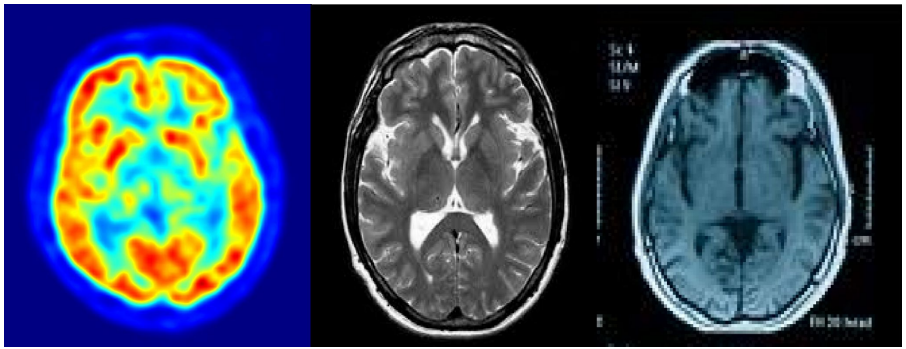
L0-Based Sparse Canonical Correlation Analysis

Ofir Lindenbaum, Moshe Salhov, Amir Averbuch, Yuval Kluger

Multimodal Learning

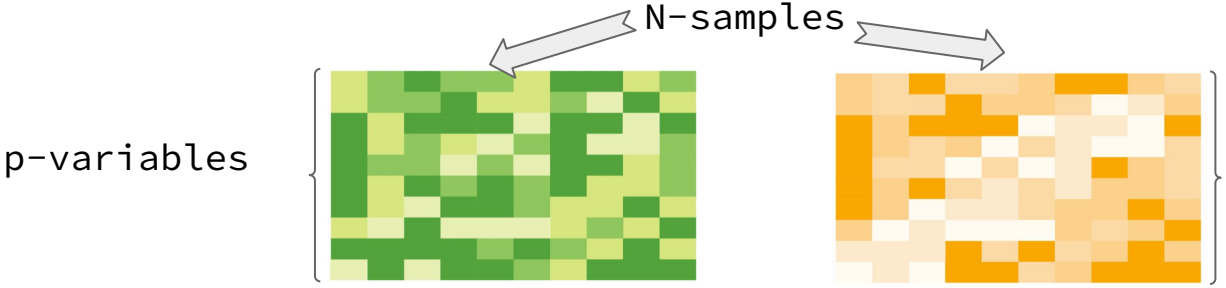
We often observe multiple coupled modalities



How can we fuse the multimodal observations?

Canonical Correlation Analysis (CCA)- Hotelling 1936

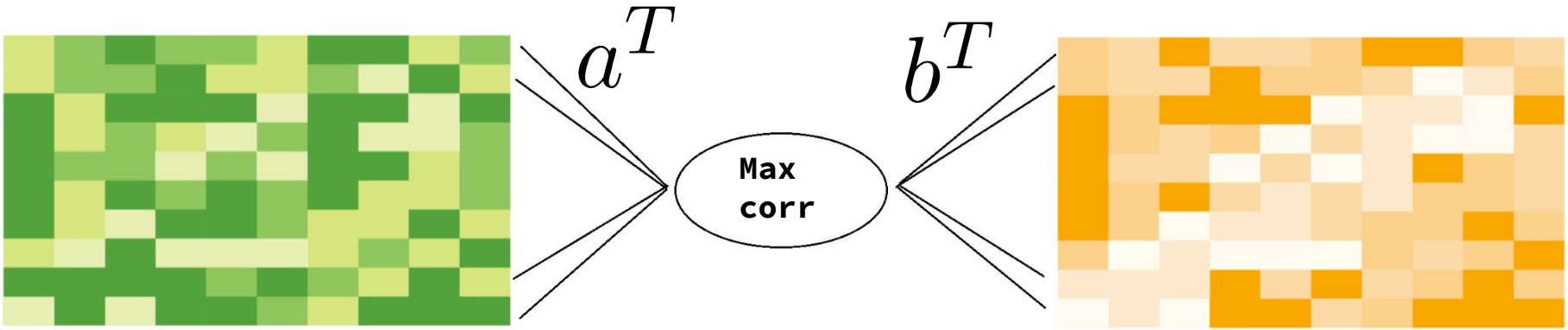
Consider an observed coupled data $X \in \mathbb{R}^{p \times N}$ and $Y \in \mathbb{R}^{q \times N}$

