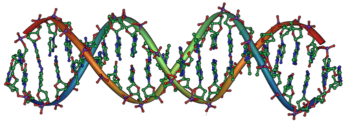
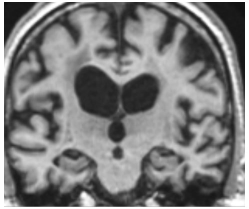
Latent variable models for the analysis of heterogeneous information

Marco Lorenzi

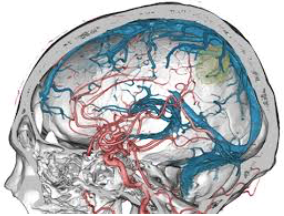
Epione Research Group, Université Côte d'Azur, Inria

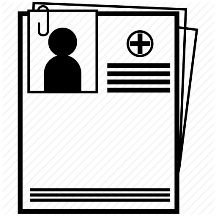
Data Science meets biomedical research





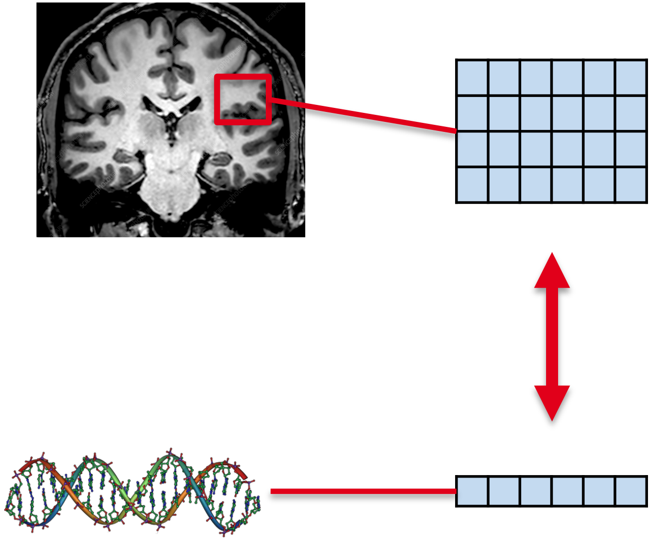
baseline





…

Variability of multivariate biomedical data
- within/between views -



Signal

biology

bias field (scanner)

noise

processing bias
(e.g. imputation)

Relationship

Causal

Mediated

Spurious

Must be Inferred

Variability of multivariate biomedical data

- between datasets -

Population

Acquisition

Processing

Data security

…

Latent variable models

1. *Multi-variate* modeling

2. Novel scalable approaches to *multi-view* data

Latent variable models

1. *Multi-variate* modeling

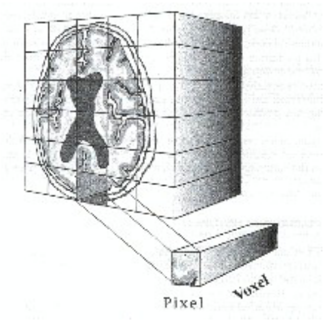
2. Novel scalable approaches to multi-view data

