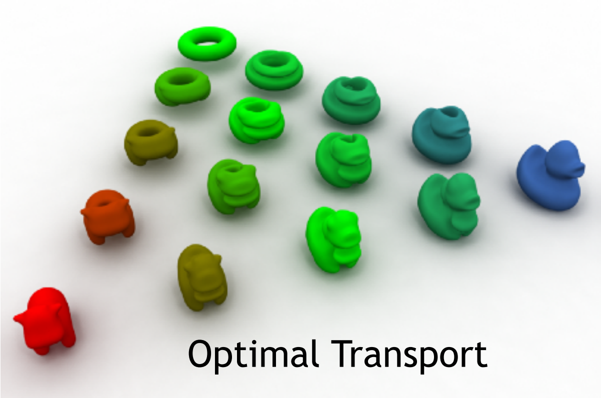
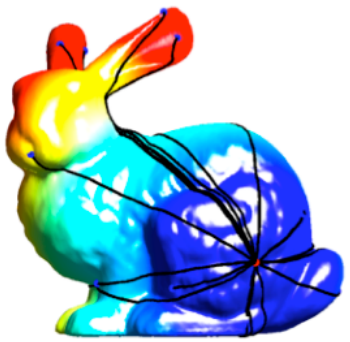
# Mathematical Coffees

Huawei-FSMP joint seminars

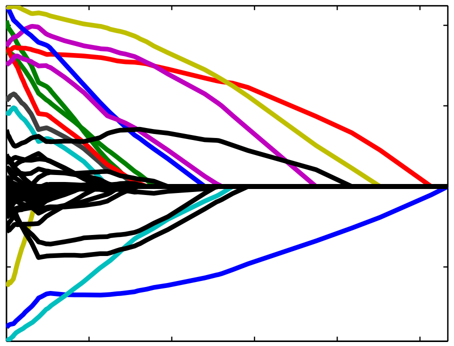
https://mathematical-coffees.github.io

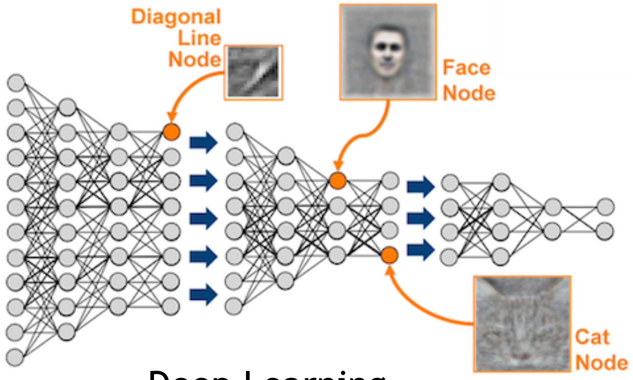
**Organized by**: Mérouane Debbah & Gabriel Peyré