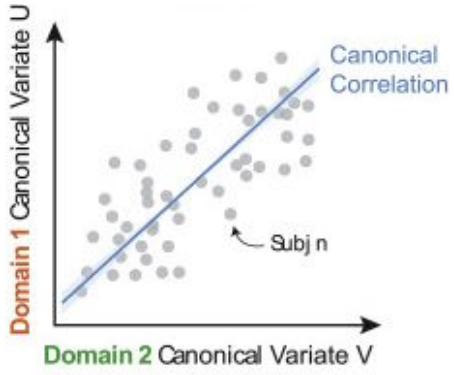


CCA seeks *canonical vectors* $a \in \mathbb{R}^{D^x}$, and $b \in \mathbb{R}^{D^y}$, such that , $u = a^T X$, and $v = b^T Y$ will maximize the sample correlations between the *canonical variates*:

$$\max_{a,b \neq 0} \quad \rho(a^T X, b^T Y) = \frac{a^T X Y^T b}{\|a^T X\|_2 \|b^T Y\|_2}.$$

Canonical Correlation Analysis (CCA)





Canonical Correlation Analysis (CCA)

The problem has a closed form solution using eigenvectors

a^i the eigenvectors of $C_x^{-1}C_{xy}C_y^{-1}C_{yx}$ are the ***canonical vectors***

b^i the eigenvectors of $C_y^{-1}C_{yx}C_x^{-1}C_{xy}$ are the ***canonical vectors***

where C_x, C_y are within view sample covariance matrices and C_{xy}, C_{yx} are cross-view sample covariance matrices.

These provide multimodal ***canonical variates***

$$u^i = (a^i)^T X \text{ and } v^i = (b^i)^T Y$$

CCA- Degeneracy

$X \in \mathbb{R}^{p \times N}$ and $Y \in \mathbb{R}^{q \times N}$ if $N < q$ or $N < p$ CCA “breaks”

$$\max_{a,b \neq 0} \quad \rho(a^T X, b^T Y) = \frac{a^T X Y^T b}{\|a^T X\|_2 \|b^T Y\|_2}.$$

a is an eigenvector of $C_x^{-1} C_{xy} C_y^{-1} C_{yx}$ but the covariances are not accurate, and not invertible

Solution: sparsify **a** and **b**

Sparse CCA

To sparsify the number of variables used in each view, we can minimize

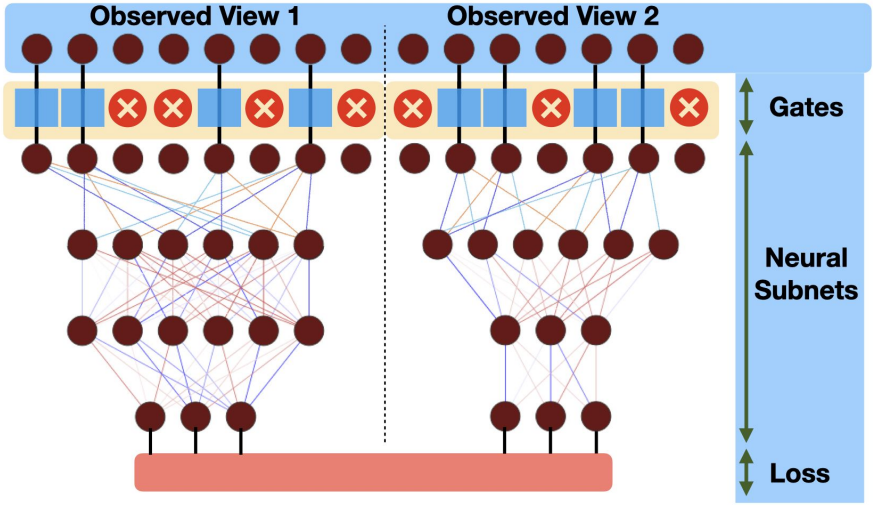
$$\min_{a,b} \quad -\rho(a^T X, b^T Y) + \lambda^x \|a\|_0 + \lambda^y \|b\|_0,$$

where λ^x and λ^y are regularization parameters which control the sparsify the input variables.