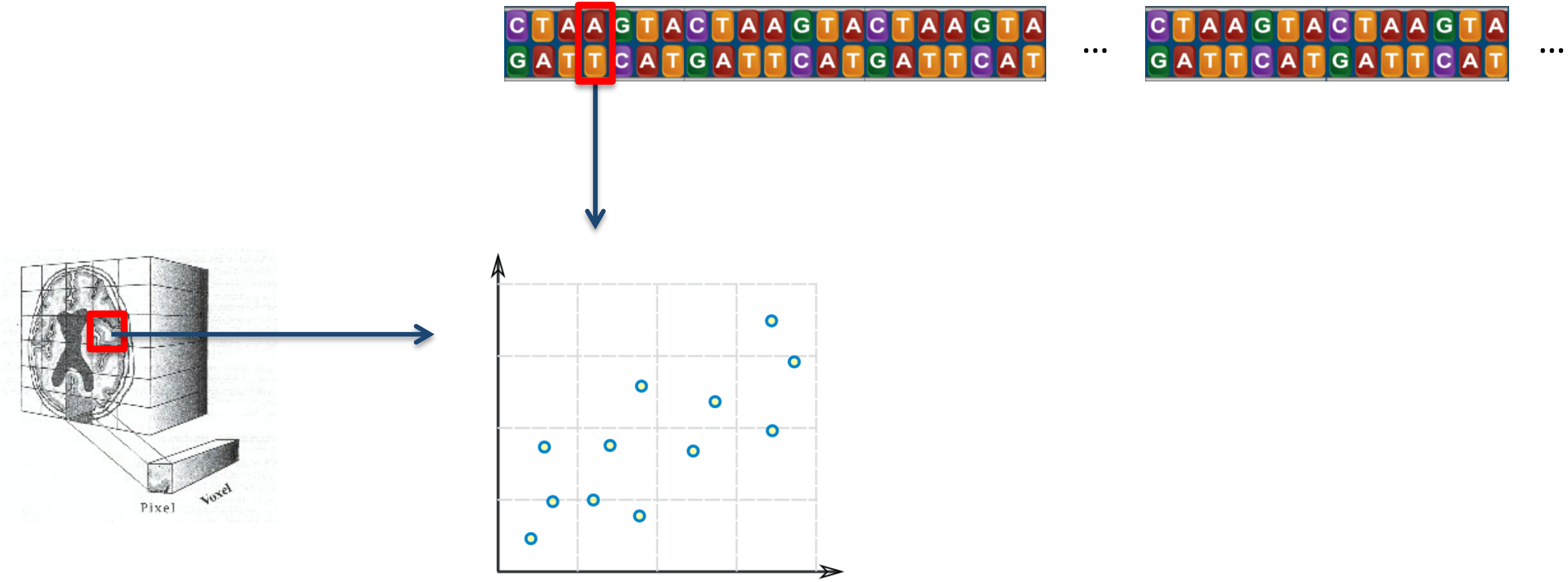

CTAAGTACTAAGTACTAAGTA
GATTCATGATTCATGATTCAT

…

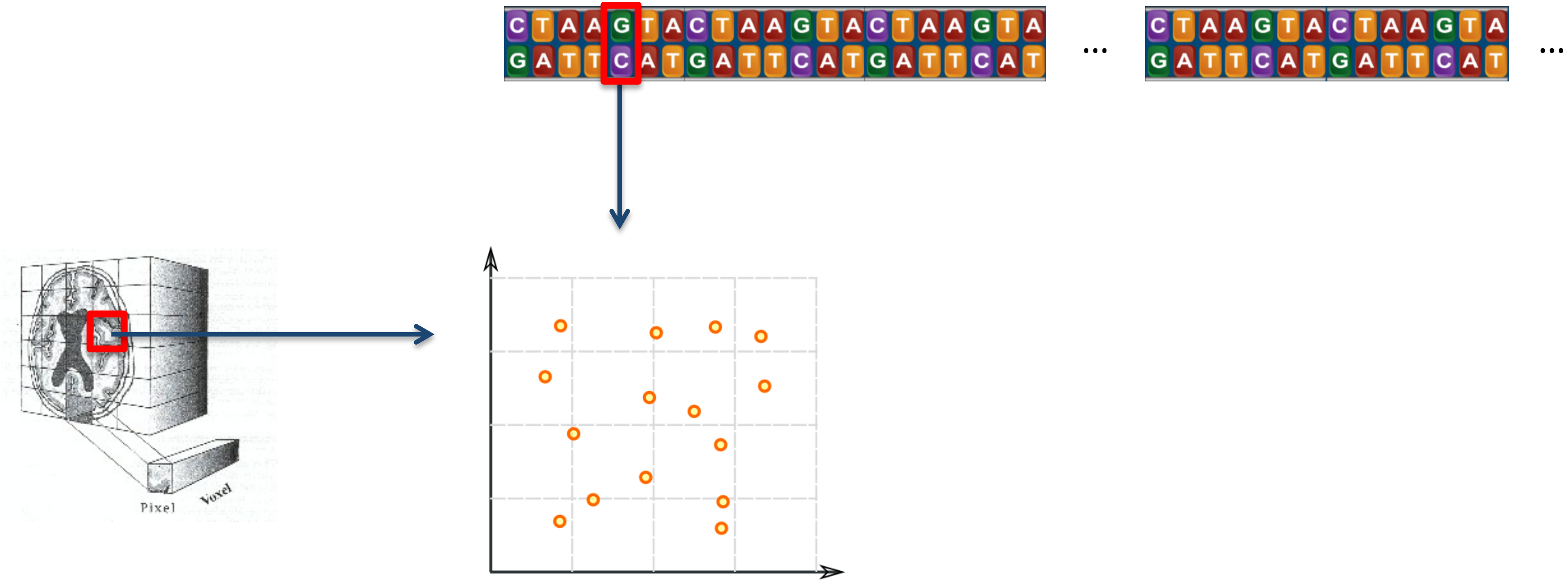

CTAAGTACTAAGTA
GATTCATGATTCAT

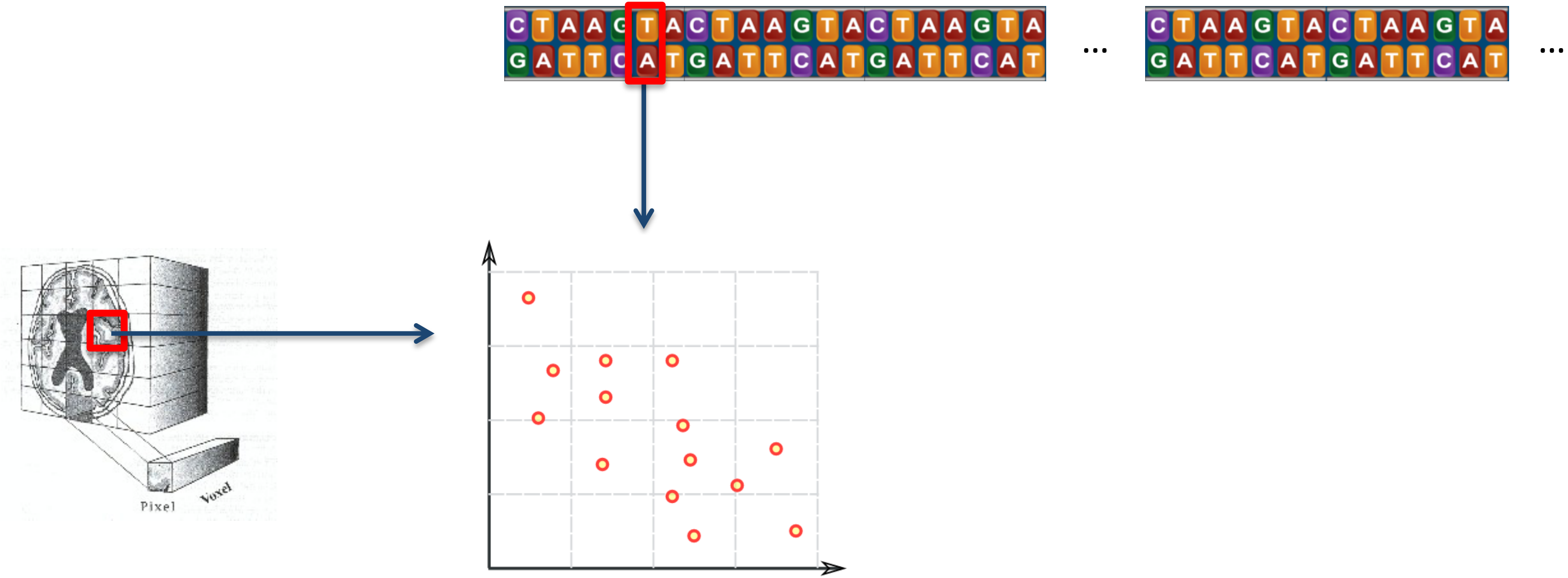
…

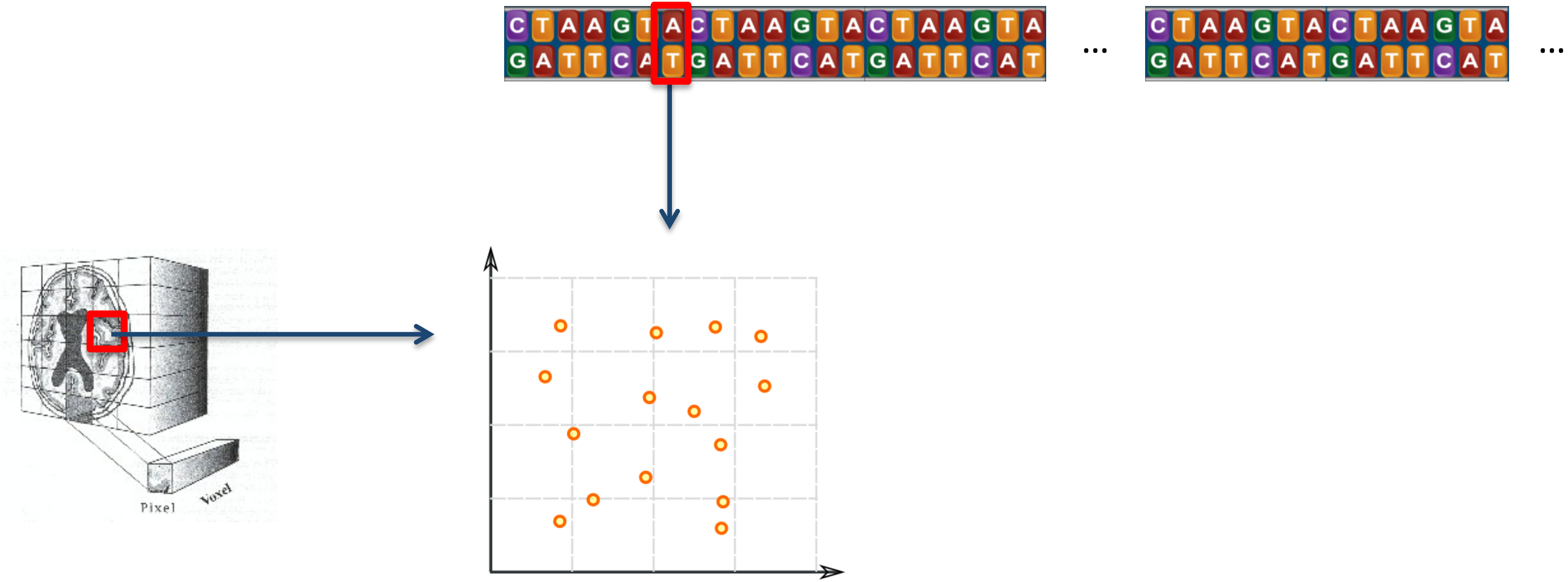

Pixel
Voxel



CTAAGTACTAAGTACTAAGTA
GATTCATGATTCATGATTCAT
…
CTAAGTACTAAGTA
GATTCATGATTCAT
…
Pixel
Voxel

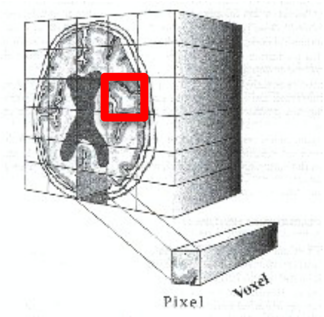

CTAAGTACTAAGTACTAAGTA
GATTCATGATTCATGATTCAT
…
CTAAGTACTAAGTA
GATTCATGATTCAT
…
Pixel
Voxel


CTAAGTACTAAGTACTAAGTA
GATTCATGATTCATGATTCAT
...
CTAAGTACTAAGTA
GATTCATGATTCAT
...
Pixel
Voxel

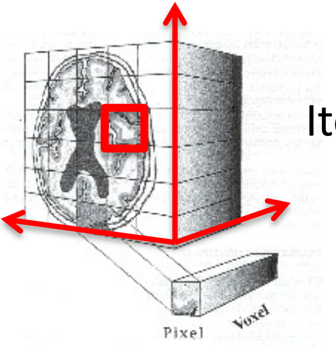


CTAAGTACTAAGTACTAAGTA
GATTCATGATTCATGATTCAT
…
CTAAGTACTAAGTA
GATTCATGATTCAT
…
Pixel
Voxel

Iterate for > 1’000’000 variants

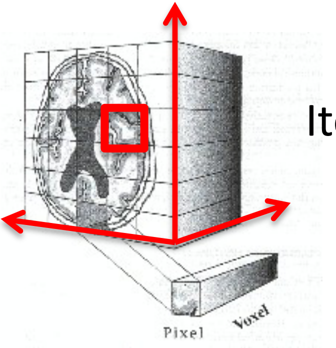


Iterate for > 1’000’000 variants



Iterate for > 1’000’000
image locations

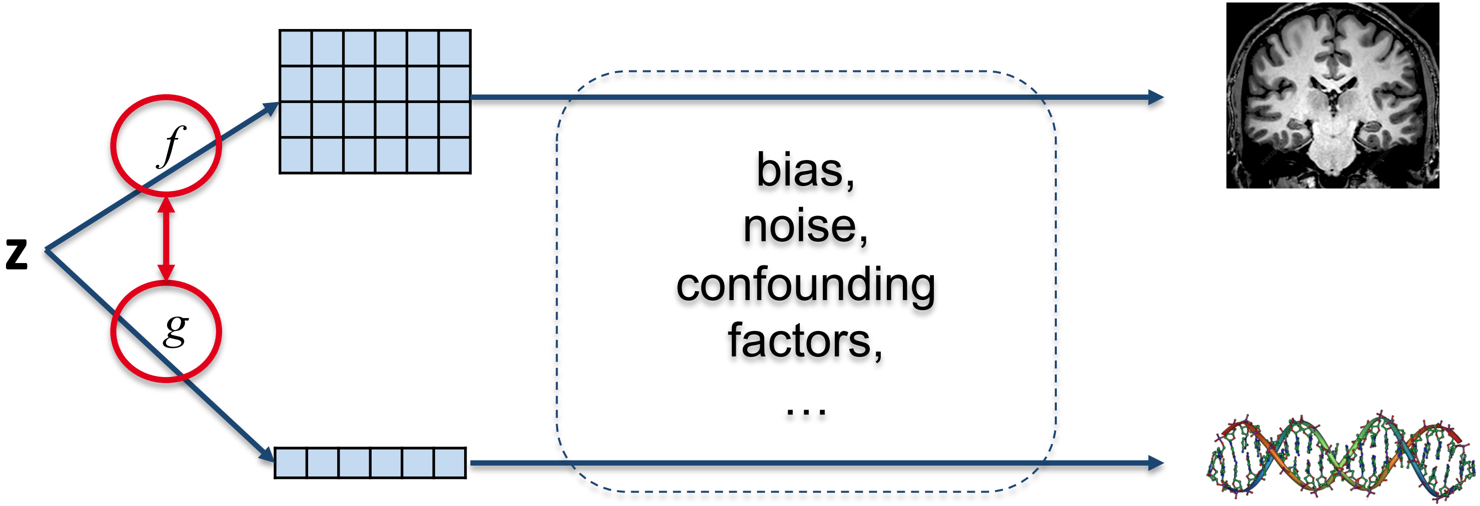
Iterate for > 1’000’000 variants



Iterate for > 1’000’000 image locations

- Hard interpretability
- False positive discoveries
- No interaction across brain and genetic areas

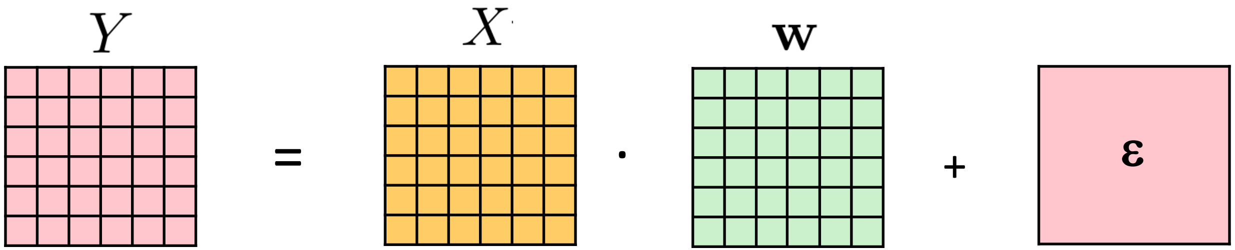
A unified statistical formulation via latent generative models



The inference problem

parameters ⟵ data

The building-block: linear model

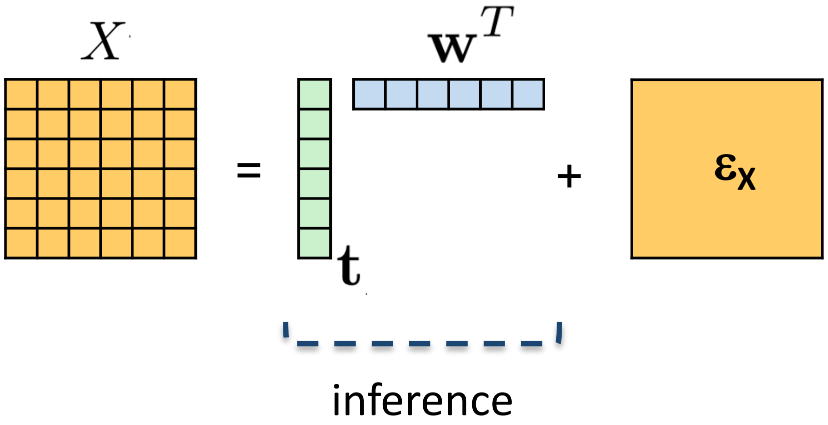


$$\nabla_{\mathbf{w}}(\|Y - X\mathbf{w}\|^2) = 0 \quad \Rightarrow \quad \mathbf{w} = (X^T X)^{-1} X^T Y$$

Classical formulation of latent variable models

Principal component analysis

$$\mathbf{t} \xrightarrow{\mathbf{w}^T} X$$



Classical formulation of latent variable models

Principal component analysis

A **variance** maximisation problem:

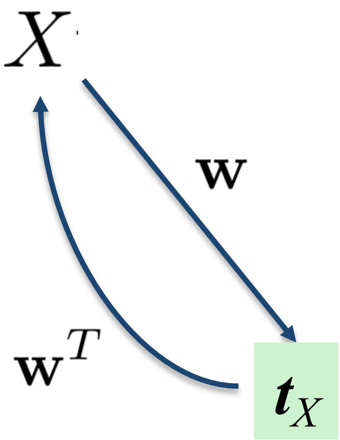
$$
\begin{tikzcd}
X \arrow[dr, "\mathbf{w}"] & \\
& \mathbf{t} \arrow[ul, bend left, "\mathbf{w}^T"]
\end{tikzcd}
$$

$$
\begin{aligned}
\mathbf{w} &= \operatorname{argmax}_{\|\mathbf{w}\|=1} (X\mathbf{w})^T (X\mathbf{w}) \\
&= \operatorname{argmax}_{\|\mathbf{w}\|=1} \mathbf{w}^T X^T X \mathbf{w} \\
&= \operatorname{argmax}_{\|\mathbf{w}\|=1} \mathbf{w}^T S_{XX} \mathbf{w}
\end{aligned}
$$

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w} S_{XX} \mathbf{w} - \lambda[\mathbf{w}_X^T \mathbf{w}_X - 1]$$

$$S_{XX}\mathbf{w} = \lambda \mathbf{w}$$

Non-linear iterative partial least squares - NIPALS
Wold 1975



1. Random initialization $\mathbf{w}$
2. Solve :

$$\mathrm{argmin}_{\mathbf{t}} ||X - \mathbf{t}\mathbf{w}^T||^2$$

$$\mathbf{t} = X\mathbf{w}(\mathbf{w}^T\mathbf{w})^{-1}$$

3. Normalize :

$$\mathbf{t} = \frac{\mathbf{t}}{||\mathbf{t}||}$$

4. Update :

$$\mathrm{argmin}_{\mathbf{w}} ||X - \mathbf{t}\mathbf{w}^T||^2$$

$$\mathbf{w} = X^T\mathbf{t}(\mathbf{t}^T\mathbf{t})^{-1}$$

5. Iterate 2-4 until convergence

Why it works:

$$4 \rightarrow const\,\mathbf{w} = X^T\mathbf{t}$$

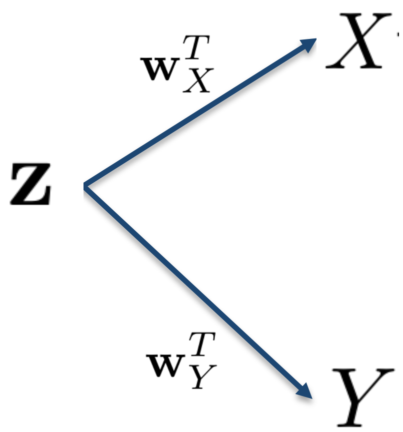
$$2 \rightarrow const\,\mathbf{t} = X\mathbf{w}$$

