GEneralised Mutual INformatIon Maximisation for Deep Clustering

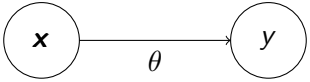
Louis Ohl
Joint work with Pierre-Alexandre Mattei, Charles Bouveyron, Frederic Precioso, Warith Harchaoui, Arnaud Droit, Mickael Leclercq

Université Côte d'Azur, I3S UMR 7271 CNRS, Inria Maasai project-team

October 26, 2021

Motivations

- Can we perform clustering with a conditional model $p_\theta(y|\boldsymbol{x})$, e.g. logistic regression or a deep neural net?
- Then, how can we perform model selection with the "right" amount of clusters?



Maximising the GEneralised Mutual INformatIon (GEMINI)

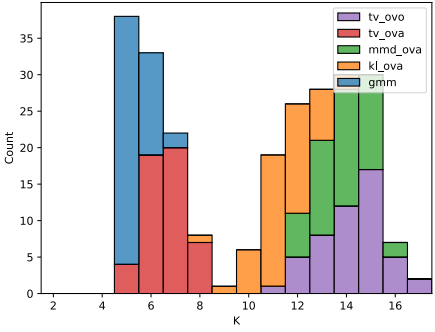
$$\mathcal{I}(X;Y) = \mathbb{E}_{y\sim p_\theta(y)}\left[D_{\mathrm{KL}}(p_\theta(\boldsymbol{x}|y)||p_{\mathrm{data}}(\boldsymbol{x}))\right] \tag{1}$$

- Replace D_{KL} by other distances D: f-divergences, integral probability metrics
- Thanks to the known discriminative model $p_\theta(y|\boldsymbol{x})$, we can estimate the cluster distribution $p_\theta(y)$. The remaining is unknown.

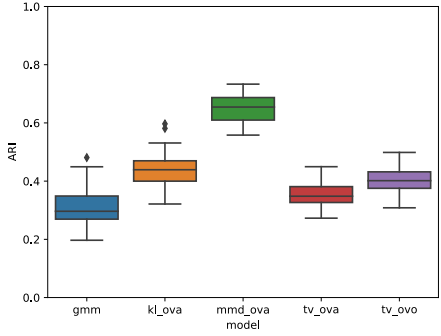
$$\mathcal{I}_D^{\mathrm{ova}}(X;Y) = \mathbb{E}_{y\sim p_\theta(y)}\left[D(p_\theta(\boldsymbol{x}|y)||p_{\mathrm{data}}(\boldsymbol{x}))\right] \tag{2}$$

$$\mathcal{I}_D^{\mathrm{ovo}}(X;Y) = \mathbb{E}_{y_a,y_b\sim p_\theta(y)}\left[D(p_\theta(\boldsymbol{x}|y_a)||p_\theta(\boldsymbol{x}|y_b))\right] \tag{3}$$

A glimpse at results: implicit model selection



(a) USPS - Cluster number selection



(b) USPS - ARI

Conclusions

- A new discriminative method objective, suitable with deep learning
- Implicit cluster number selection

Thank you for your attention!

(More with P.A. Mattei on Thursday!)