Variational Canonical Correlation Analysis with Kernel Methods

Polina Barabanshchikova

MIPT

January 13, 2024

Variational Canonical Correlation Analysis

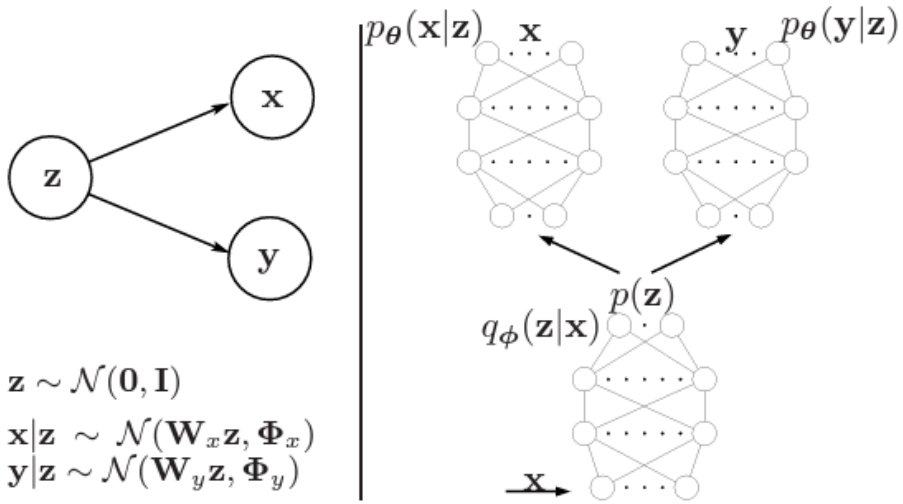

Figure 1: Left: Probabilistic latent variable interpretation of CCA. Right: Deep variational CCA[2].

ELBO:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}; \theta, \phi) := -\mathcal{D}_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_{\theta(\mathbf{x}|\mathbf{z})} + \log p_{\theta(\mathbf{y}|\mathbf{z})}]$$

[1] Weiran Wang, Honglak Lee, and Karen Livescu. "Deep Variational Canonical Correlation Analysis". In: *CoRR* abs/1610.03454 (2016). arXiv: 1610.03454. URL: http://arxiv.org/abs/1610.03454.

[2] Wang, Lee, and Livescu, "Deep Variational Canonical Correlation Analysis".

Kernel KL divergence

- **Problem**: KL divergence is hard to estimate beyond Gaussian distributions
- **Covariance operator**[3]: $\Sigma_p = \int_{\mathcal{X}} \phi(x)\phi(x)^* dp(x)$
- **Kernel KL divergence**: $\mathcal{D}(\Sigma_p||\Sigma_q) = \operatorname{tr}\left[\Sigma_p(\log \Sigma_p - \log \Sigma_q)\right]$
- **Properties**:
  - Always non-negative, and equal to 0 for $p = q$
  - Jointly convex in (p, q)
  - $0 \leq \mathcal{D}(p||q) - \mathcal{D}(\Sigma_p||\Sigma_q) \leq E(p, q) \times \Delta(k)$
- **Estimators**:
  - $\hat{\Sigma}_p = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i)\phi(x_i)^*$, i.i.d. $x_1, \ldots, x_n \sim p(x)$
  - $\operatorname{tr}\left[\hat{\Sigma}_p \log \hat{\Sigma}_p\right] = \operatorname{tr}\left[\frac{1}{n}K \log\left(\frac{1}{n}K\right)\right]$, where K – kernel matrix

[3] Francis Bach. *Information Theory with Kernel Methods*. 2022. arXiv: 2202.08545 [cs.IT].

Project description

Title: Variational Canonical Correlation Analysis with Kernel Methods
Problem: Given multi-view data, extract common information between the views
Data: Noisy MNIST
Reference: DVCCA[4], Variational Interpretable CCA[5], Kernel Methods[6]
Basic solution: Deep Variational CCA
Proposed solution: Enrich variational family by using Kernel KL divergence instead of $D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$.
Novelty: Address the limitation of VCCA. Propose a use case for Kernel KL divergence.

---

[4] Wang, Lee, and Livescu, "Deep Variational Canonical Correlation Analysis".

[5] Lin Qiu, Vernon M. Chinchilli, and Lin Lin. "Variational Interpretable Deep Canonical Correlation Analysis". In: *ICLR2022 Machine Learning for Drug Discovery*. 2022. URL: https://openreview.net/forum?id=Gzare7_sTAJ.

[6] Bach, *Information Theory with Kernel Methods*.