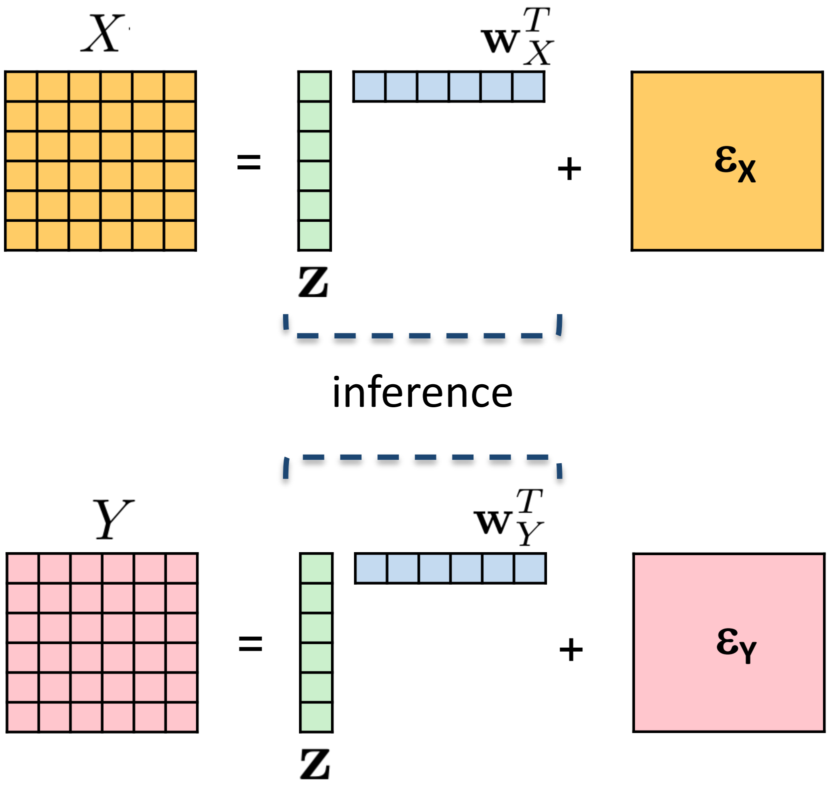
Then

$$const\,\mathbf{w} = S_{XX}\mathbf{w}$$

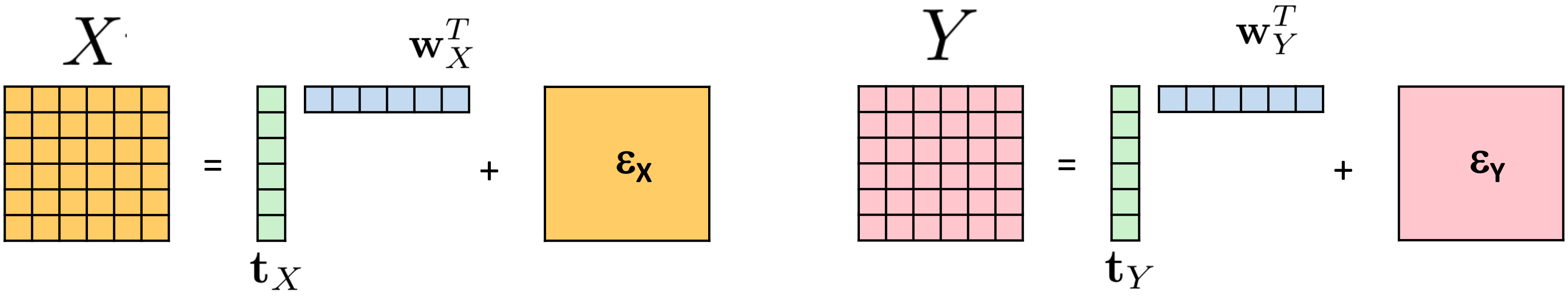
eigen-solution of the covariance matrix

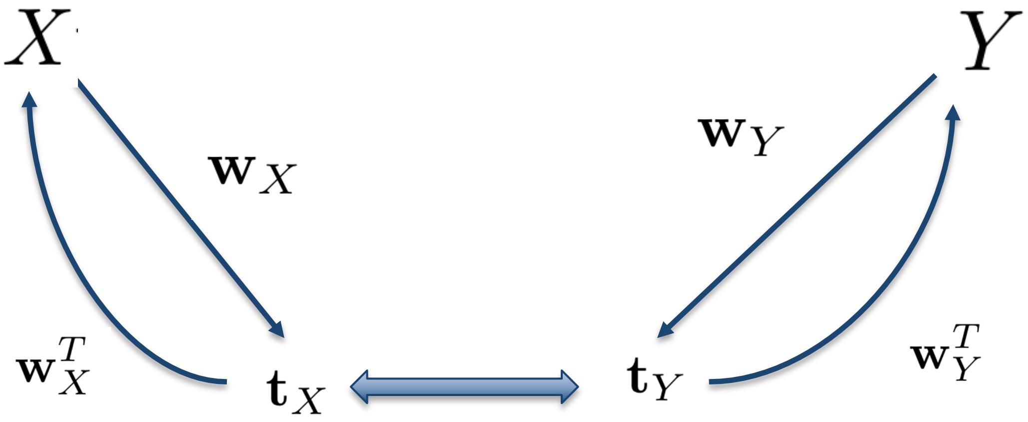
Multi-modal latent variable models





Multi-modal latent variable models



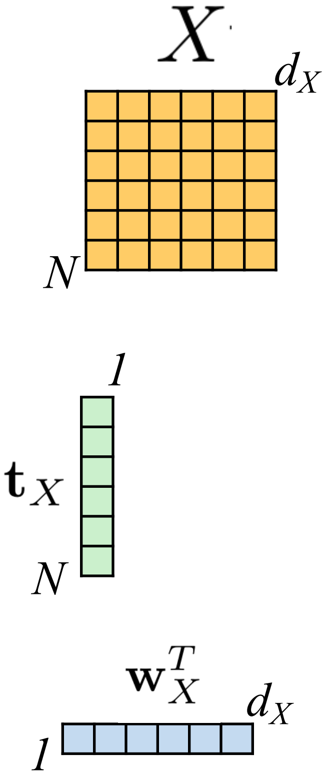


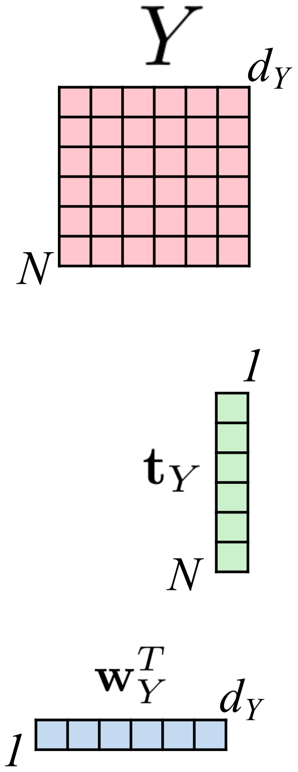
Goal:

Identifying
- Projections (loadings) $\mathbf{w}$
- Latent representation (scores) $\mathbf{t}$

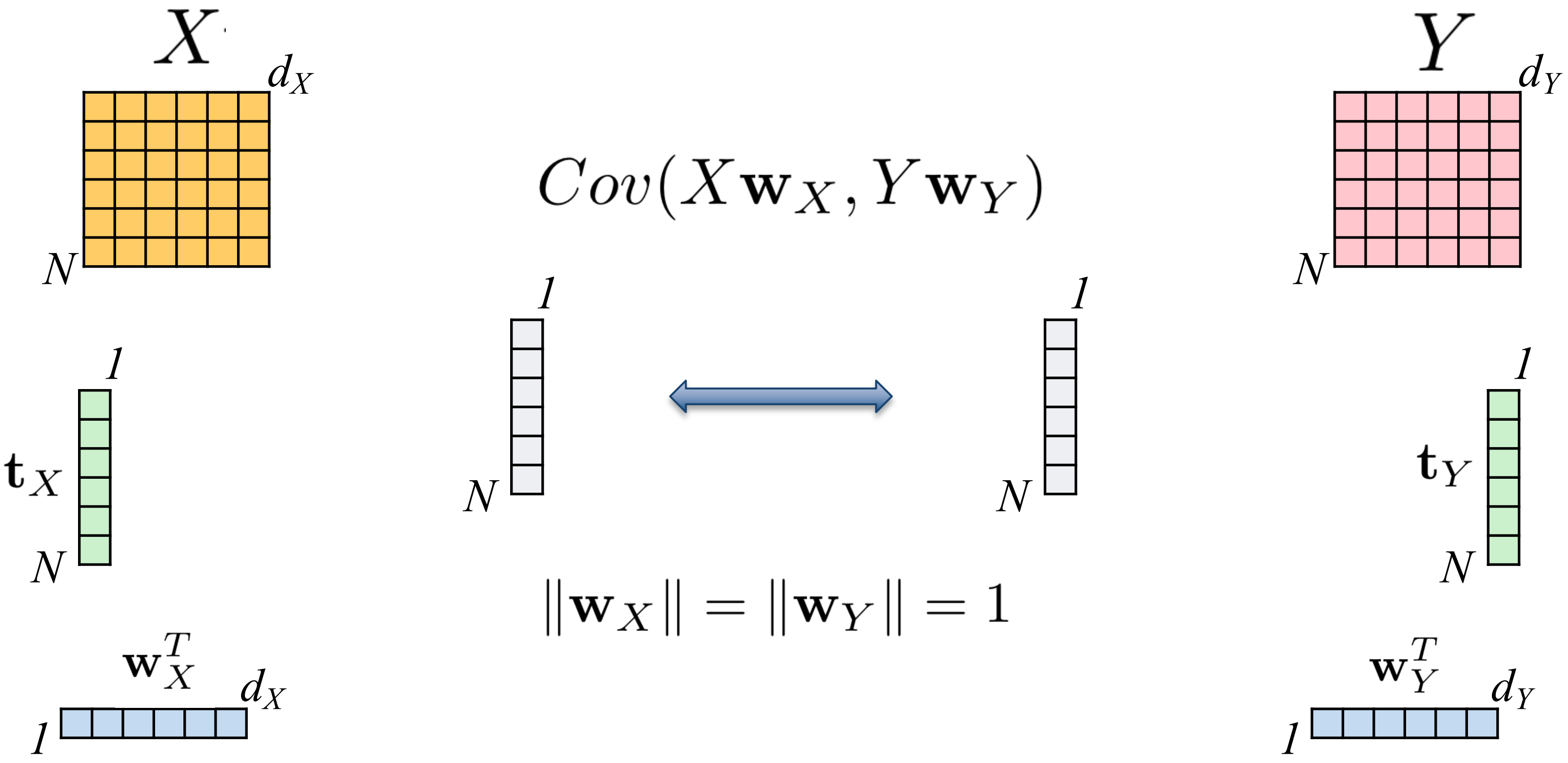
Geladi and Kowalski, Analytica Chimica Acta, 1985

Multi-modal latent variable models





Multi-modal latent variable models



Partial Least Squares

A **covariance** maximisation problem:

$$\operatorname{argmax}_{\mathbf{w}_X, \mathbf{w}_Y} Cov(X\mathbf{w}_X, Y\mathbf{w}_Y)$$

$$Cov(X\mathbf{w}_X, Y\mathbf{w}_Y) = \frac{\mathbf{w}_X^T S_{XY} \mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T \mathbf{w}_X}\sqrt{\mathbf{w}_Y^T \mathbf{w}_Y}}$$

Partial Least Squares

$$\mathcal{L}(\mathbf{w}_X, \mathbf{w}_Y, \lambda_X, \lambda_Y) = \mathbf{w}_X^T S_{XY} \mathbf{w}_Y - \lambda_X[\mathbf{w}_X^T \mathbf{w}_X - 1] - \lambda_Y[\mathbf{w}_Y^T \mathbf{w}_Y - 1]$$

$$\begin{cases} S_{XY}\mathbf{w}_Y = \lambda_X \mathbf{w}_X \\ S_{YX}\mathbf{w}_X = \lambda_Y \mathbf{w}_Y \end{cases} \Rightarrow \quad \begin{aligned} \lambda_X \mathbf{w}_X^T \mathbf{w}_X &= \mathbf{w}_X^T S_{XY} \mathbf{w}_Y \\ &= \mathbf{w}_Y^T S_{YX} \mathbf{w}_X \\ &= \lambda_Y \mathbf{w}_Y^T \mathbf{w}_Y \end{aligned}$$

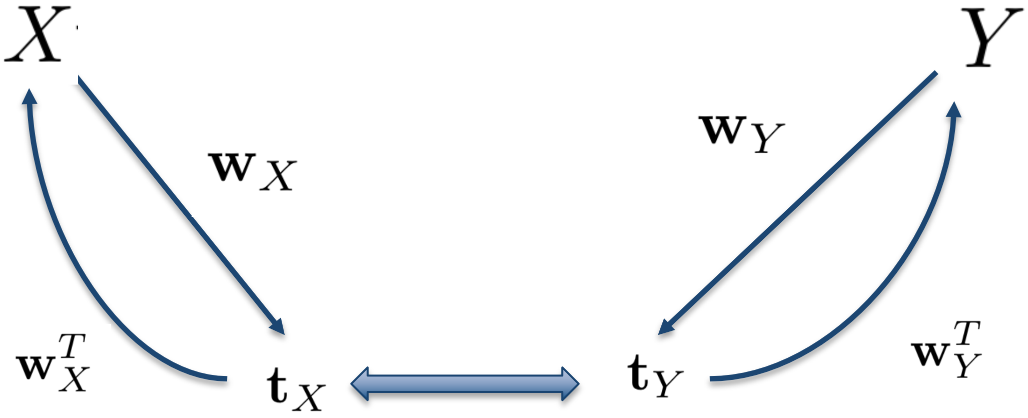
$$\lambda_X = \lambda_Y = \lambda$$

$$\begin{bmatrix} 0 & S_{XY} \\ S_{YX} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w}_X \\ \mathbf{w}_Y \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{w}_X \\ \mathbf{w}_Y \end{bmatrix}$$

The PLS problem is solved via singular value decomposition (SVD) of the covariance matrix

Non-linear iterative partial least squares - NIPALS

scikit-learn/sklearn/cross_decomposition



1. Random initialization $\mathbf{t}_X$
2. Update $\mathbf{w}_Y$:

$$\text{argmin}_{\mathbf{w}_Y} ||Y - \mathbf{t}_X \mathbf{w}_Y^T||^2$$

$$\mathbf{w}_Y = Y^t \mathbf{t}_X (\mathbf{t}_X^T \mathbf{t}_X)^{-1}$$

3. Normalize :

$$\mathbf{w}_Y = \frac{\mathbf{w}_Y}{||\mathbf{w}_Y||}$$

4. $\mathbf{t}_Y = Y \mathbf{w}_Y$
5. Update $\mathbf{w}_X$:

$$\text{argmin}_{\mathbf{w}_X} ||X - \mathbf{t}_Y \mathbf{w}_X^T||^2$$

$$\mathbf{w}_X = X^T \mathbf{t}_Y (\mathbf{t}_Y^T \mathbf{t}_Y)^{-1}$$

6. Normalize :

$$\mathbf{w}_X = \frac{\mathbf{w}_X}{||\mathbf{w}_X||}$$

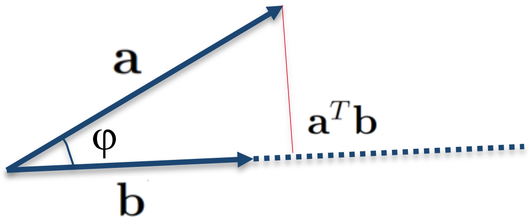
7. $\mathbf{t}_X = X \mathbf{w}_X$
8. Iterate 2-7 until convergence