If $\|a\|_0$ and $\|b\|_0$ are smaller than N, we remove the degeneracy

LO DEEP CCA

$$f(\hat{X}) = f(\mathbf{z}^x \odot X|\theta^x) \in \mathbb{R}^{d\times N} \qquad g(\hat{Y}) = g(\mathbf{z}^y \odot Y|\theta^y) \in \mathbb{R}^{d\times N}$$

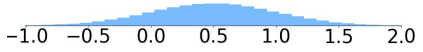


$$\mathbb{E}\big[-\bar{\rho}(f(\hat{X}), g(\hat{Y})) + \lambda^x \|z^x\|_0 + \lambda^y \|z^y\|_0\big]$$

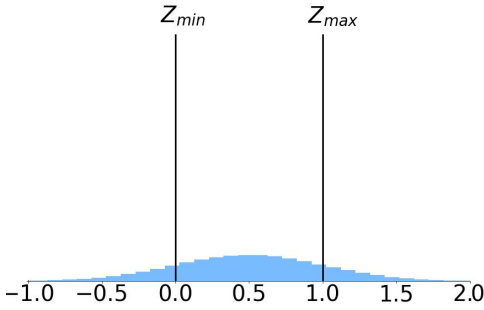
10 Deep CCA- Stochastic Gates (STG)-Yamada et al. 2020

The idea: use a truncated Gaussian to relax the Bernoulli distribution

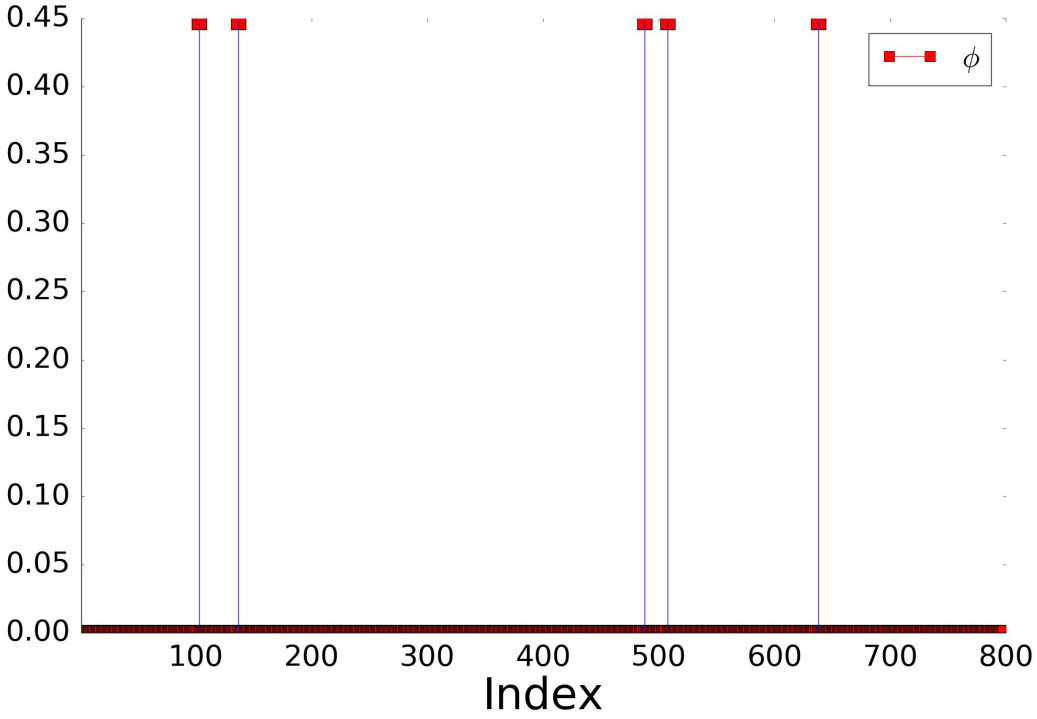
Draw from a Gaussian $\epsilon \sim N(0, 0.5)$, shift by $\mu = 0.5$