Optimal Transport

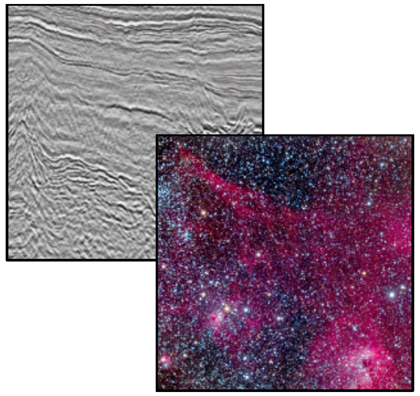

Geodesics

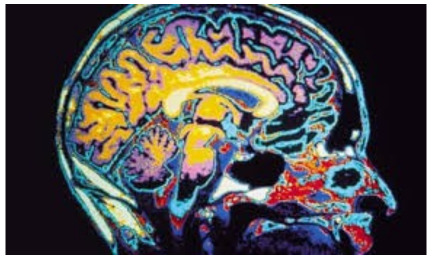

Meshes

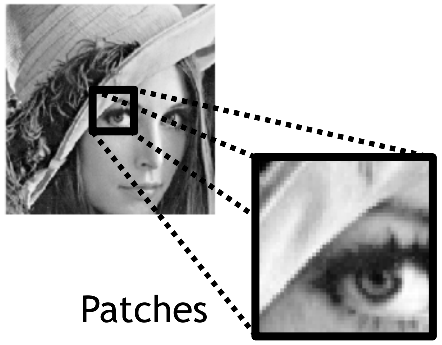

Optimization

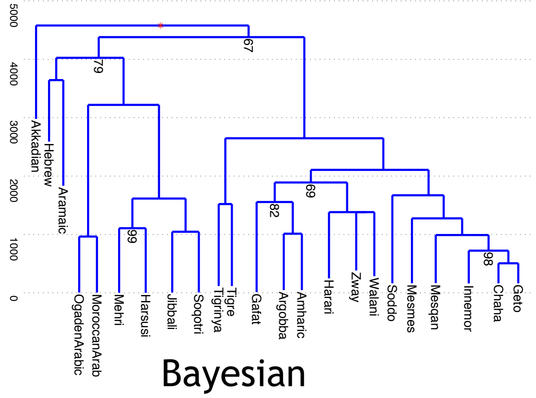


Deep Learning

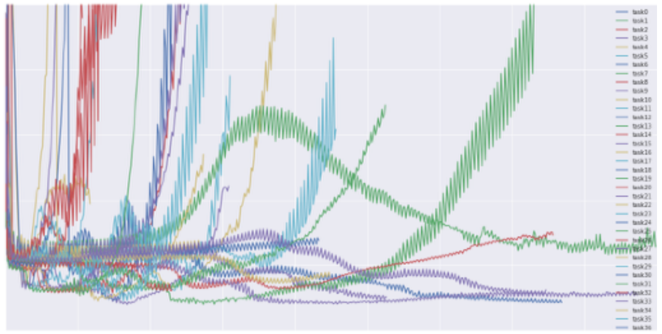

Sparsity


Neuro-imaging


Patches



Bayesian

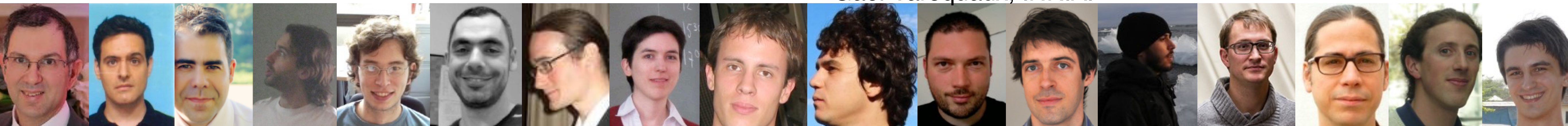

Parallel/Stochastic

Alexandre Allauzen, Paris-Sud.
Pierre Alliez, INRIA.
Guillaume Charpiat, INRIA.
Emilie Chouzenoux, Paris-Est.

Nicolas Courty, IRISA.
Laurent Cohen, CNRS Dauphine.
Marco Cuturi, ENSAE.
Julie Delon, Paris 5.

Fabian Pedregosa, INRIA.
Guillaume Lecué, CNRS ENSAE
Julien Tierny, CNRS and P6.
Robin Ryder, Paris-Dauphine.
Gael Varoquaux, INRIA.

Jalal Fadili, ENSICaen.
Alexandre Gramfort, INRIA.
Matthieu Kowalski, Supelec.
Jean-Marie Mirebeau, CNRS,P-Sud.



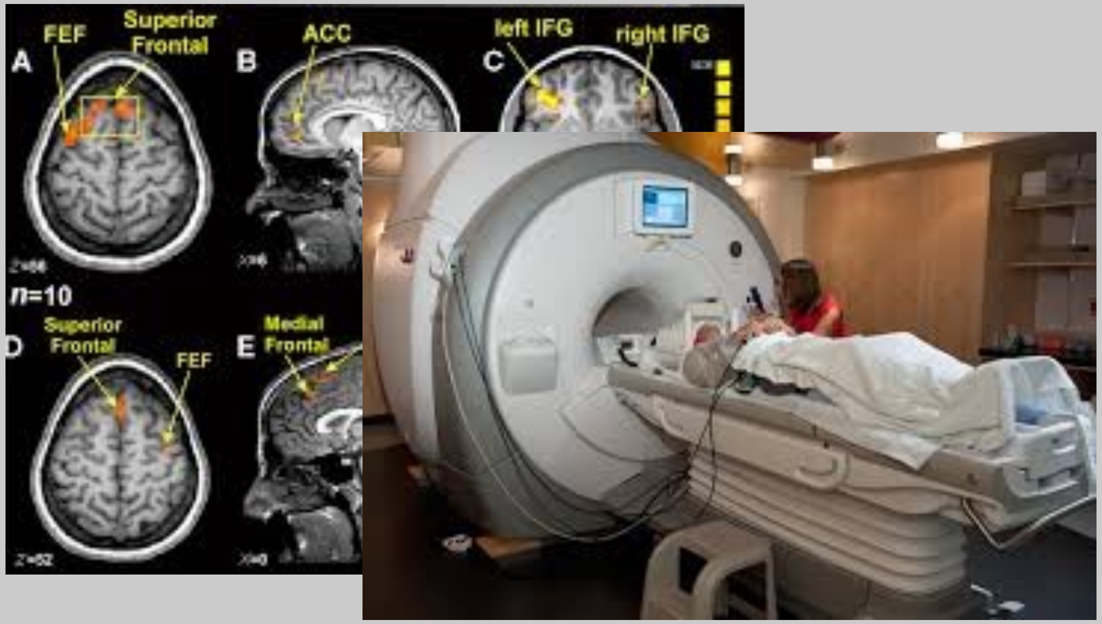
# Curses and Blessings of High Dimension