Canonical Correlation Analysis

A **correlation** maximisation problem:

$$\underset{\mathbf{w}_X, \mathbf{w}_Y}{\operatorname{argmax}}\, \rho(X\mathbf{w}_X, Y\mathbf{w}_Y)$$

$$\rho(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T\mathbf{b}/(\sqrt{\mathbf{a}^T\mathbf{a}}\sqrt{\mathbf{b}^T\mathbf{b}})$$



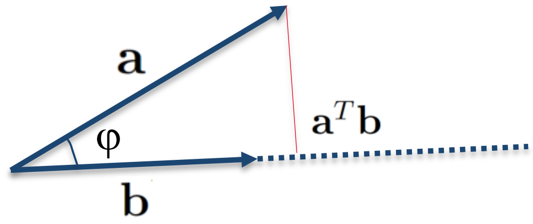
$$\rho(X\mathbf{w}_X, Y\mathbf{w}_Y) = \frac{\mathbf{w}_X^T S_{XY} \mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T S_{XX} \mathbf{w}_X}\sqrt{\mathbf{w}_Y^T S_{YY} \mathbf{w}_Y}}$$

Canonical Correlation Analysis

A **correlation** maximisation problem:

$$\underset{\mathbf{w}_X, \mathbf{w}_Y}{\operatorname{argmax}} \, \rho(X\mathbf{w}_X, Y\mathbf{w}_Y)$$

$$\rho(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T\mathbf{b}/(\sqrt{\mathbf{a}^T\mathbf{a}}\sqrt{\mathbf{b}^T\mathbf{b}})$$



Cross-covariance

$$\rho(X\mathbf{w}_X, Y\mathbf{w}_Y) = \frac{\mathbf{w}_X^T S_{XY} \mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T S_{XX} \mathbf{w}_X}\sqrt{\mathbf{w}_Y^T S_{YY} \mathbf{w}_Y}}$$

Within-modality covariance

Canonical Correlation Analysis

$$\mathcal{L}(\mathbf{w}_X, \mathbf{w}_Y, \lambda_X, \lambda_Y) = \mathbf{w}_X^T S_{XY} \mathbf{w}_Y - \lambda_X [\mathbf{w}_X^T S_{XX} \mathbf{w}_X - 1] - \lambda_Y [\mathbf{w}_Y^T S_{YY} \mathbf{w}_Y - 1]$$

Canonical Correlation Analysis

$$\mathcal{L}(\mathbf{w}_X, \mathbf{w}_Y, \lambda_X, \lambda_Y) = \mathbf{w}_X^T S_{XY} \mathbf{w}_Y - \lambda_X [\mathbf{w}_X^T S_{XX} \mathbf{w}_X - 1] - \lambda_Y [\mathbf{w}_Y^T S_{YY} \mathbf{w}_Y - 1]$$

$$\begin{cases} S_{XY}\mathbf{w}_Y = \lambda_X S_{XX} \mathbf{w}_X \\ S_{YX}\mathbf{w}_X = \lambda_Y S_{YY} \mathbf{w}_Y \end{cases}$$

Canonical Correlation Analysis

$$\mathcal{L}(\mathbf{w}_X, \mathbf{w}_Y, \lambda_X, \lambda_Y) = \mathbf{w}_X^T S_{XY} \mathbf{w}_Y - \lambda_X [\mathbf{w}_X^T S_{XX} \mathbf{w}_X - 1] - \lambda_Y [\mathbf{w}_Y^T S_{YY} \mathbf{w}_Y - 1]$$

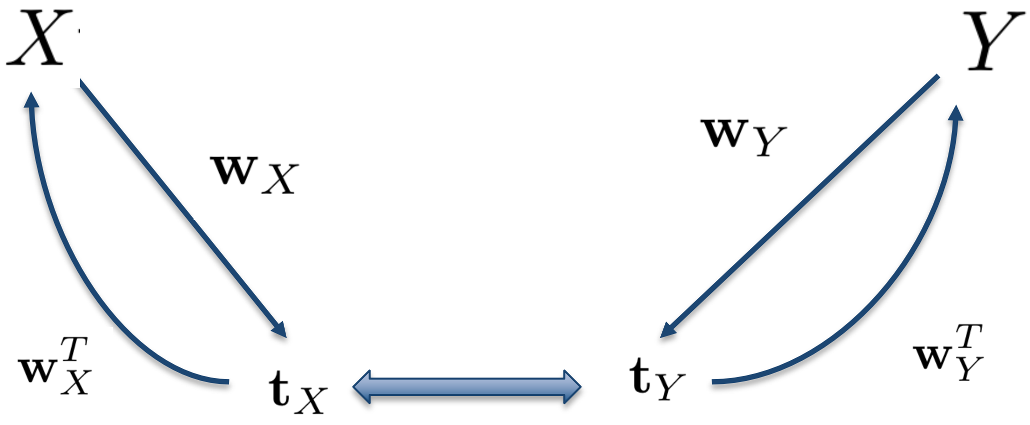
$$\begin{cases} S_{XY}\mathbf{w}_Y = \lambda_X S_{XX} \mathbf{w}_X \\ S_{YX}\mathbf{w}_X = \lambda_Y S_{YY} \mathbf{w}_Y \end{cases}$$

$$\begin{bmatrix} 0 & S_{XY} \\ S_{YX} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w}_X \\ \mathbf{w}_Y \end{bmatrix} = \lambda \begin{bmatrix} S_{XX} & 0 \\ 0 & S_{YY} \end{bmatrix} \begin{bmatrix} \mathbf{w}_X \\ \mathbf{w}_Y \end{bmatrix}$$

CCA is solved as a generalized eigenvalue problem

Non-linear iterative partial least squares - NIPALS

scikit-learn/sklearn/cross_decomposition



1. Random initialization $\mathbf{t}_X$
2. Update $\mathbf{w}_Y$:

$$\text{argmin}_{\mathbf{w}_Y} ||Y\mathbf{w}_Y - \mathbf{t}_X||^2$$

$$\mathbf{w}_Y = (Y^T Y)^{-1} Y^T \mathbf{t}_X$$

3. Normalize :

$$\mathbf{w}_Y = \frac{\mathbf{w}_Y}{||\mathbf{w}_Y||}$$

4. $\mathbf{t}_Y = Y\mathbf{w}_Y$
5. Update $\mathbf{w}_X$:

$$\text{argmin}_{\mathbf{w}_X} ||X\mathbf{w}_X - \mathbf{t}_Y||^2$$

$$\mathbf{w}_X = (X^T X)^{-1} X^T \mathbf{t}_Y$$

6. Normalize :

$$\mathbf{w}_X = \frac{\mathbf{w}_X}{||\mathbf{w}_X||}$$

7. $\mathbf{t}_X = X\mathbf{w}_X$
8. Iterate 2-7 until convergence

Non-linear iterative partial least squares - NIPALS
Deflation

$$\mathbf{X}^{(i+1)} = \mathbf{X}^{(i)} - \boldsymbol{t}^{(i)} \frac{\boldsymbol{t}^{(i)T} \mathbf{X}^{(i)}}{\boldsymbol{t}^{(i)T} \boldsymbol{t}^{(i)}},$$

$$\mathbf{Y}^{(i+1)} = \mathbf{Y}^{(i)} - \boldsymbol{u}^{(i)} \frac{\boldsymbol{u}^{(i)T} \mathbf{Y}^{(i)}}{\boldsymbol{u}^{(i)T} \boldsymbol{u}^{(i)}}.$$

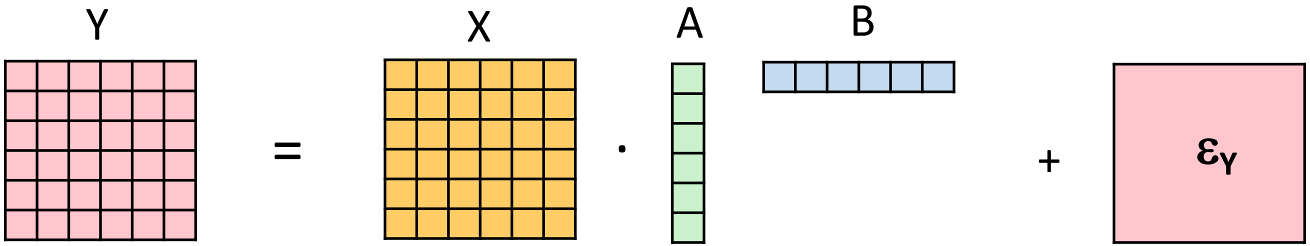
Iterate until

- residual component negligible epsilon
- Difference between consecutive residual components negligible

Reduced Rank Regression

$$\mathbf{Y} = \mathbf{X}\mathbf{C} + \epsilon.$$

$$\mathbf{C} = \mathbf{A}\mathbf{B}$$



$$f(\mathbf{A}, \mathbf{B}) = tr\{(\mathbf{Y} - \mathbf{X}\mathbf{A}\mathbf{B})\Gamma(\mathbf{Y} - \mathbf{X}\mathbf{A}\mathbf{B})^T\}$$

Reduced Rank Regression

Solution associated to the eigen-decomposition of the matrix

$$\mathbf{R} = \Gamma^{1/2}\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{XY}}\Gamma^{1/2}$$

Matrix encoding prior knowledge on Y

Reduced Rank Regression

Solution associated to the eigen-decomposition of the matrix

$$\mathbf{R} = \Gamma^{1/2}\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{XY}}\Gamma^{1/2}$$

Matrix encoding prior knowledge on Y

RRR solutions:

$$\mathbf{A} = \Gamma^{-1/2}\mathbf{U}, \qquad \mathbf{B} = \mathbf{U}^T\Gamma^{1/2}\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}$$

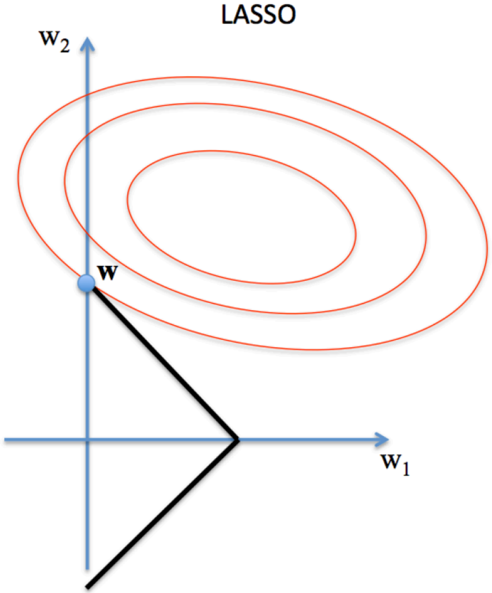
Special case:

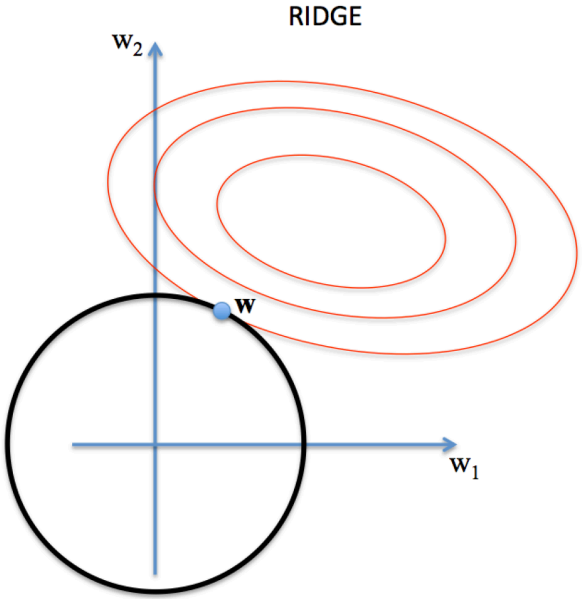
$$\Gamma = \mathbf{S}_{\mathbf{YY}} \Longrightarrow \text{CCA}$$

Sparsity in latent variable models

$$\underset{\mathbf{w}}{\operatorname{argmin}}\, f(\mathbf{w}) + \lambda ||\mathbf{w}||_1$$

$$\underset{\mathbf{w}}{\operatorname{argmin}}\, f(\mathbf{w}) + \lambda ||\mathbf{w}||_2^2$$





Algorithm Regularization of projections parameters $\mathbf{w}_x$ and $\mathbf{w}_y$ in NIPALS.

Given current estimates of $\mathbf{w}_x$ and $\mathbf{w}_y$.