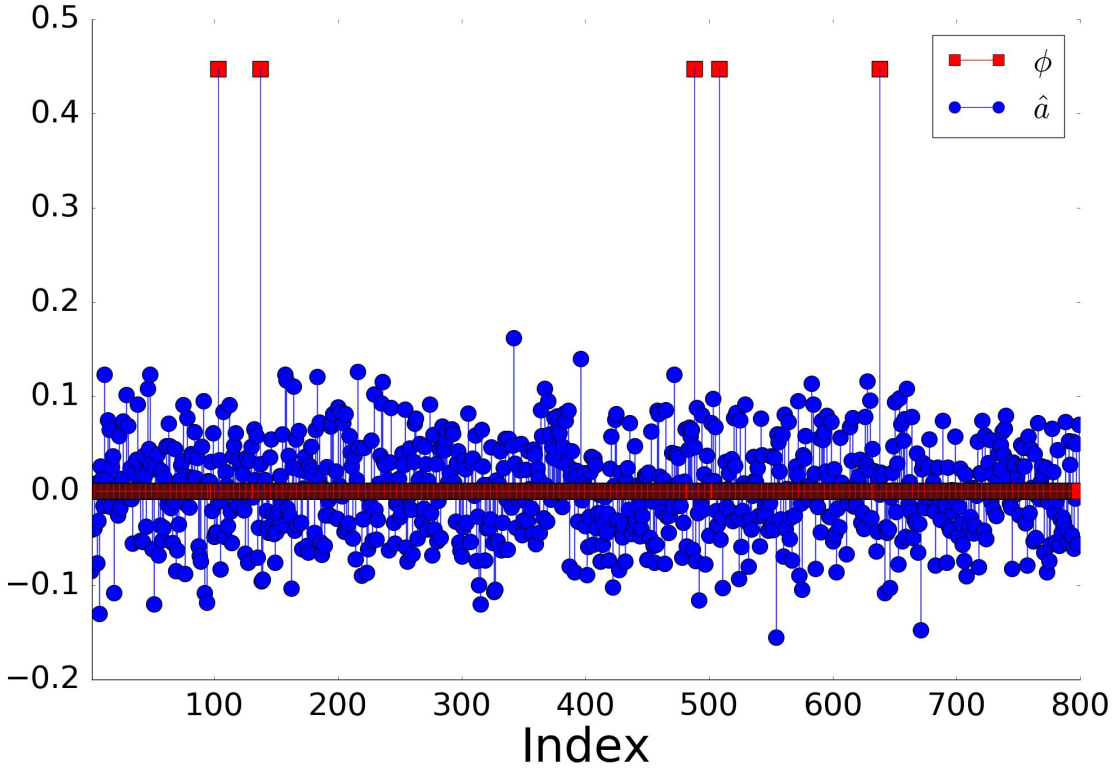
Truncate into $[0, 1]$ using $z = \max(0, \min(1, \mu + \epsilon))$



