MULTIVIEW CANONICAL CORRELATION ANALYSIS

CANONICAL CORRELATION ANALYSIS

- Given **N sources,** let be a dataset $\{X_m \in R^{D_m \times N}\}_{m=1}^{M}$ collected from M= 2 views.

- The **goal of CCA** is to find lower dimensional $d \ll D_m$ representation of these two views through a linear projection U_1 and U_2 while preserving cross information between two views.

$$[U_1^*, , U_2^*] = \boldsymbol{argmax}_{\{U_1, U_2\}} Corr(U_1 X_1, U_2 X_2)$$

- **It has been shown that CCA can increase the quality of clustering and various machine learning tasks.**

- CCA remains limited because it can only deal with two views and captures linear relationships.

- This original problem leads to various extensions more or less difficult to incorporate :

Multiviews M>2 / Non-linear extension /Awareness of potential geometric structure on sources etc.

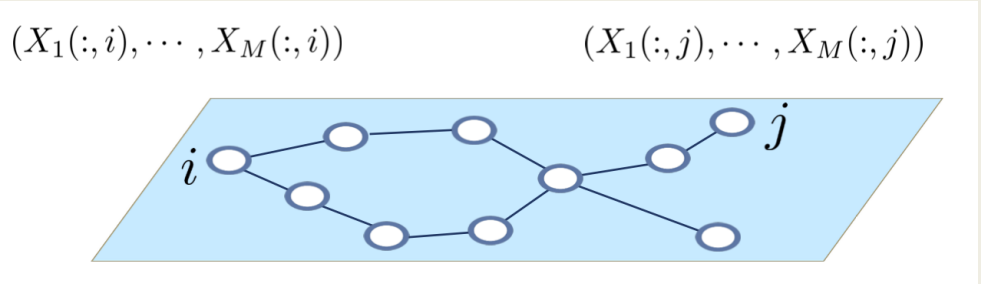
- These extensions can be caracterised by some key properties :

Method	Complexity	Non Linear	>2 views	Graph
CCA	$O(n)$	✗	✗	✗

MULTIVIEW CANONICAL CORRELATION ANALYSIS

Multiview Varitional Graph CANONICAL CORRELATION ANALYSIS

- Sources can rely on a graph. **Taking care of this graph can improve results** but it's increase complexity.



- We propose a probabilistic model based on an existing **equivalence between original CCA and a bayesian problem** and **solved with a variational auto encoders.**

- $p(X_1, X_2, \ldots, X_M, A) = \int P(X_1, A|Z) \ldots P(X_M, A|Z)\, P(Z) dZ$

- This model is the only model which deals simultaneously with :

  - M $>$ 2 views,
  - Non linear
  - Accounting for geometric structure
  - Scalable
  - Robust to missing views