While not converged do:

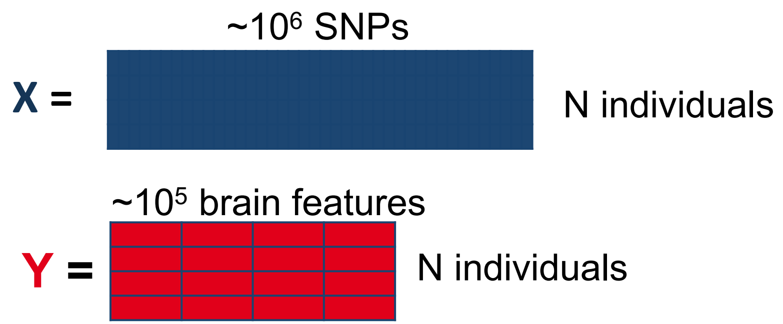
1. compute $\mathbf{t} = \mathbf{X}\mathbf{w}_x$,
2. compute $\mathbf{u} = \mathbf{Y}\mathbf{w}_y$,
3. compute $\overline{\mathbf{w}_x}$ by solving the Elastic-Net regression:
$$\overline{\mathbf{w}_x} = \arg\min_{\mathbf{v}} (\mathbf{t} - \mathbf{X}\mathbf{v})^2 + \lambda_{x2}\|\mathbf{v}\|_2^2 + \lambda_{x1}\|\mathbf{v}\|_1,$$
4. compute $\overline{\mathbf{w}_y}$ by solving the Elastic-Net regression:
$$\overline{\mathbf{w}_y} = \arg\min_{\mathbf{v}} (\mathbf{u} - \mathbf{Y}\mathbf{v})^2 + \lambda_{y2}\|\mathbf{v}\|_2^2 + \lambda_{y1}\|\mathbf{v}\|_1,$$
3. Normalize $\overline{\mathbf{w}_x}$ and $\overline{\mathbf{w}_x}$,
4. Set $\mathbf{w}_x = \overline{\mathbf{w}_x}$, $\mathbf{w}_y = \overline{\mathbf{w}_y}$.

PLS in practice

Application to imaging-genetics analysis
in Alzheimer's disease

Multivariate Association studies

Maximizing the joint relationship between genetic variants and brain features



Partial least squares (PLS)

$$\max_{p,q} \mathrm{Cov}(\, X \cdot p, Y \cdot q \,)$$

Multivariate Association studies

Maximizing the joint relationship between genetic variants and brain features

~10^6 SNPs

X = [matrix] N individuals

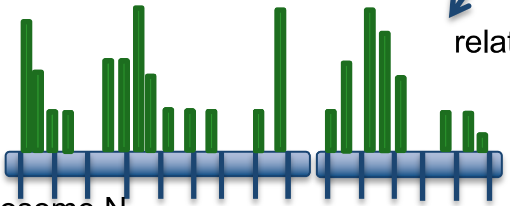
~10^5 brain features

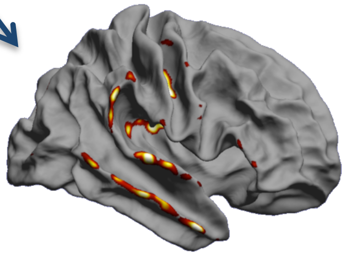
Y = [matrix] N individuals

Partial least squares (PLS)

$$\max_{p,q} \mathrm{Cov}(X \cdot p, Y \cdot q)$$

PLS weights = relative importance





Multivariate Association studies

Maximizing the joint relationship between genetic variants and brain features

~10^6 SNPs

X = N individuals

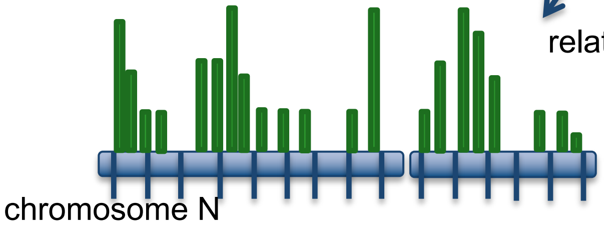
~10^5 brain features

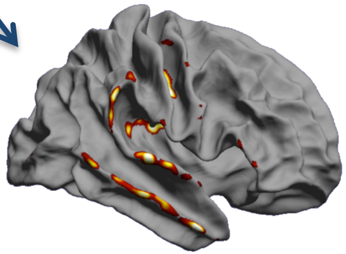
Y = N individuals

Partial least squares (PLS)

$$\max_{p,q} \mathrm{Cov}(\mathbf{X} \cdot p, \mathbf{Y} \cdot q)$$

PLS weights
=
relative importance





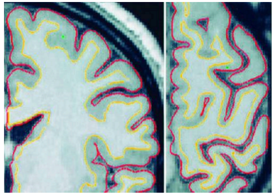
Pros. Overcomes issues of mass univariate analysis

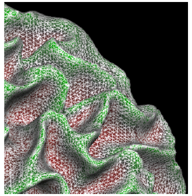
- Avoiding independent **multiple testing**
- Exploring **SNP-SNP interaction** (epistatic effects)

Study cohort

	Healthy	AD
N	401	238
Age (years)	74.45	74.72
Sex (% females)	49	45
MMSE	29.1	23.2
Apoe4 (% 0/1/2)	72/26/2	31/48/21

adni.loni.usc.edu





X = Phenotype features

- Freesurfer **brain cortical thickness** maps (327,684 mesh points)
- **Radial distance** of hippocampi and amygdalae (27,120 mesh points) [Gutman et al, NeuroImage 2013]

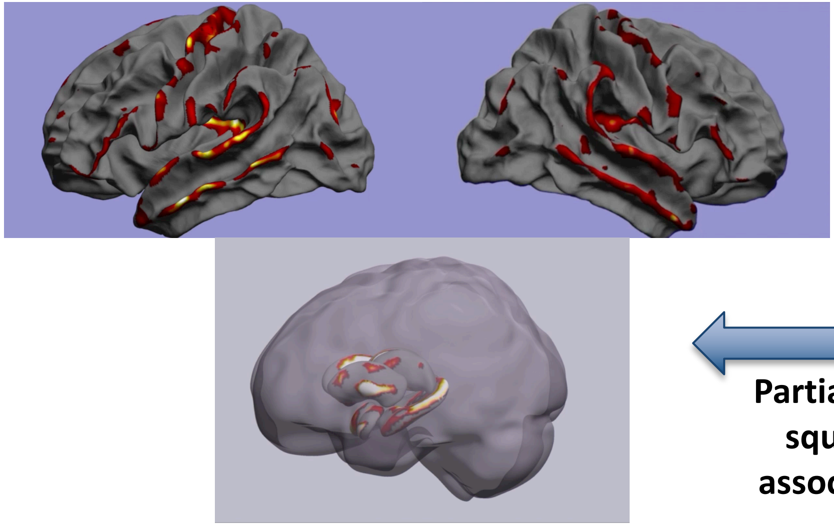
Y = Genotype features

- Individuals' minor allele counts for **1,167,126 SNPs** in chromosomes 1 to 22

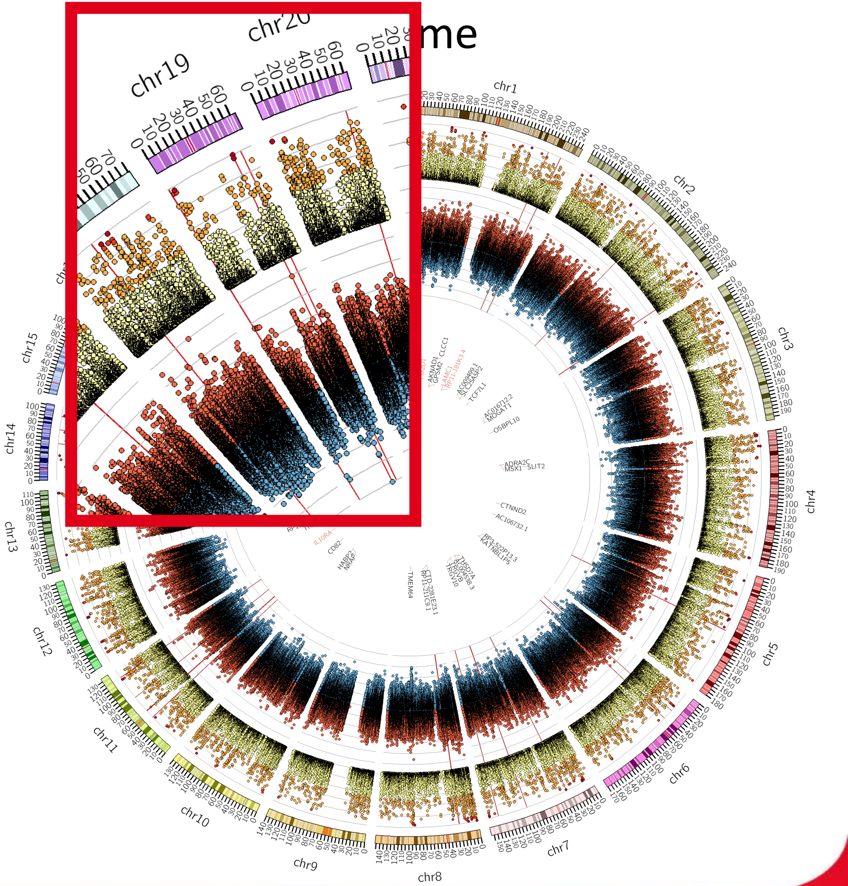
Standard quality control: MAF < 0.01, Genotype Call Rate <95% , Hardy-Weinberg Equilibrium < 1x10^{-6}.
Imputation to HapMap III reference panel, quality controlled (MAF > 0.01 and R-squared > 0.3)

Application to multivariate Imaging-genetics

Atrophy profile from brain imaging



Partial least squares association



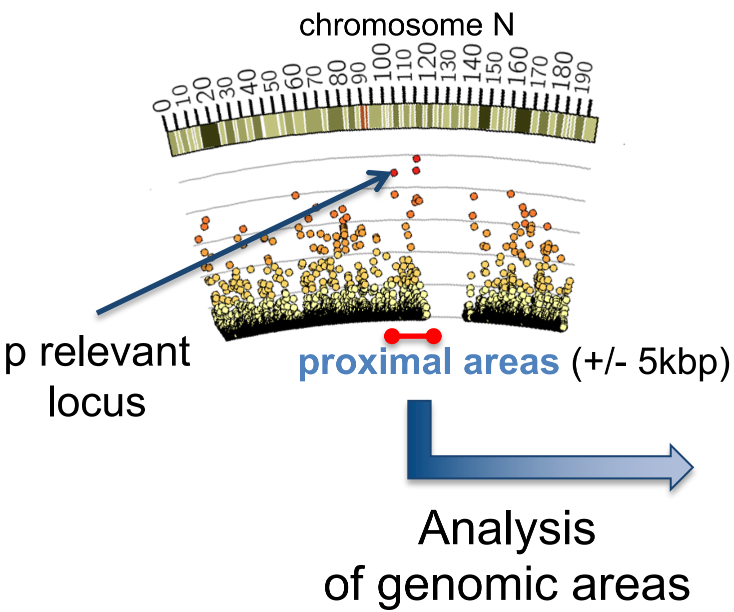
639 individuals

401 healthy

238 Alzheimer's

Investigating biological mechanisms through Meta-analysis

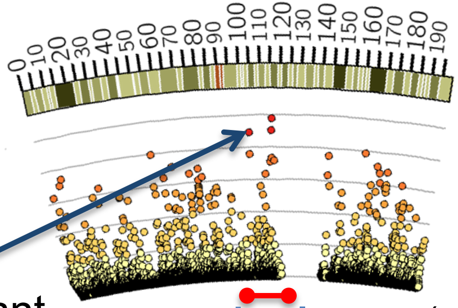
PLS statistical result



Investigating biological mechanisms through Meta-analysis

PLS statistical result

chromosome N

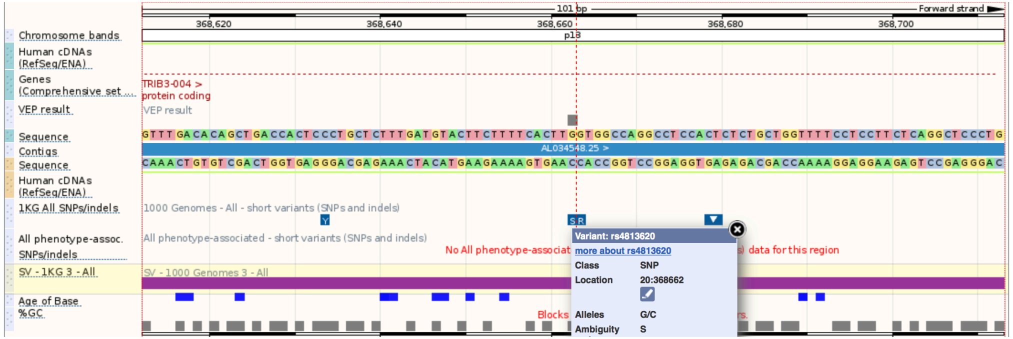


p relevant locus

proximal areas (+/- 5kbp)

Analysis
of genomic areas

Querying gene annotation databases



McLaren et al. The Ensembl Variant Effect Predictor. Genome Biology, 2016

Investigating biological mechanisms through Meta-analysis

Ve!P → 148 SNP-gene combinations

6 tested tissues

hippocampus, whole blood,
Adipose subcutaneous, artery tibia, nerve tibial, treated fibroblast

14 Significantly expressed genes

TM2D1 (amyloid-beta binding protein),
IL10RA (increase in hippo in mouse model),
TRIB3
(neuronal cell death, modulates PSEN1 stability, interacts with APP)

	Significance (p-value) training	Significance (p-value) testing
TM2D1	0.005	0.053
IL10RA	0.107	0.620
TRIB3	0.003	0.003
ZBTB7A	0.036	0.913
LYSMD4	0.000	0.206
CRYL1	0.621	0.118
FAM135B	0.000	0.559
IP6K3	0.000	0.465
ITGA1	0.099	0.731
KIN	0.001	0.206
LAMC1	0.002	0.062
LINC00941	0.000	0.690
RBPMS2	0.000	0.215
RP11-181K3.4	0.002	0.053