**Gabriel Peyré**

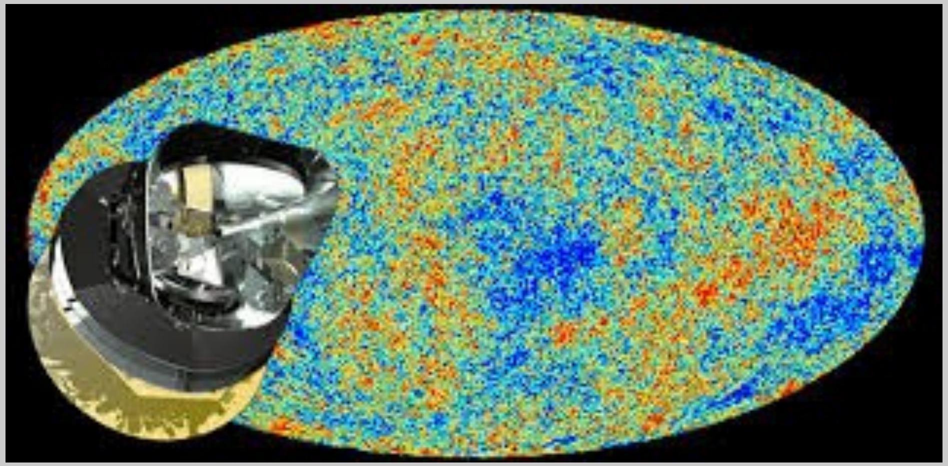
**www.numerical-tours.com**

# Big Datasets (large *n*)

Giga pixels camera



fMRI imaging



Planck mission

Imaging sciences

Machine learning

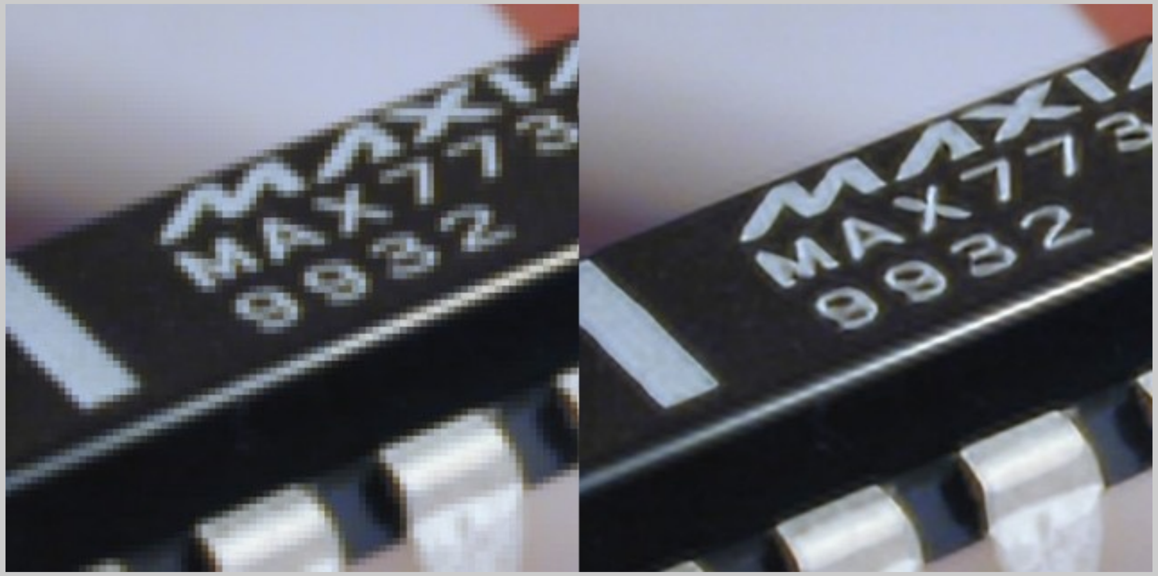
DNA microarrays

Network monitoring

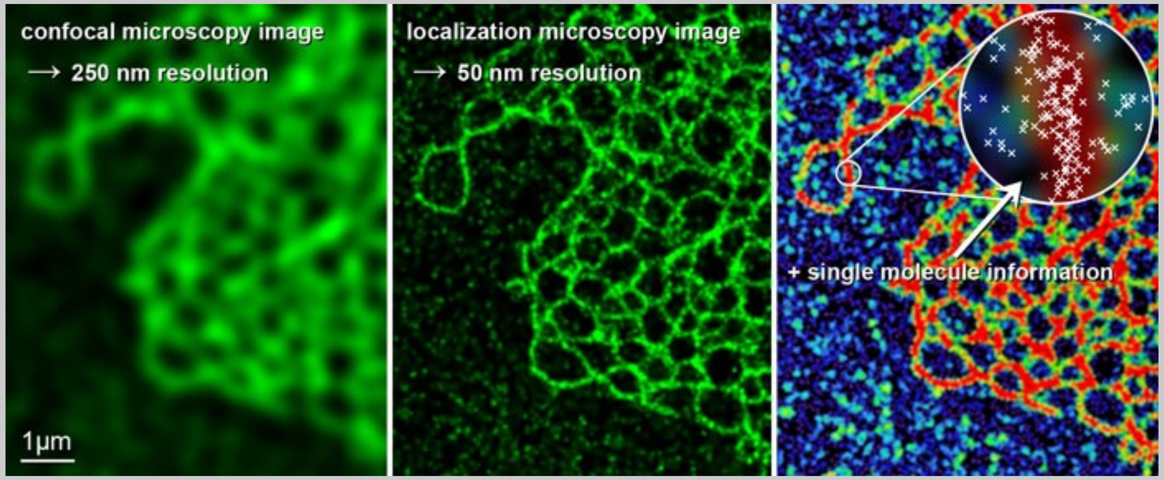
# Big Models (large $p$)

