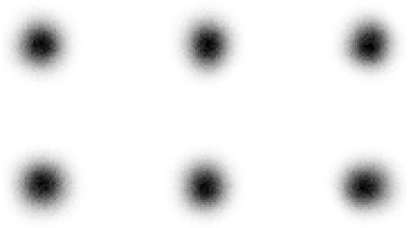
Applications of Kernel-based Information Theoretical Measures

Zoltán Szabó

PhD Open Day, LSE
Oct 14, 2021
(morning)

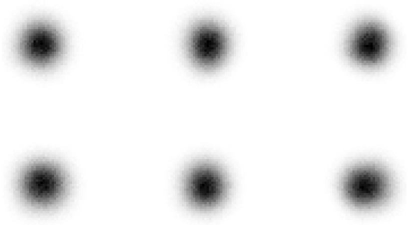
Puzzle

Observation:



Puzzle

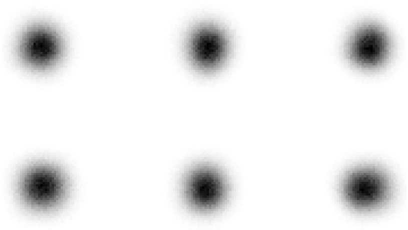
Observation:



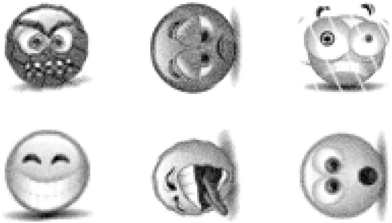
Let us estimate hidden independent 2d sources

Puzzle

Observation:



Let us estimate hidden independent 2d sources:



Independent subspace analysis

Observation:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \qquad \mathbf{s} = \left[\mathbf{s}^1; \ldots; \mathbf{s}^M\right].$$

Goal: $\hat{\mathbf{s}}$ from $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$. Assumptions:

1. independent groups: $\mathrm{I}\left(\mathbf{s}^1, \ldots, \mathbf{s}^M\right) = 0$,
2. $\mathbf{s}^m$-s: non-Gaussian,
3. $\mathbf{A}$: invertible.

In the example: $\mathbf{s}_m \in \mathbb{R}^2$, $M = 6$.

Independent subspace analysis

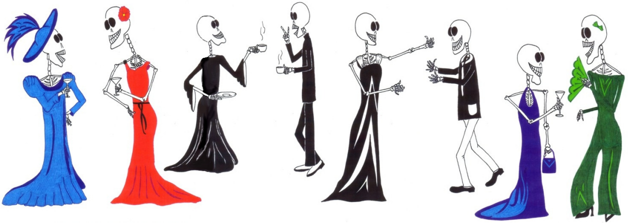
Observation:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \qquad \mathbf{s} = \left[\mathbf{s}^1; \ldots ; \mathbf{s}^M\right].$$

Goal: $\hat{\mathbf{s}}$ from $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$. Assumptions:

1. independent groups: $I\left(\mathbf{s}^1, \ldots, \mathbf{s}^M\right) = 0$,
2. $\mathbf{s}^m$-s: non-Gaussian,
3. $\mathbf{A}$: invertible.

In the example: $\mathbf{s}_m \in \mathbb{R}^2$, $M = 6$. Various applications:



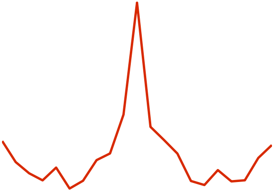
Outlier-robust image registration

Given two images:

Goal: find the transformation which takes the right one to the left.

Given two images:



Goal: find the transformation which takes the right one to the left.

In equations

- Reference image: $\mathbf{y}_{\text{ref}}$,
- test image: $\mathbf{y}_{\text{test}}$,
- possible transformations: $\boldsymbol{\Theta}$.

Trick

maximization of statistical dependence

Objective:

$$J(\boldsymbol{\theta}) = \underbrace{\mathrm{I}(\mathbf{y}_{\text{ref}}, \mathbf{y}_{\text{test}}(\boldsymbol{\theta}))}_{\text{similarity}} \to \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}}.$$

Feature selection

- **Goal**: find
  - the feature subset (# of rooms, criminal rate, local taxes)
  - most relevant for house price prediction (y).

Feature selection

- **Goal**: find
  - the feature subset (# of rooms, criminal rate, local taxes)
  - most relevant for house price prediction (y).

- Features: $x^1, \ldots, x^F$. Subset: $S \subseteq \{1, \ldots, F\}$.

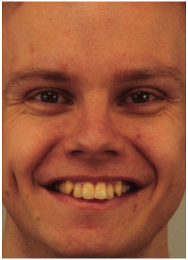
Feature selection

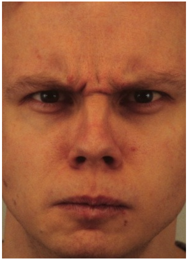
- **Goal**: find
  - the feature subset (# of rooms, criminal rate, local taxes)
  - most relevant for house price prediction (y).

- Features: $x^1, \ldots, x^F$. Subset: $S \subseteq \{1, \ldots, F\}$.
- MaxRelevance - MinRedundancy principle:

$$J(S) = \frac{1}{|S|} \sum_{i \in S} \mathrm{I}\left(x^i, y\right) - \frac{1}{|S|^2} \sum_{i,j \in S} \mathrm{I}\left(x^i, x^j\right) \to \max_{S \subseteq \{1,\ldots,F\}}.$$

Differentiating positive and negative emotions





- Given: two sets of faces (happy, angry).
- Task:

Do $\{x_i\}$ and $\{y_j\}$ come from the same distribution, i.e. $\mathbb{P}_x = \mathbb{P}_y$?

Determine the most discriminative features/regions.