Latent variable models

1. *Multi-variate* modeling

2. **Novel scalable approaches to *multi-view* data**

Latent variable models via Variational Autoencoders

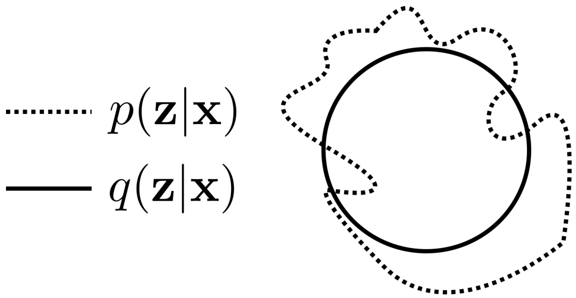
Kingma &Welling, 2014; Rezende et al. 2014

$$\mathbf{z} \longrightarrow \mathbf{x}$$

Posterior $p(\mathbf{z}|\mathbf{x})$ $\qquad$ $p(\mathbf{x}|\mathbf{z})$ Likelihood

$$p(\mathbf{z}|\mathbf{x}) = \int_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

Difficult to compute



Idea: find a “close enough” and simple approximation $q(\mathbf{z}|\mathbf{x})$

Latent variable models via Variational Autoencoders

Kingma &Welling, 2014; Rezende et al. 2014

$$\mathbf{z} \longrightarrow \mathbf{x}$$

$$D_{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z}|\mathbf{x})] \quad = \quad \mathbf{E}_{\mathbf{z}\sim q}\log[q(\mathbf{z}|x)] - \mathbf{E}_{\mathbf{z}\sim q}\log[p(\mathbf{z}|x)]$$

C. M. Bishop, Pattern Recognition and Machine Learning, Ch.10, Ed. 2006

Latent variable models via Variational Autoencoders

Kingma &Welling, 2014; Rezende et al. 2014

$$\mathbf{z} \longrightarrow \mathbf{x}$$

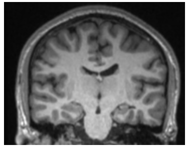
$$D_{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z}|\mathbf{x})] \;=\; \mathbf{E}_{\mathbf{z}\sim q}\log[q(\mathbf{z}|x)] - \mathbf{E}_{\mathbf{z}\sim q}\log[p(\mathbf{z}|x)]$$

Evidence lower bound (ELBO)

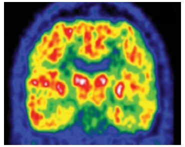
$$\mathcal{L} = \underbrace{\mathbf{E}_{\mathbf{z}\sim q}\log[p(\mathbf{x}|\mathbf{z})]}_{\text{reconstruction}} - \underbrace{D_{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})]}_{\text{regularization}}$$

C. M. Bishop, Pattern Recognition and Machine Learning, Ch.10, Ed. 2006

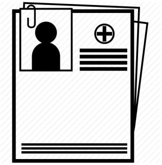
Generative representation of multimodal data



X_1



X_2

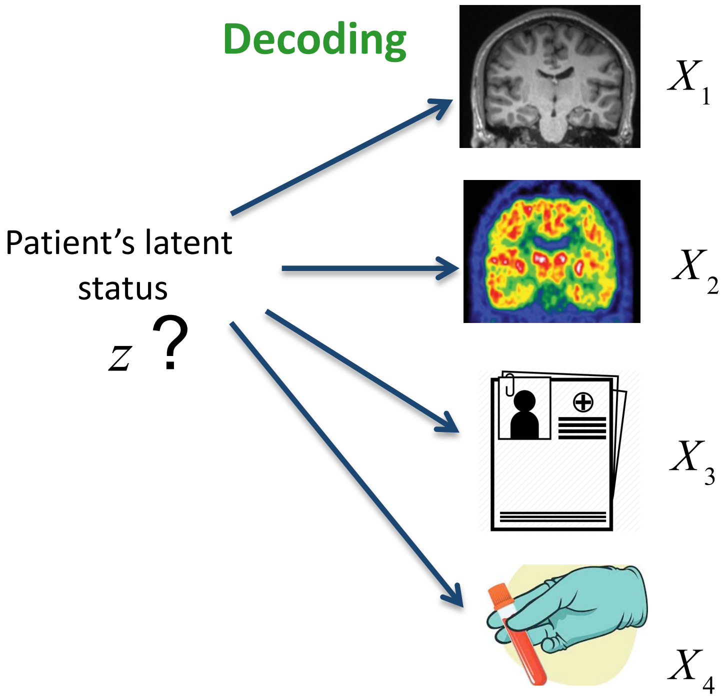


X_3

X_4

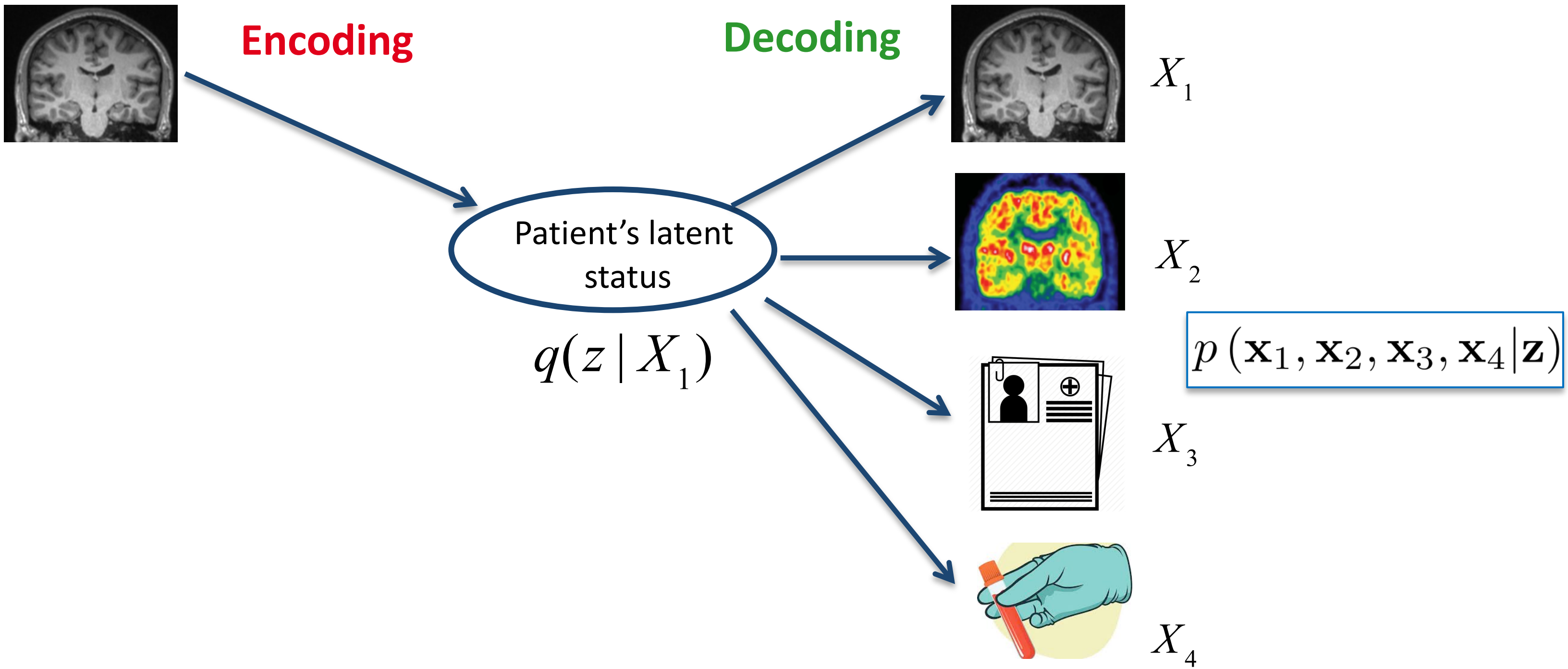
Generative representation of multimodal data

$$p\left(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 | \mathbf{z}\right)$$



Decoding: data reconstruction from the latent representation

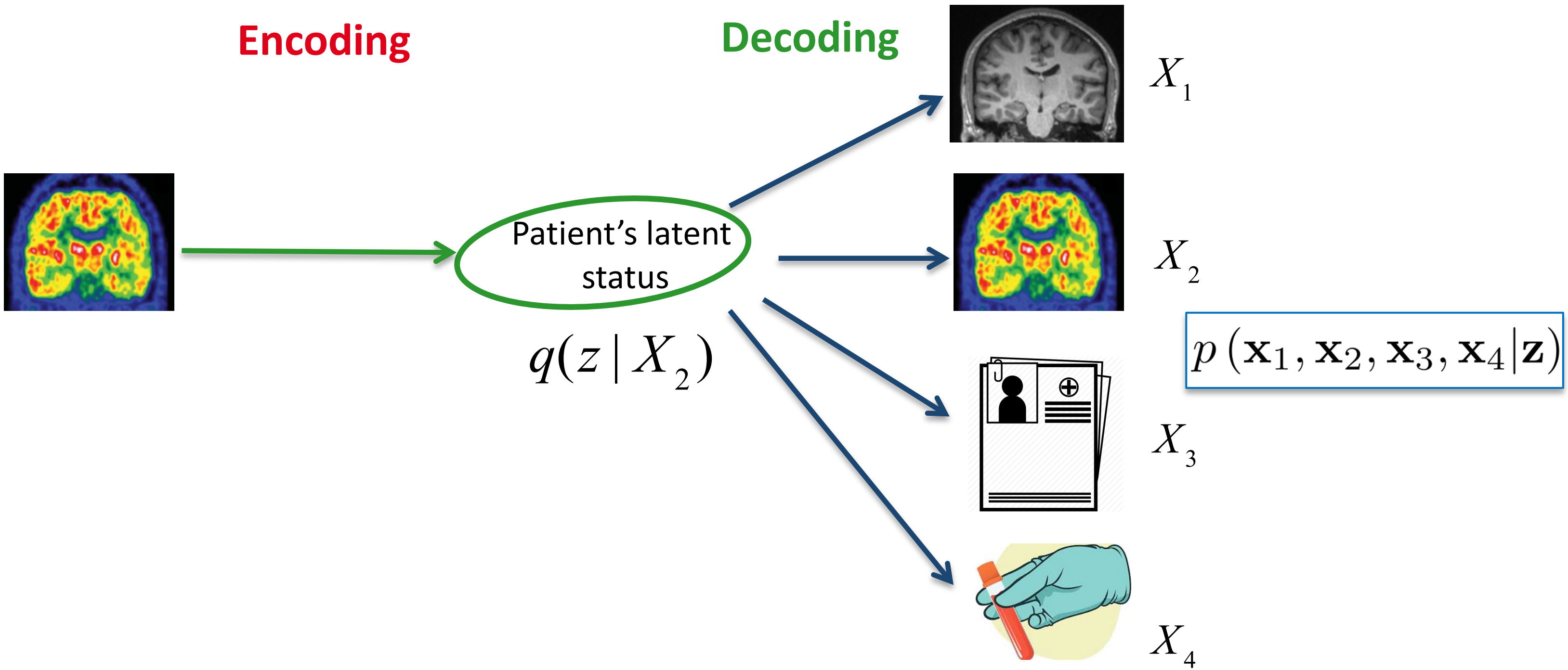
Generative representation of multimodal data



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

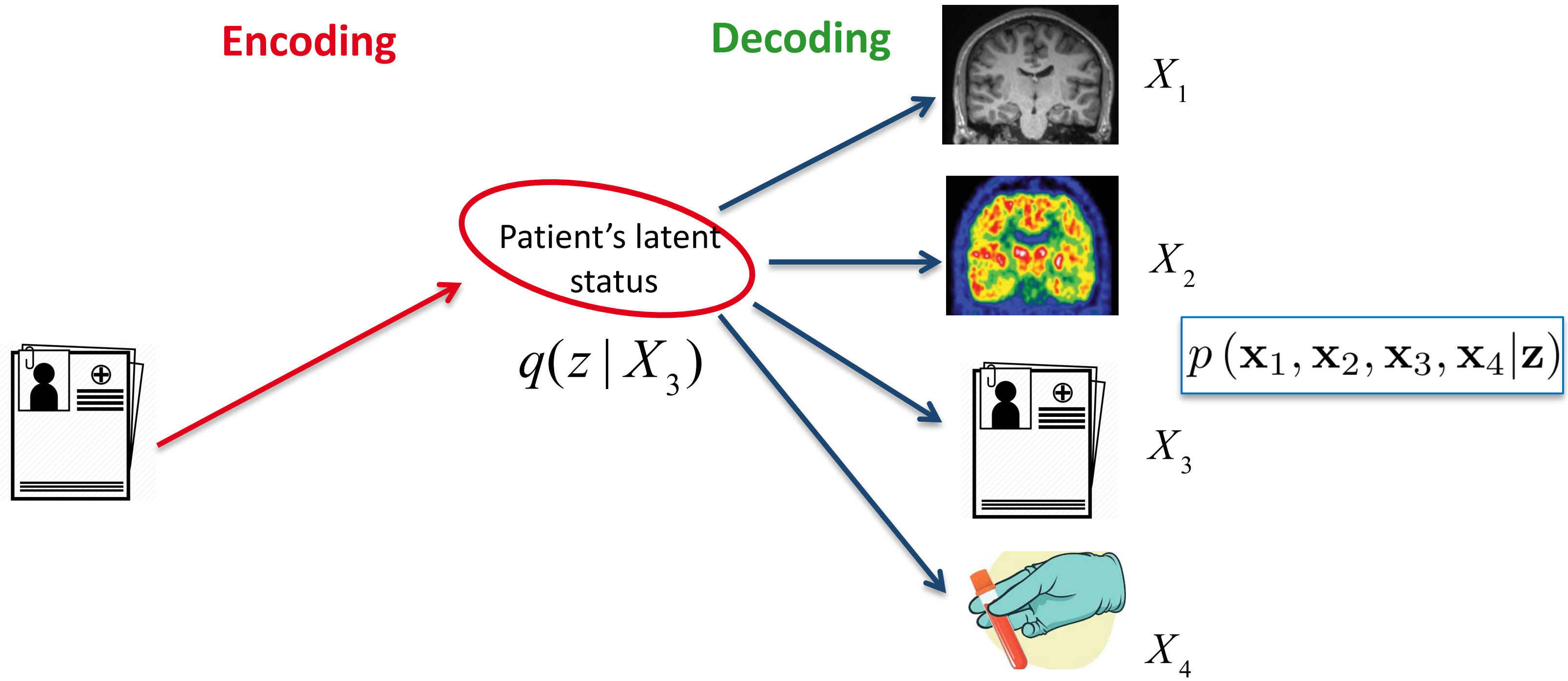
Generative representation of multimodal data