Imaging sciences

Super-resolution



Single molecule imaging



Machine learning

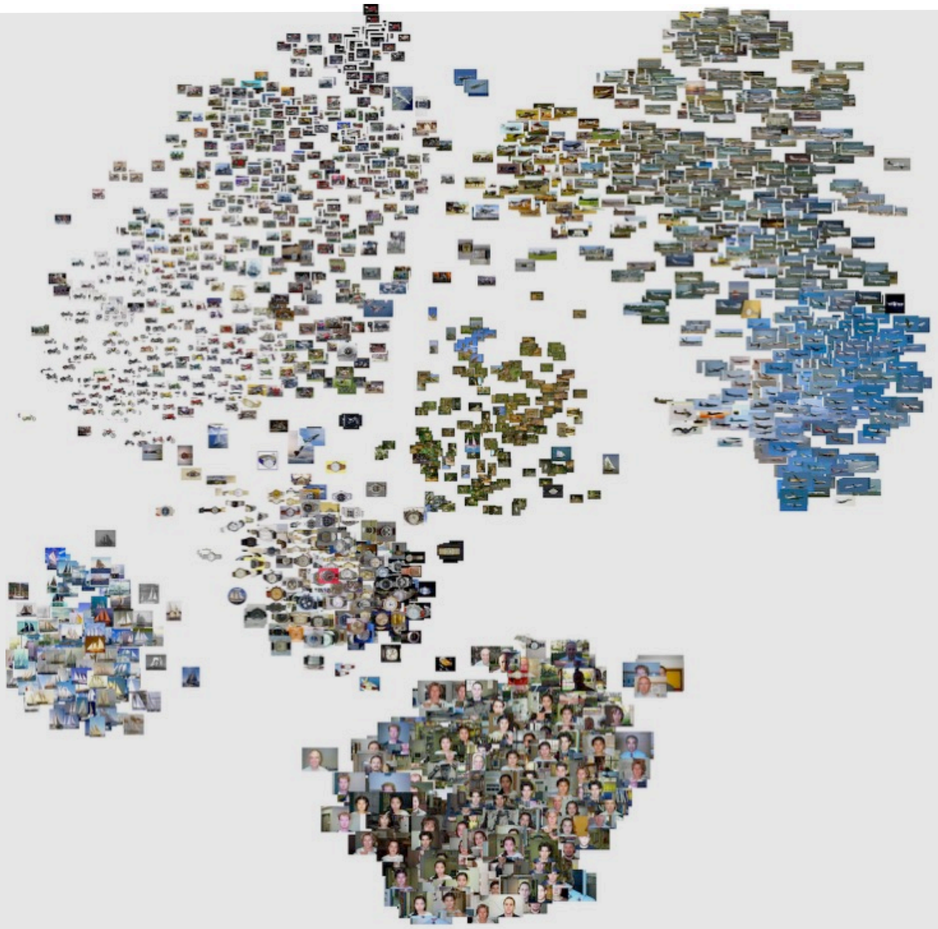
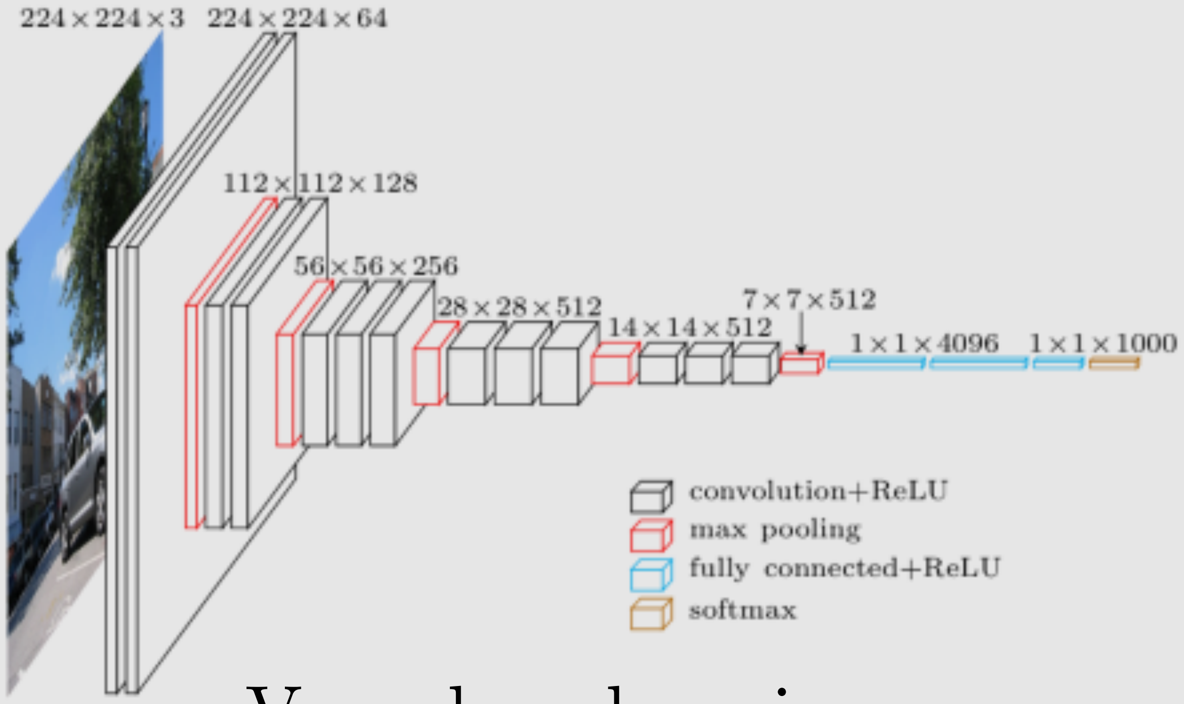


Image classification

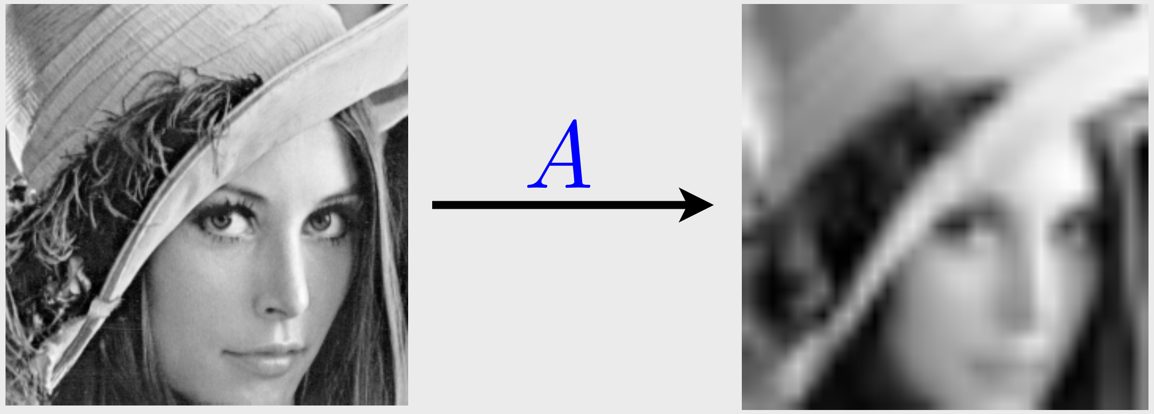


Very deep learning

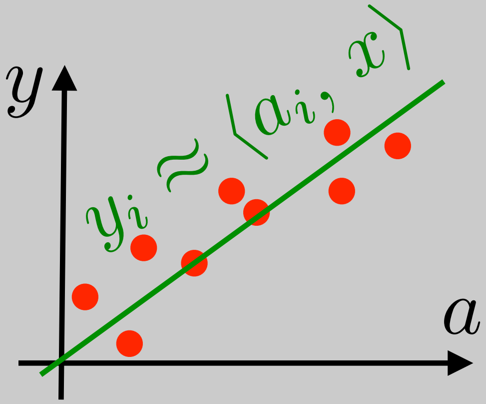
# Linear Models

Solving $y \underset{\approx}{=} A\,x \in \mathbb{R}^n$ $\qquad A \in \mathbb{R}^{n\times p}$