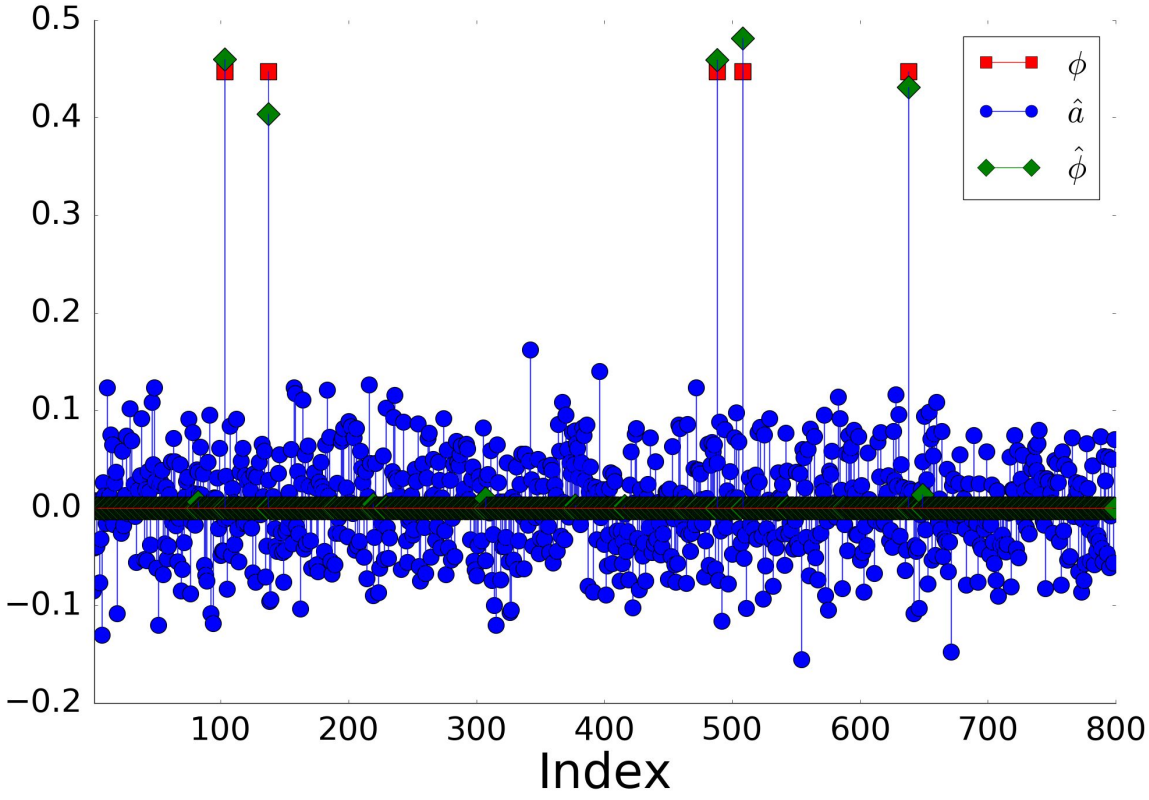
LO CCA- Linear Example



LO CCA- Linear Example

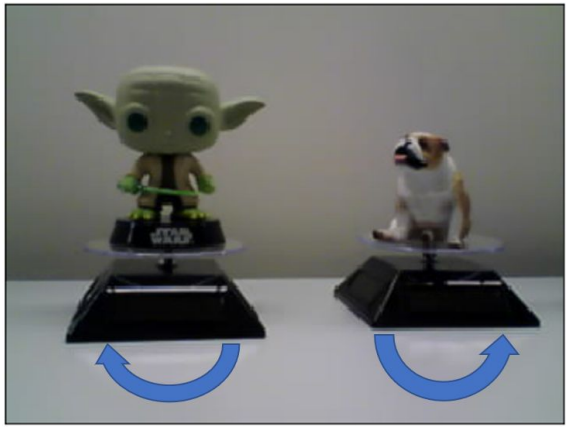


LO CCA- Linear Example



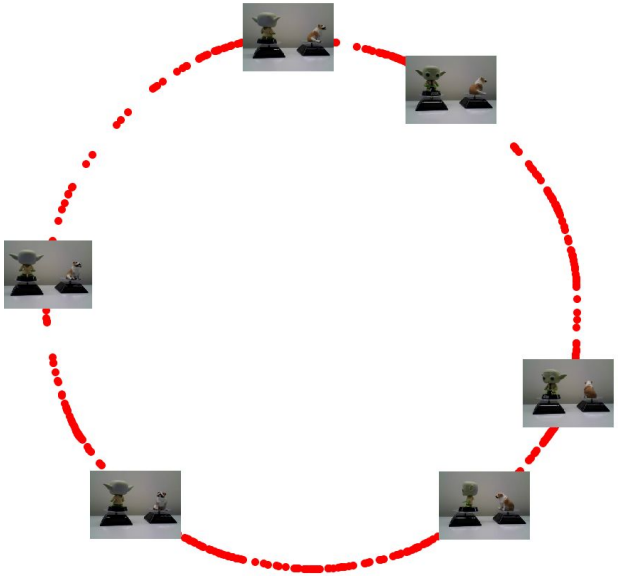
Toy Example

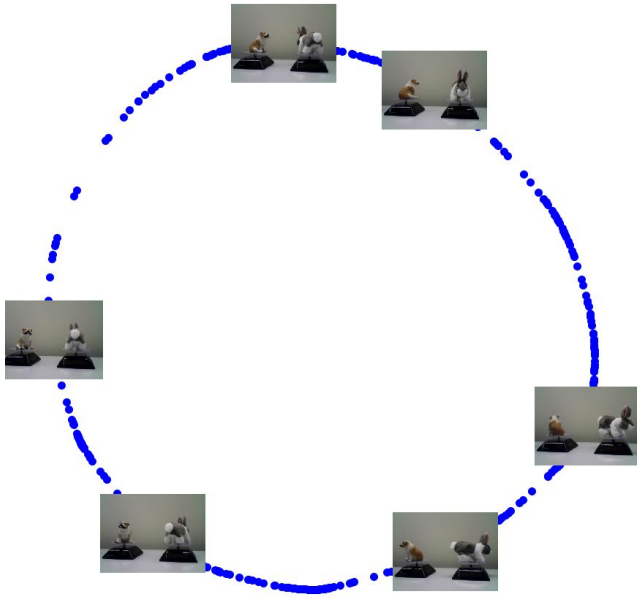
Toy Example



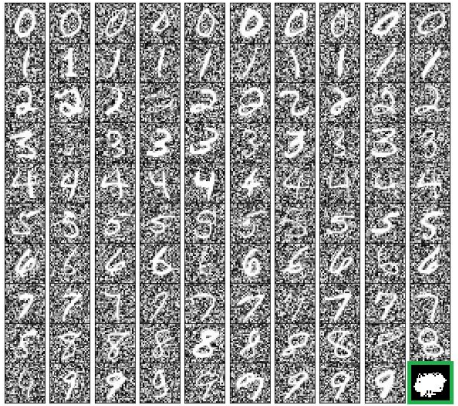
BIU- Ofir Lindenbaum

Toy Example



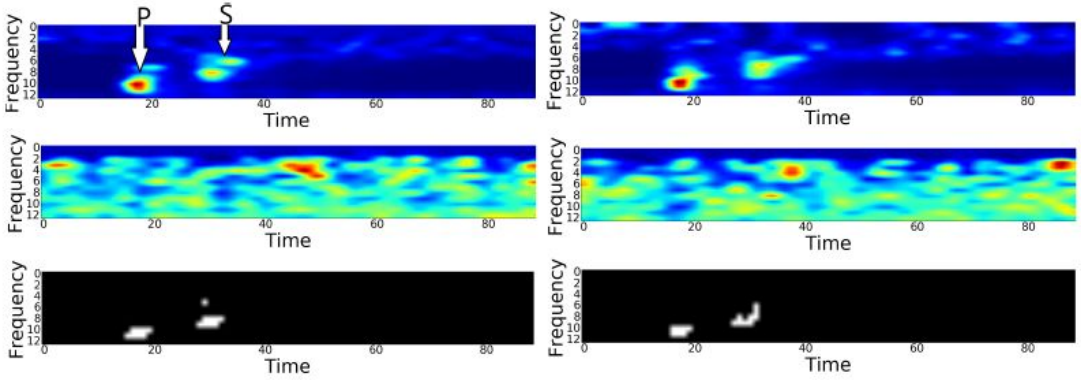


Noisy MNIST



Method	Noisy MNIST (LeCun et al., 2010)		
	MI	KM (%)	SVM (%)
Raw Data	0.130	16.6	86.6
PCA (Pearson, 1901)	0.130	16.6	89.3
CCA (Chaudhuri et al., 2009)	1.290	66.4	75.8
mod-SCCA (Suo et al., 2017)	0.342	23.9	63.1
SCCA-HSIC (Uurtio et al., 2018)	NA	NA	NA
KCCA (Bach & Jordan, 2002)	0.943	50.2	85.3
grad-KCCA (Uurtio et al., 2019)	NA	NA	NA
multiview-ICA (Richard et al., 2020)	1.750	88.0	90.0
NCCA (Michaeli et al., 2016)	1.030	47.5	77.2