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

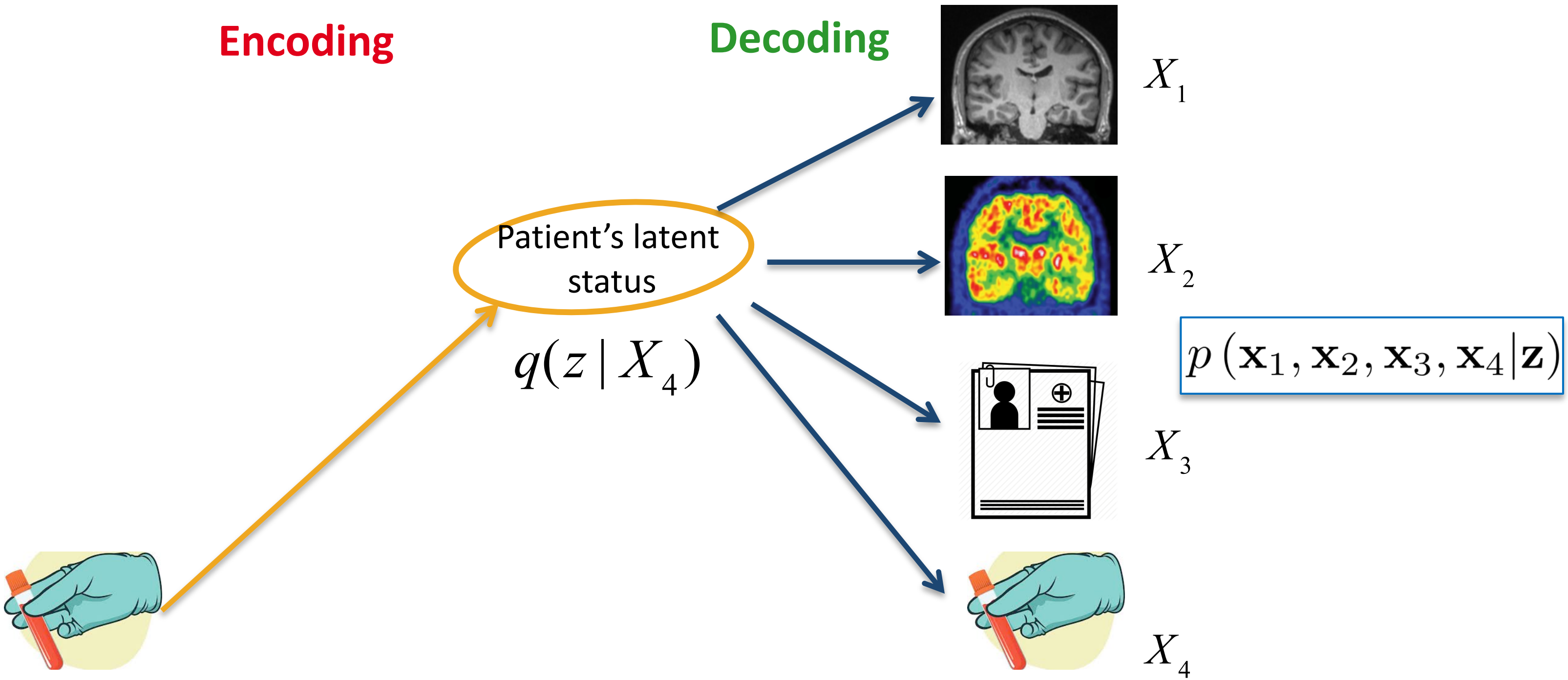
Generative representation of multimodal data



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

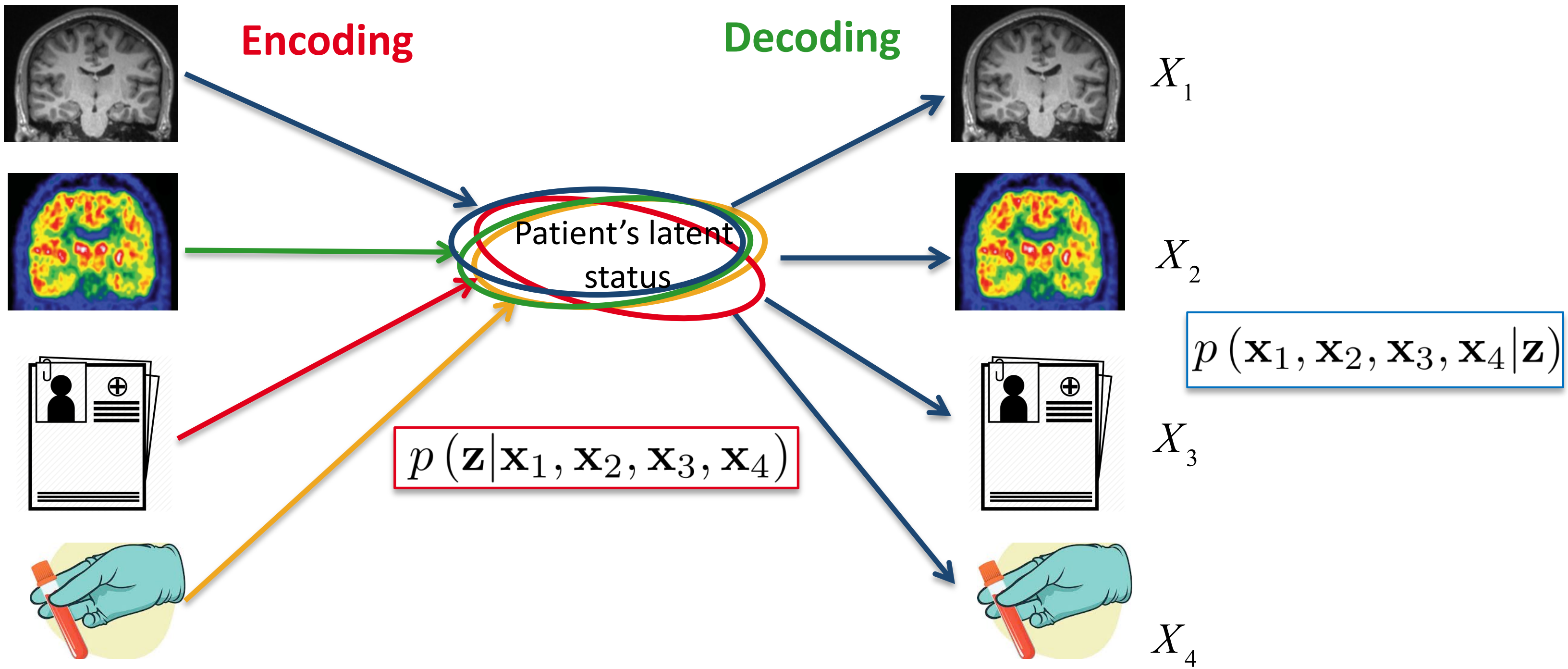
Generative representation of multimodal data



Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

Generative representation of multimodal data

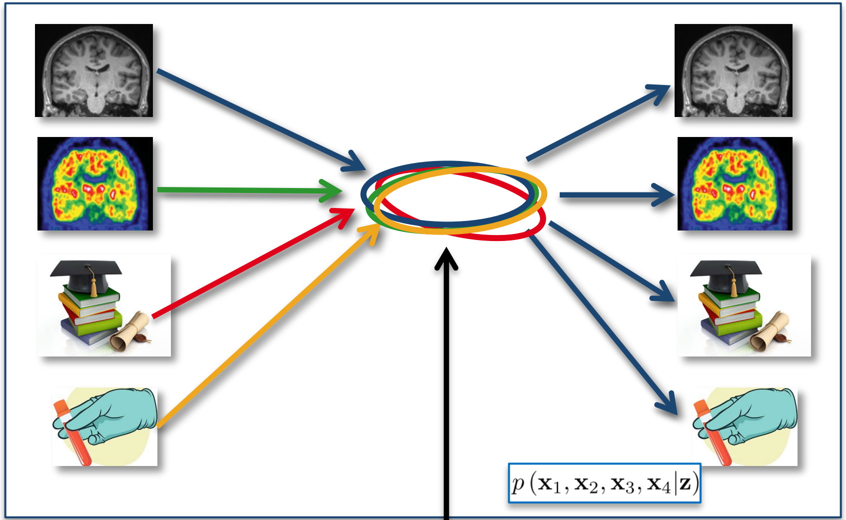


Decoding: data reconstruction from the latent representation

Encoding: latent representation from the data

Antelmi, Ayache, Robert and Lorenzi, ICML 2019

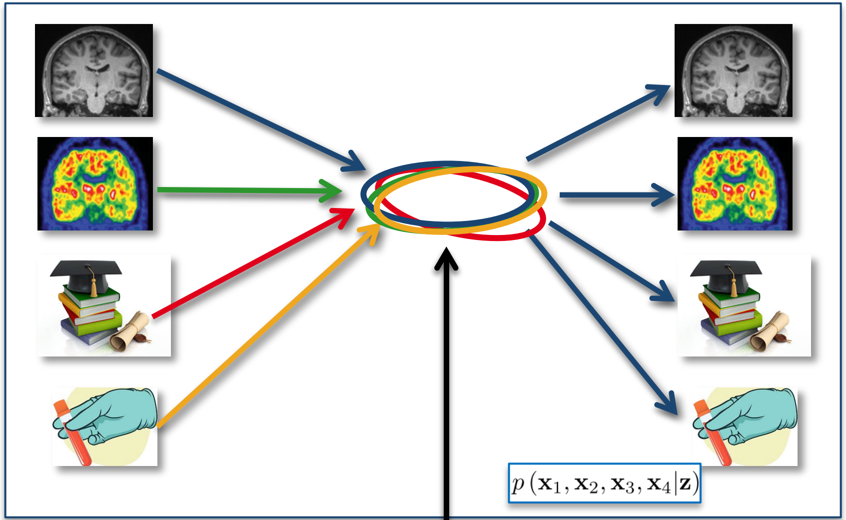
Generative representation of multimodal data



minimize

$$\frac{1}{C}\,\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}|\mathbf{x}_1,\ldots,\mathbf{x}_C\right)\right)$$

Generative representation of multimodal data



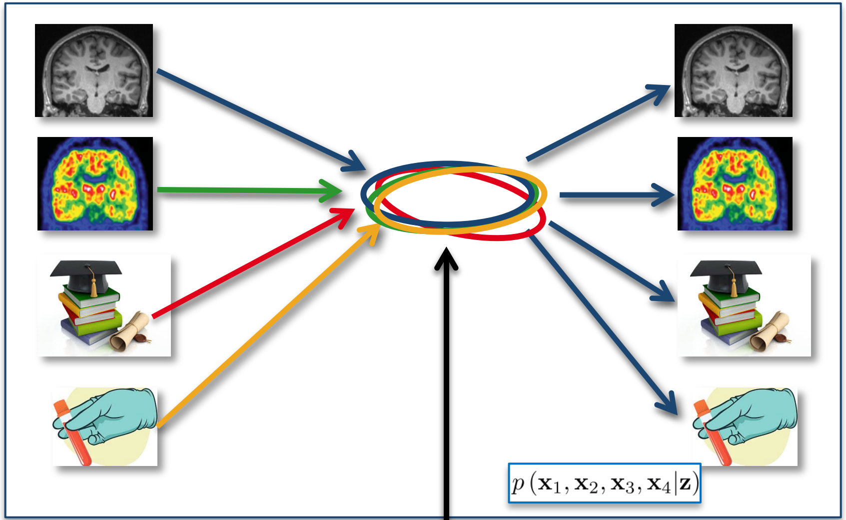
Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right] - \mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\right)$$

minimize

$$\tfrac{1}{C}\,\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}|\mathbf{x}_1,\ldots,\mathbf{x}_C\right)\right)$$

Generative representation of multimodal data



minimize

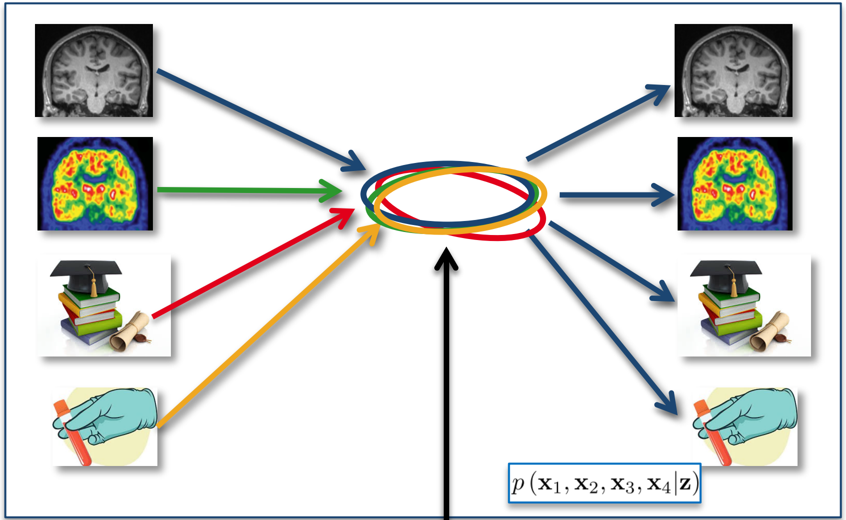
$$\frac{1}{C}\,\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}|\mathbf{x}_1,\ldots,\mathbf{x}_C\right)\right)$$