Imaging sciences:



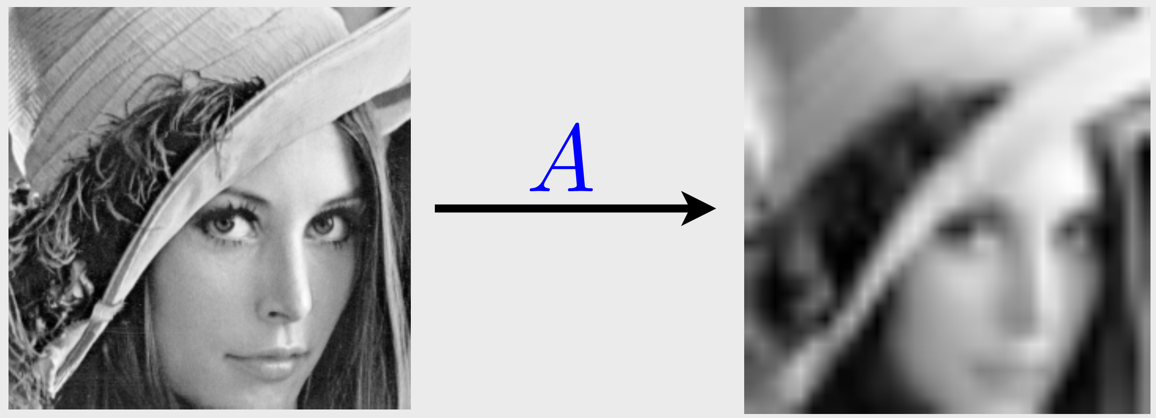
Machine learning:
Observations $(a_i, y_i)_{i=1}^p$,



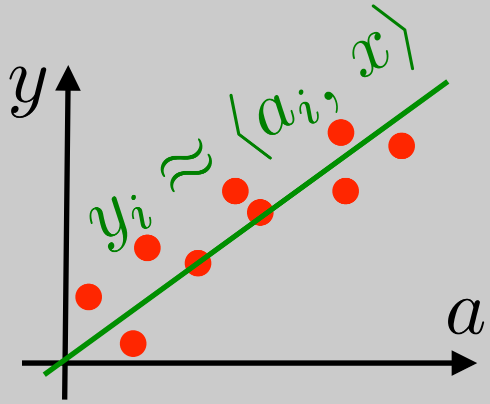
# Linear Models

Solving $y \underset{\approx}{=} A\, x \in \mathbb{R}^n$ $\quad A \in \mathbb{R}^{n \times p}$

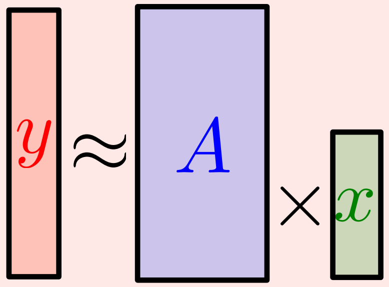
Imaging sciences:



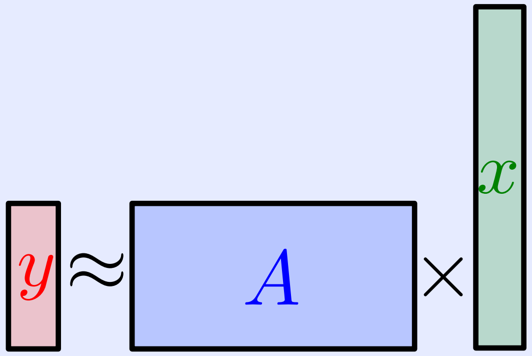
Machine learning:
Observations $(a_i, y_i)_{i=1}^p$,

