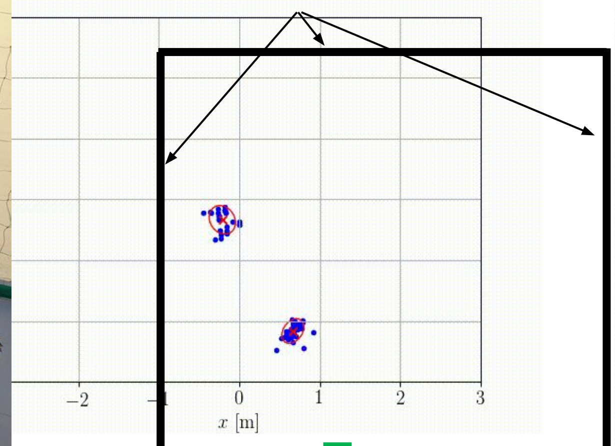


subject ID



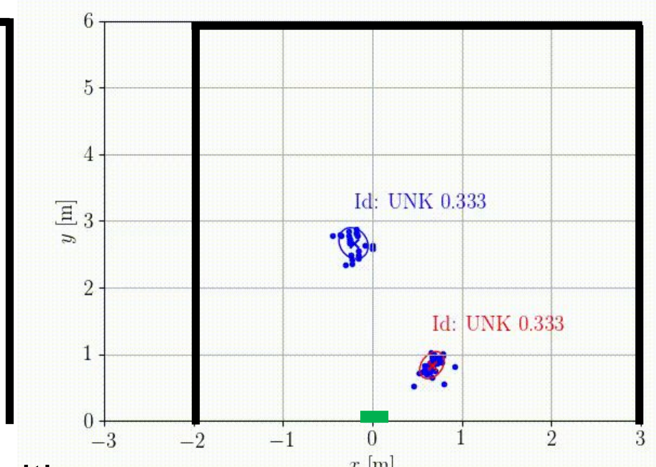
radar

Room walls



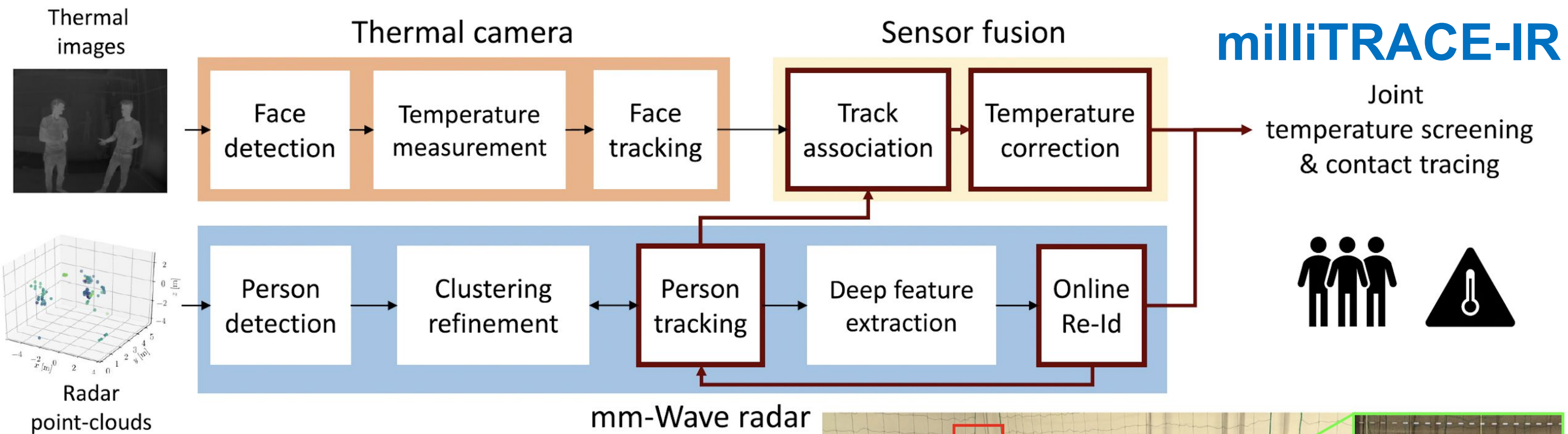
Radar position