DCCA (Andrew et al., 2013)	1.970	93.2	93.2
DCCAE (Wang et al., 2015b)	1.940	91.8	94.0
ℓ_0-CCA (linear)	1.73	87.1	88.4
ℓ_0-DCCA (non-linear)	**2.05**	**95.4**	**95.5**

Seismic Data



Method	Seismic (Lindenbaum et al., 2018)		
	MI	KM (%)	SVM (%)
Raw Data	0.001	35.7	41.3
PCA (Pearson, 1901)	0.002	38.8	41.3
CCA (Chaudhuri et al., 2009)	0.003	38.1	40.4
mod-SCCA (Suo et al., 2017)	0.610	71.7	86.9
SCCA-HSIC (Uurtio et al., 2018)	0.003	38.7	49.5
KCCA (Bach & Jordan, 2002)	0.006	38.4	92.5
grad-KCCA (Uurtio et al., 2019)	0.005	40.9	41.4
multiview-ICA (Richard et al., 2020)	0.748	90.1	94.2
NCCA (Michaeli et al., 2016)	0.700	86.8	91.4
DCCA (Andrew et al., 2013)	0.830	94.9	94.6
DCCAE (Wang et al., 2015b)	0.92	97.0	97.0
ℓ_0-CCA (linear)	0.85	93.7	94.9
ℓ_0-DCCA (non-linear)	**0.97**	**98.1**	**97.2**

BIU- Ofir Lindenbaum