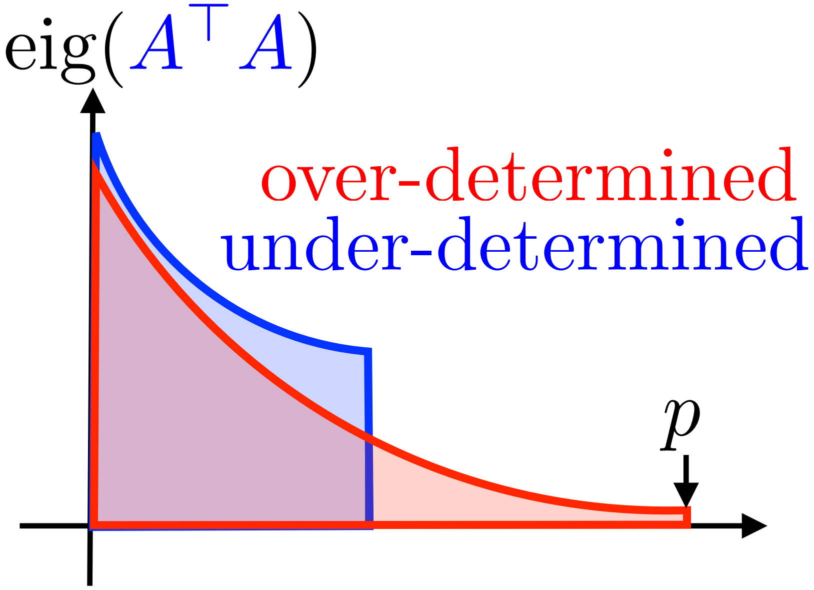


Over-determined $(n > p)$



Under-determined $(n < p)$

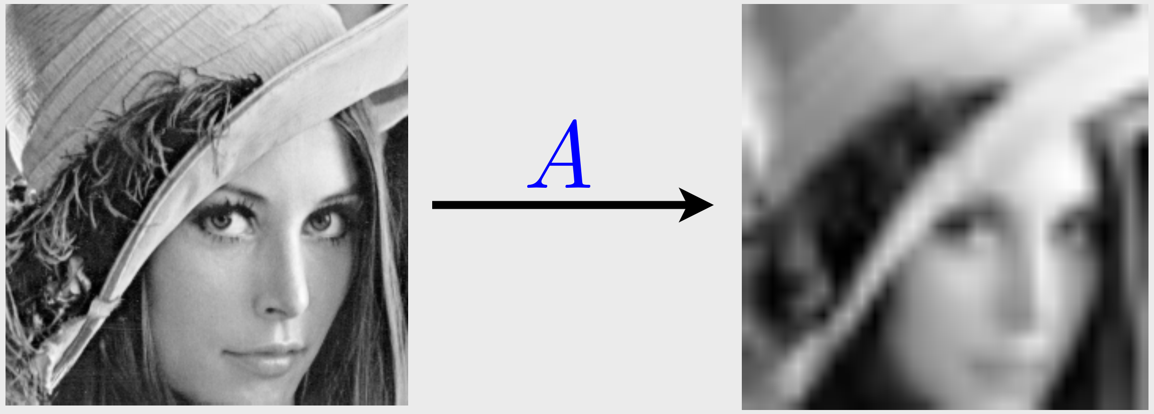




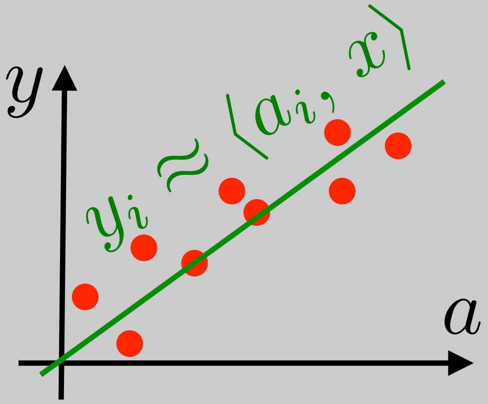
# Linear Models

Solving $y \underset{\approx}{=} A\, x \in \mathbb{R}^n$ $A \in \mathbb{R}^{n \times p}$

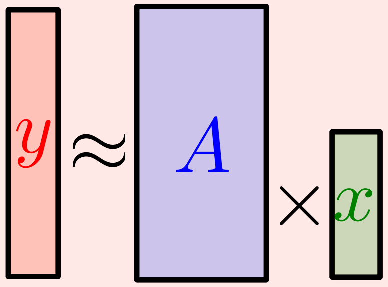
Imaging sciences:



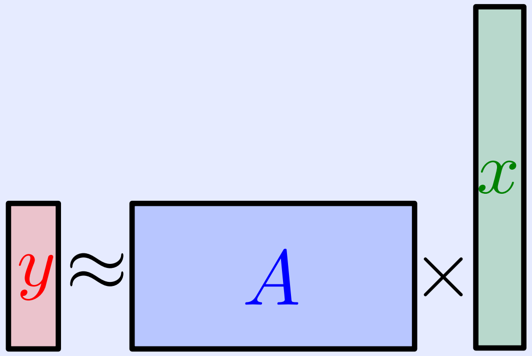
Machine learning:
Observations $(a_i, y_i)_{i=1}^p$,

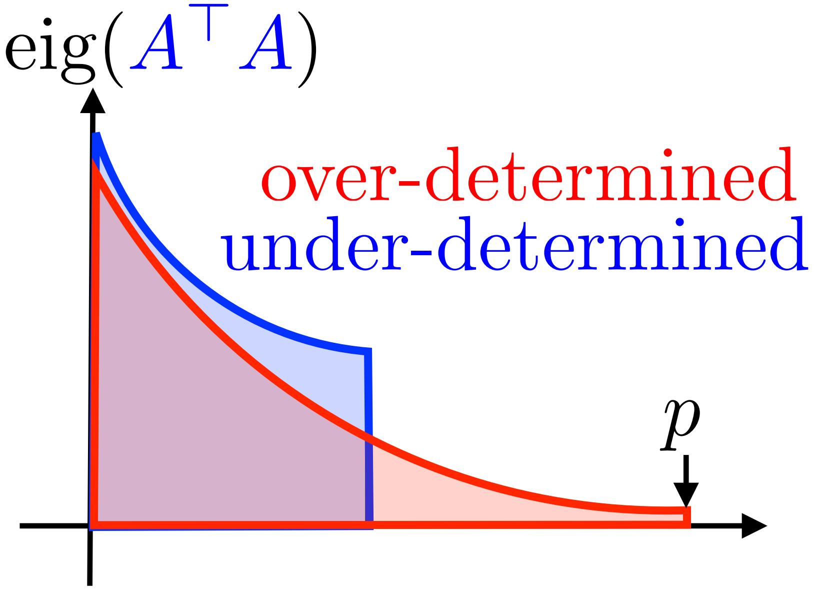


Over-determined $(n > p)$



Under-determined $(n < p)$





*Curse:* Ill-posed, noisy, large size $(n, p)$.

*Blessing:* unreasonable effectiveness of regularization in high dimension.

# Algorithms for large (n,p)

Regularized least square / empirical risk minimization:

$$\min_f \|Ax - y\|^2 + \lambda \|x\|^2$$

$$x = \underbrace{(A^\top A + \lambda \mathrm{Id}_p)^{-1} A^\top y}_{\substack{\text{If } n > p \\ \text{(over-determined)}}} = \underbrace{A^\top (A A^\top + \lambda \mathrm{Id}_n)^{-1} y}_{\substack{\text{If } n < p \\ \text{(under-determined)}}}$$

(kernel methods: $p = +\infty$)

# Algorithms for large (n,p)

Regularized least square / empirical risk minimization:

$$\min_f \|Ax - y\|^2 + \lambda\|x\|^2$$

$$x = \underbrace{(A^\top A + \lambda \mathrm{Id}_p)^{-1} A^\top y}_{\substack{\text{If } n > p \\ \text{(over-determined)}}} = \underbrace{A^\top (AA^\top + \lambda \mathrm{Id}_n)^{-1} y}_{\substack{\text{If } n < p \\ \text{(under-determined)}}}$$

(kernel methods: $p = +\infty$)

*Large but finite* $(n, p)$*:* use first order methods.