Translation checking

- We have paired samples:

x_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

x_2: No doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development.

...

y_1: Honorables sénateurs, ma question s'adresse au leader du gouvernement au Sénat et concerne l'aide financiére qu'on a annoncée pour les agriculteurs. La plupart des agriculteurs n'ont encore rien reu de cet argent.

y_2: Il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants.

...

Translation checking

- We have paired samples:

x_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

x_2: No doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development.

...

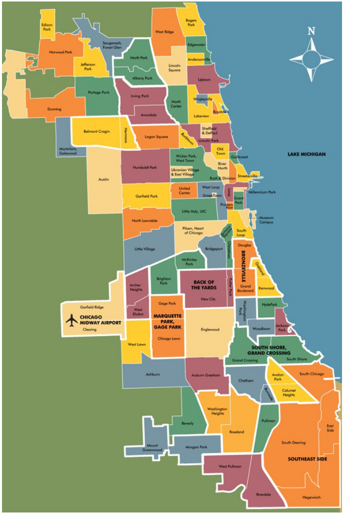
y_1: Honorables sénateurs, ma question s'adresse au leader du gouvernement au Sénat et concerne l'aide financiére qu'on a annoncée pour les agriculteurs. La plupart des agriculteurs n'ont encore rien reu de cet argent.

y_2: Il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants.

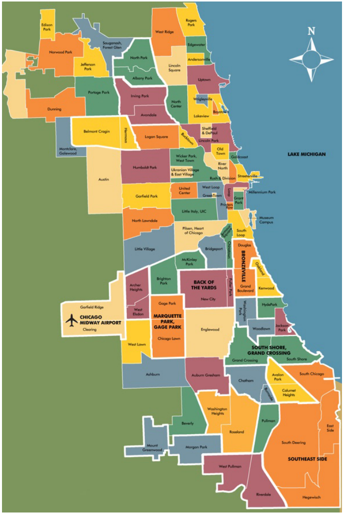
...

Are the French paragraphs translations of the English ones, or have nothing to do with it, i.e. $\mathbb{P}_{xy} = \mathbb{P}_x \mathbb{P}_y$?

- Given:
    - Density/model: p.

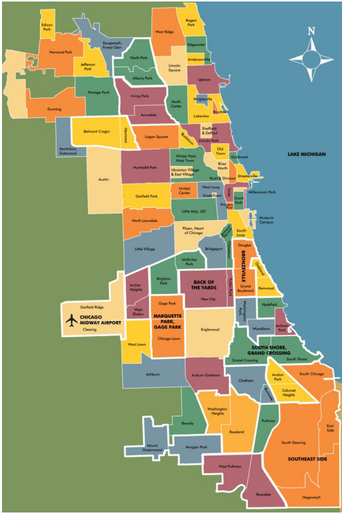


- Given:
  - Density/model: p.
  - Samples: $X = \{\mathbf{x}_i\}_{i=1}^n \sim q$ (unknown).



- Given:
  - Density/model: p.
  - Samples: $X = \{\mathbf{x}_i\}_{i=1}^n \sim q$ (unknown).
- Task: using p, X test

$$H_0 : p = q, \quad \text{vs}$$
$$H_1 : p \neq q.$$



Kernels → mean embedding, MMD, HSIC

- Kernel:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

Trick

⇒ rich and computationally tractable function classes ($\mathcal{H}_k$).

Kernels $\to$ mean embedding, MMD, HSIC

- Kernel:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

Trick

$\Rightarrow$ rich and computationally tractable function classes ($\mathcal{H}_k$).

- Mean embedding:

$$\mathbb{P} \mapsto \mu_k(\mathbb{P}) = \int_{\mathcal{X}} \underbrace{\varphi(x)}_{\text{example: } e^{i\langle \cdot, x\rangle} \text{ in } \mathbb{R}^d} \mathrm{d}\mathbb{P}(x).$$

Kernels $\to$ mean embedding, MMD, HSIC

- Kernel:

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

Trick

$\Rightarrow$ rich and computationally tractable function classes ($\mathcal{H}_k$).

- Mean embedding:

$$\mathbb{P} \mapsto \mu_k(\mathbb{P}) = \int_{\mathcal{X}} \underbrace{\varphi(x)}_{\text{example: } e^{i\langle \cdot, x\rangle} \text{ in } \mathbb{R}^d} \mathrm{d}\mathbb{P}(x).$$

- Maximum mean discrepancy (KGOF similarly):

$$\mathrm{MMD}_k(\mathbb{P}, \mathbb{Q}) = \|\mu_k(\mathbb{P}) - \mu_k(\mathbb{Q})\|_{\mathcal{H}_k}.$$

Kernels $\to$ mean embedding, MMD, HSIC

- Kernel:

$$k(x,y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}.$$

Trick

$\Rightarrow$ rich and computationally tractable function classes ($\mathcal{H}_k$).

- Mean embedding:

$$\mathbb{P} \mapsto \mu_k(\mathbb{P}) = \int_{\mathcal{X}} \underbrace{\varphi(x)}_{\text{example: } e^{i\langle \cdot, x\rangle} \text{ in } \mathbb{R}^d} \mathrm{d}\mathbb{P}(x).$$

- Maximum mean discrepancy (KGOF similarly):

$$\text{MMD}_k(\mathbb{P}, \mathbb{Q}) = \|\mu_k(\mathbb{P}) - \mu_k(\mathbb{Q})\|_{\mathcal{H}_k}.$$

- Hilbert-Schmidt independence criterion:

$$\text{HSIC}_k(\mathbb{P}) = \text{MMD}_k\left(\mathbb{P}, \otimes_{m=1}^M \mathbb{P}_m\right).$$

Thank you for the attention!

ITE on BitBucket! (playground)