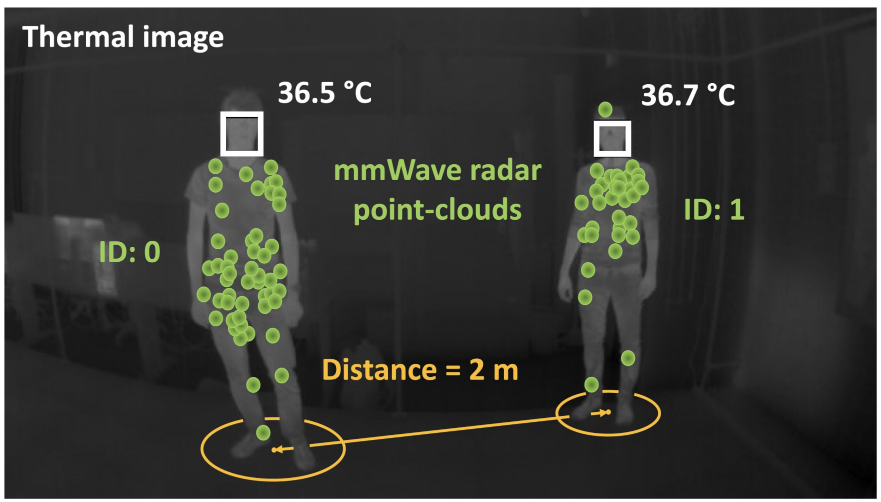
Identification

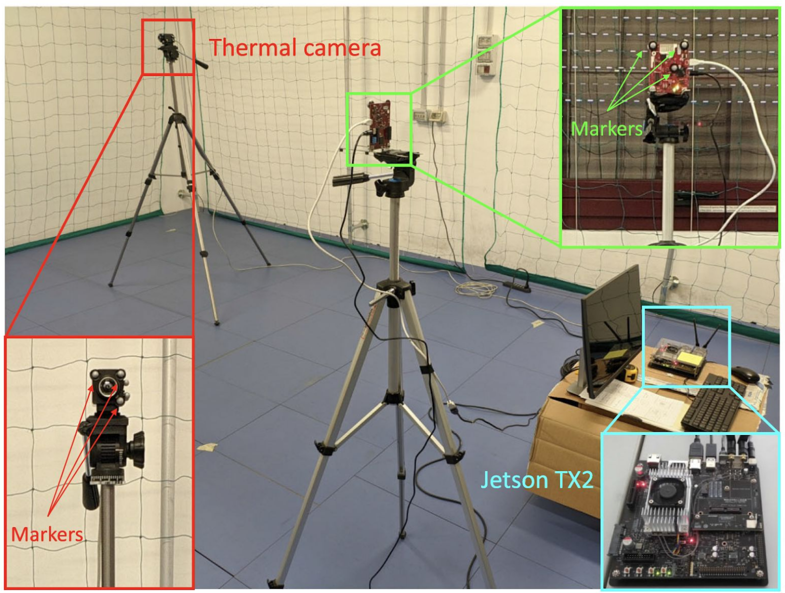


DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Temperature and contact tracing