Gradient descent, CG, BFGS, proximal splittings.

$\longrightarrow$ $O(np)$ or even $O(p)$ cost per iterate.

$\longrightarrow$ Extends to non-smooth regularization (e.g. $\ell^1$).

# Algorithms for large (n,p)

Regularized least square / empirical risk minimization:

$$\min_f \|Ax - y\|^2 + \lambda\|x\|^2$$

$$x = \underbrace{(A^\top A + \lambda \mathrm{Id}_p)^{-1} A^\top y}_{\substack{\text{If } n > p \\ \text{(over-determined)}}} = \underbrace{A^\top (A A^\top + \lambda \mathrm{Id}_n)^{-1} y}_{\substack{\text{If } n < p \\ \text{(under-determined)}}}$$

(kernel methods: $p = +\infty$)

*Large but finite $(n, p)$:* use first order methods.

Gradient descent, CG, BFGS, proximal splittings.

$\longrightarrow$ $O(np)$ or even $O(p)$ cost per iterate.

$\longrightarrow$ Extends to non-smooth regularization (e.g. $\ell^1$).

*Very large or infinite $n$:* use stochastic descent methods.

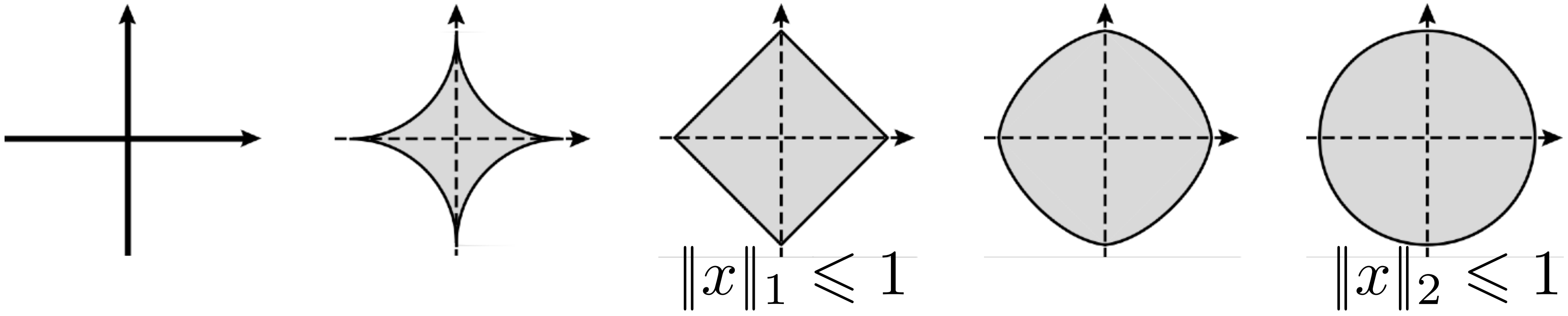
Draw $(y_i, a_i)$ at random, then $x \leftarrow (1 - \tau_k \lambda) x - \tau_k (\langle a_i,\, x\rangle - y_i) a_i$

decays to 0

# L1 and Dimensionality Reduction

Sparsity / model selection: replace $\|x\|^2$ by $\|x\|_1$.

$$\min_x \|Ax - y\|^2 + \lambda \|x\|_1$$



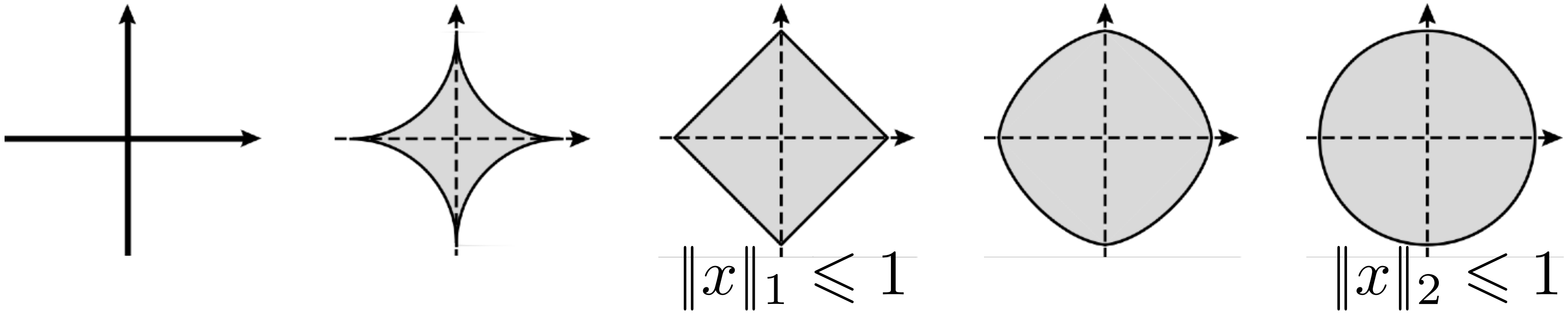
misleading
in high-dimension

$\longrightarrow$ Better model in imaging sciences. $\longrightarrow$ Support recovery with very large $p$.

