Supervised Clustering in the Data Cube

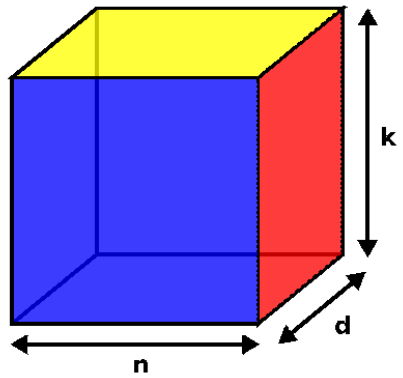
Vincent Roulet, Fajwel Fogel, Francis Bach, Alexandre D'Aspremont

INRIA, ENS Ulm, CNRS

Data Cube

- Large data in computer vision, bioinformatics, advertising...

- Large d, n and k
  - d : dimension of each observation
  - n : number of observations
  - k : number of tasks (classes)



- **Goal :** Help prediction by providing additional cluster structure on the features, the observations or the tasks

Data Cube

- Large data in computer vision, bioinformatics, advertising...
- Large d, n and k
  - d : dimension of each observation

  $\rightarrow$ *group words that share the same meaning for a given task*
  - n : number of observations

  $\rightarrow$ *group individuals that are well regressed together*
  - k : number of tasks (classes)

  $\rightarrow$ *group similar classes of images which share similarities*
- **Goal :** Help prediction by providing additional cluster structure on the features, the observations or the tasks

Supervised Clustering

- Minimize empirical loss

$$\min_{w \in \mathbb{R}^{d \times k}} \frac{1}{n} \sum_{i=1}^{n} l(y_i, w^T \phi(x_i))$$

  with w either
  - Constrained to have a clustered structure
  - or Regularized by a clustered structure
- Clustered structure
  - m objects (features, samples or tasks), Q clusters
  - Z matrix of assignment, $Z_{ij} = 1$ if i, j are in the same clusters, $Z_{ij} = 0$ otherwise
  - $C = (c_1, \ldots, c_Q)$ matrix of centers of each cluster
- See poster for detailed formulation

Supervised Clustering

- Non convex optimization
  - Projected Gradient whose projection reduced to K-means step
  - Conditional Gradient whose oracle reduced to K-means step
- Results
  - Fast optimization using K-means steps
  - Competitive results for topics prediction or sentiment analysis
  - Provide additional information on the datas