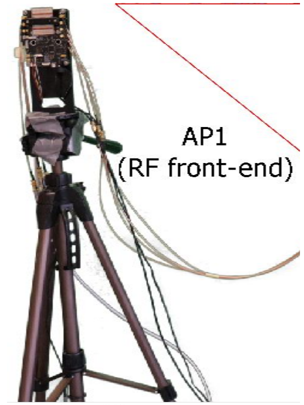


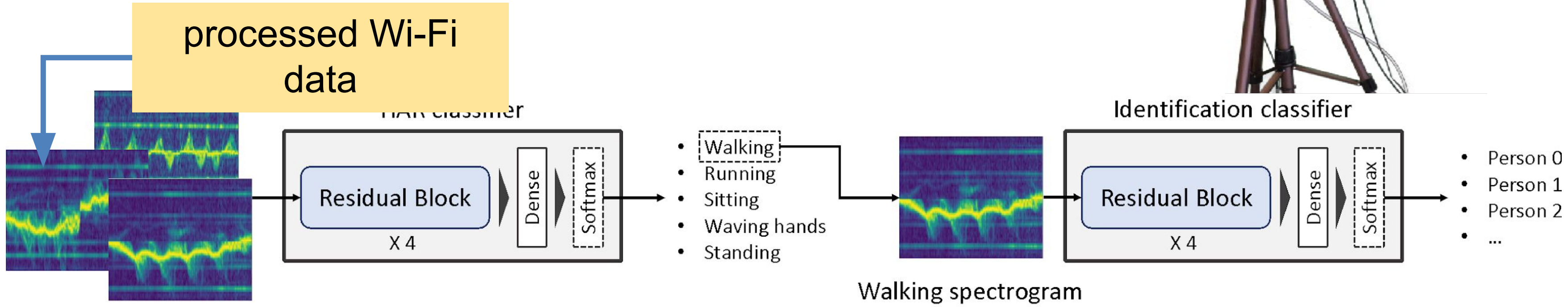


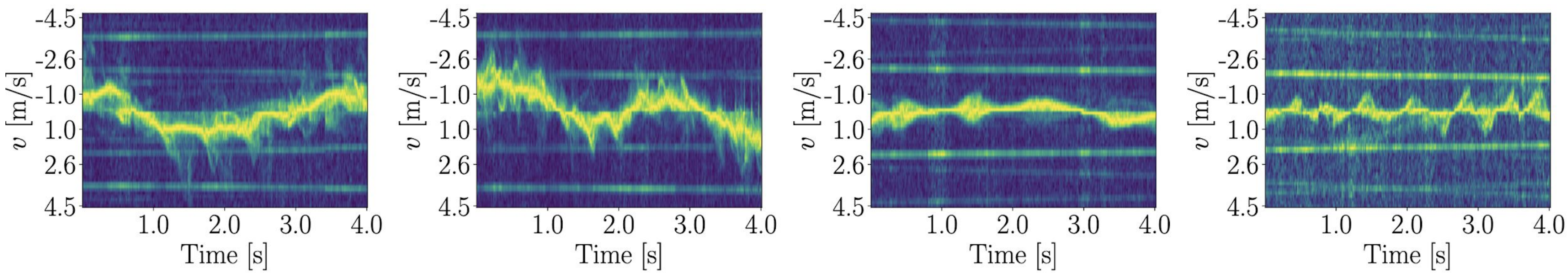
DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Wi-Fi sensing

- **RAPID** – indoor human detection and sensing through a testbed implementing the new IEEE 802.11ax Wi-Fi standard at 60 GHz



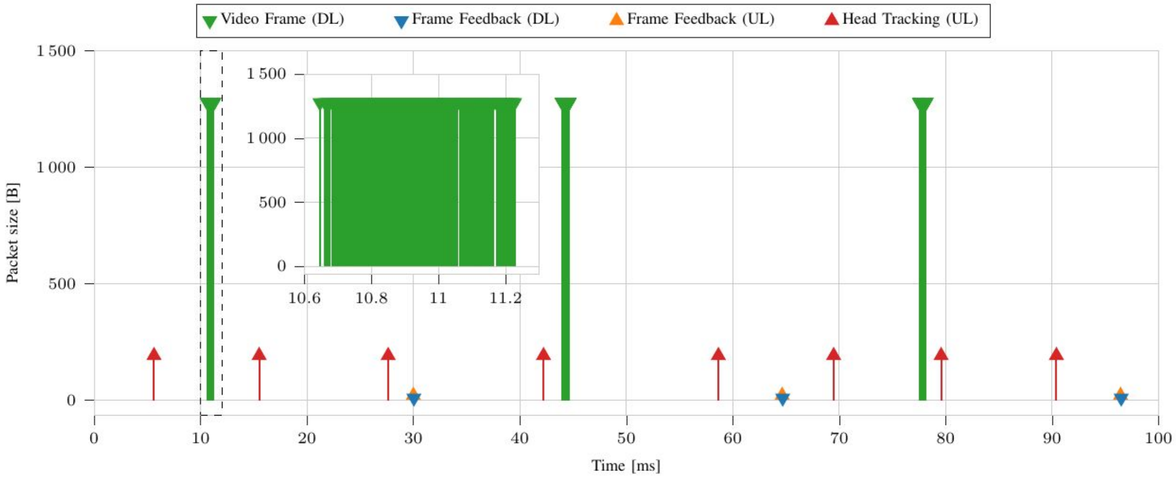




Walking. Running. Sitting down. Waving hands.

VR trace analysis

- How do we predict VR traffic?
- Frames depend on activity: what is the user doing?
- ML applied on traffic traces: capture and analysis tools



Semantic communications

- Adapting communications to only send the most relevant information
- Mix of ML styles: reinforcement, supervised, unsupervised
- Theory of mind: how do we model other agents?