# L1 and Dimensionality Reduction

Sparsity / model selection: replace $\|x\|^2$ by $\|x\|_1$.

$$\min_{x} \|Ax - y\|^2 + \lambda \|x\|_1$$

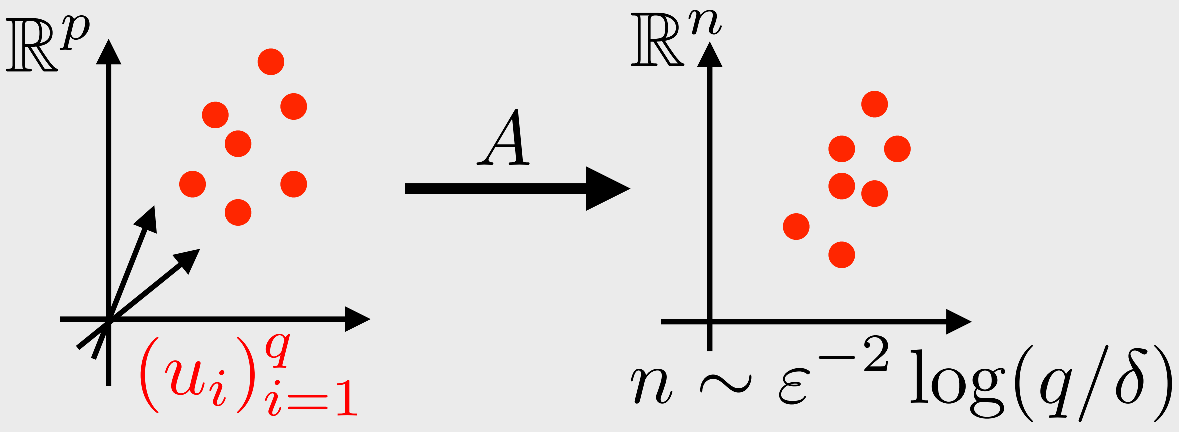


⚠ misleading
in high-dimension

$\longrightarrow$ Better model in imaging sciences. $\longrightarrow$ Support recovery with very large $p$.

"Optimal" setting: choose $A \in \mathbb{R}^{n \times p}$ random.

Johnson-Lindenstrauss lemma

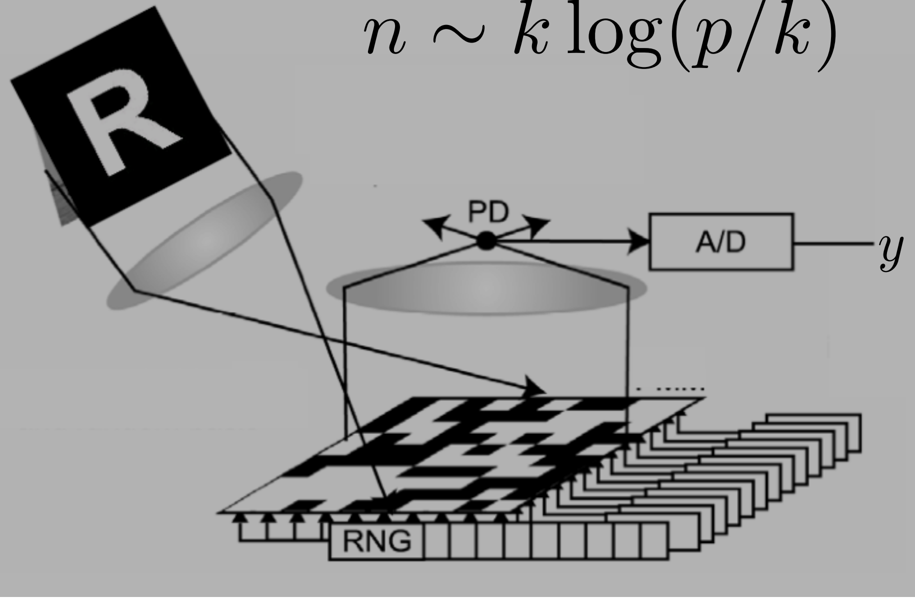


$$\frac{\|A(u_i - u_j)\|^2_{\mathbb{R}^n}}{\|u_i - u_j\|^2_{\mathbb{R}^p}} \in [(1-\varepsilon), (1+\varepsilon)]$$

Compressed sensing

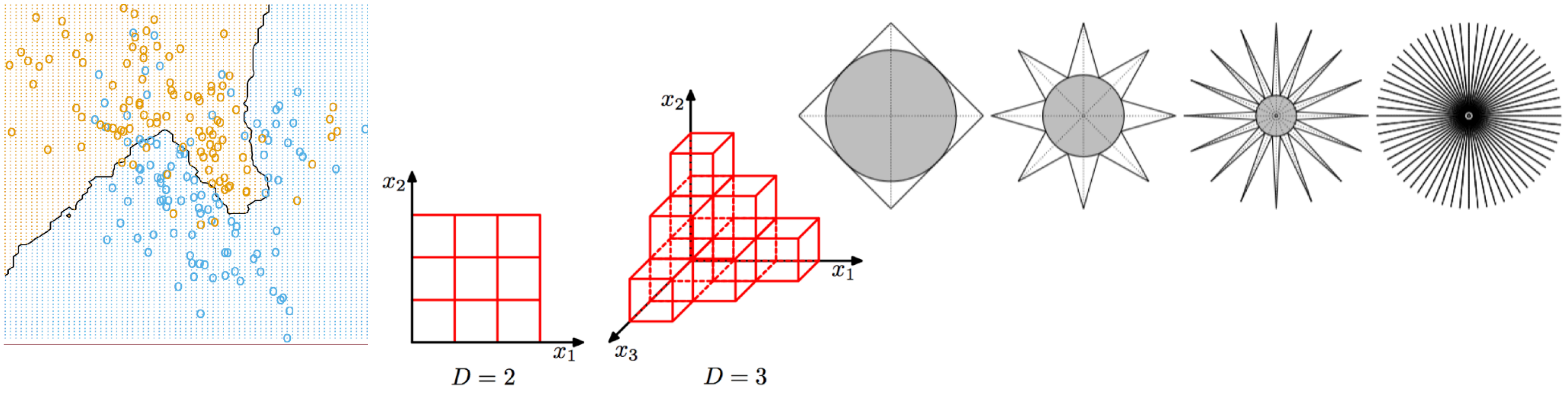
Perfect recovery of $k$-sparse input

$$n \sim k \log(p/k)$$



# What’s Next

**Julie Delon:** not so intuitive phenomena in high dimension.



**Jalal Fadili:** model selection in high-dimension.