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right]-\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\right)$$

Encoding for given channel

Generative representation of multimodal data



minimize

$$\frac{1}{C}\, \mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right) || \, p\left(\mathbf{z}|\mathbf{x}_1, \ldots, \mathbf{x}_C\right)\right)$$

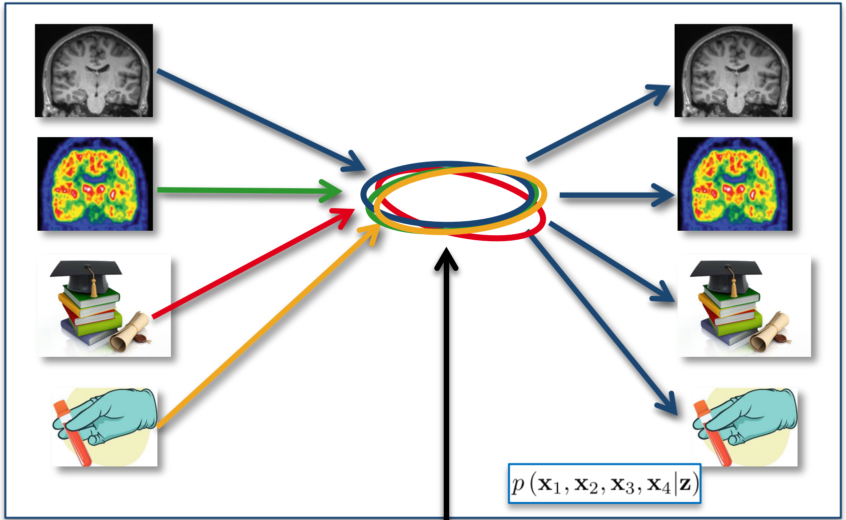
Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C} \ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right] - \mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right) || \, p\left(\mathbf{z}\right)\right)$$

Encoding for given channel

Reconstruction of all channels

Generative representation of multimodal data



minimize

$$\frac{1}{C}\,\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}|\mathbf{x}_1,\ldots,\mathbf{x}_C\right)\right)$$

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right] - \mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\right)$$

Encoding for given channel

Reconstruction of all channels

Regularization: sparsity inducing prior

[Kingma et al, NIPS, 2015; Molchanov et al, ICML 2017]

Classic Implementation (non sparse)

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right]-\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\right)$$

Reconstructions: $p\left(\mathbf{x}_i|\mathbf{z}\right)=\mathcal{N}\left(\mathbf{x}_i|\boldsymbol{\mu}_{\mathbf{z}};\boldsymbol{\Sigma}_{\mathbf{z}}\right)$

Encodings: $q\left(\mathbf{z}|\mathbf{x}_c\right)=\mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_{\mathbf{x}_c};\boldsymbol{\Sigma}_{\mathbf{x}_c}\right)$

Prior: $p\left(\mathbf{z}\right)=\mathcal{N}\left(\mathbf{0};\mathbf{I}\right)$

Sparse Implementation

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right]-\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\right)$$

Reconstructions: $p\left(\mathbf{x}_i|\mathbf{z}\right)=\mathrm{same}$

Encodings: $q\left(\mathbf{z}|\mathbf{x}_c\right)=\mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_{\mathbf{x}_c};\alpha\odot\boldsymbol{\mu}_{\mathbf{x}_c}^2\right)$

Prior: $p\left(\mathbf{z}\right)\propto {}^{1}/_{|\mathbf{z}|}$

Sparse Implementation: why it works?

Evidence Lower bound (ELBO)

$$\frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c)}\left[\sum_{i=1}^{C}\ln p\left(\mathbf{x}_i|\mathbf{z}\right)\right]-\mathcal{D}_{\mathrm{KL}}\left(q\left(\mathbf{z}|\mathbf{x}_c\right)||\,p\left(\mathbf{z}\right)\right)$$

Encodings: $q\left(\mathbf{z}|\mathbf{x}_c\right)=\mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_{\mathbf{x}_c};\alpha\odot\boldsymbol{\mu}^2_{\mathbf{x}_c}\right)$

Prior: $p\left(\mathbf{z}\right)\propto {}^1\!/_{|\mathbf{z}|}$

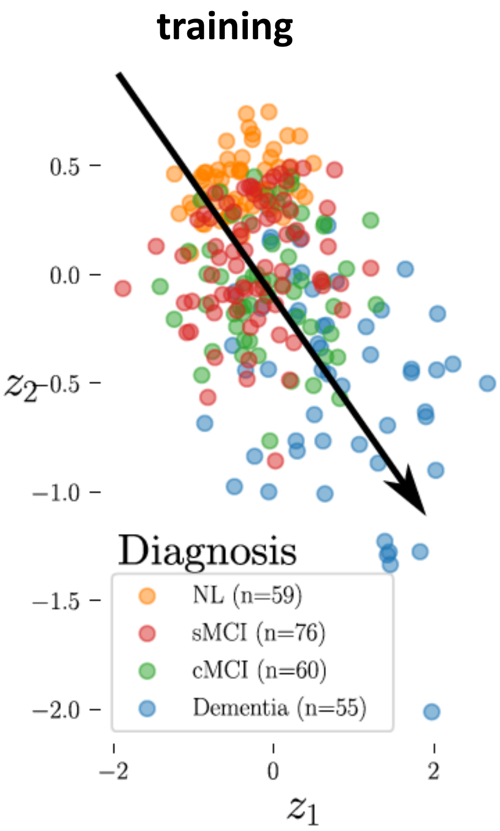
Prior and encodings act together such that (element-wise):

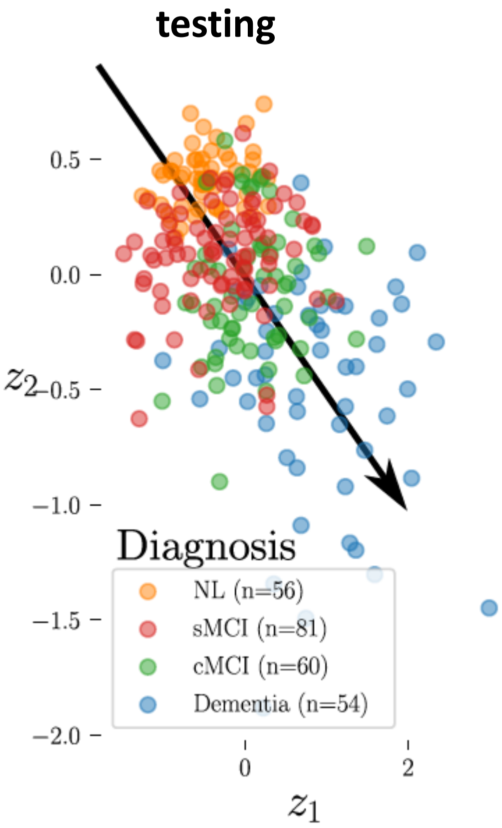
$$\lim_{\mu_i\to 0}\mathcal{N}\left(z_i|\mu_i;\alpha_i\cdot\mu_i^2\right)=\delta(0)$$

Relationship between α and the probability of pruning the *i*-th dimension:

$$\alpha_i=\frac{p_i}{1-p_i}$$

Prediction from latent space



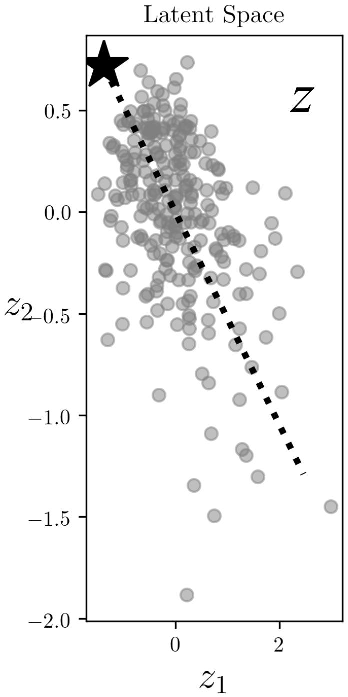


Joint modeling of

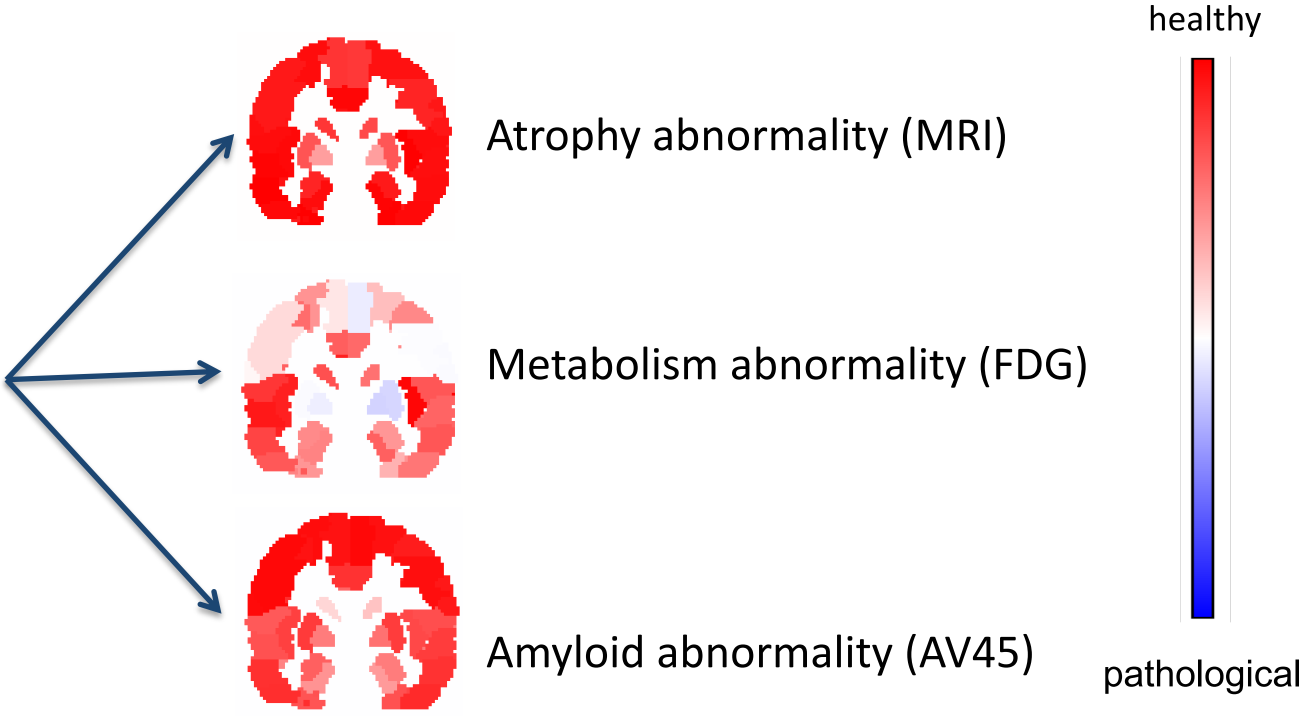
- Brain imaging:
  - Structural (T1 MRI)
  - Molecular (FDG-PET + Amy-PET)
- Socio-demographic factors
- Clinical scores

	Accuracy (SD)
Cognitively Healthy	**0.89** (**0.03**)
Stable Mild Cognitive Impairment (sMCI)	**0.75** (**0.02**)
MCI converting to Dementia (cMCI)	**0.70** (**0.05**)
Dementia	**0.94** (**0.05**)

Generation from latent space



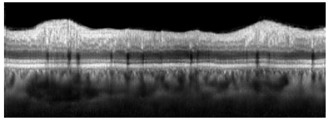
Patient's latent status
from cognitive scores only

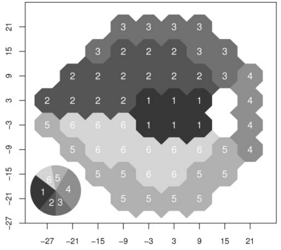


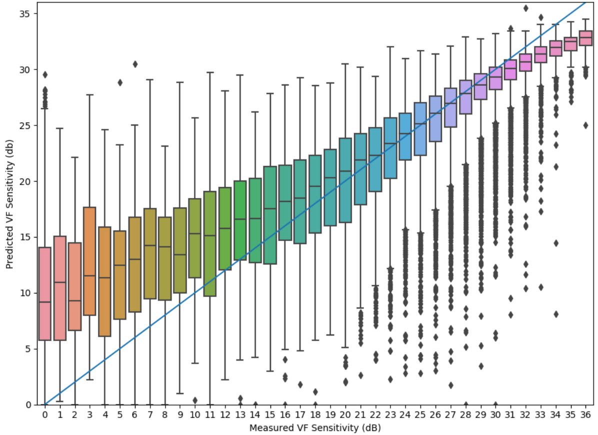
- Improved interpretability
- Multi-channel: working with missing data/data imputation
- Simulations for clinical trials

Large-scale applications

From RNFL thickness to visual fields

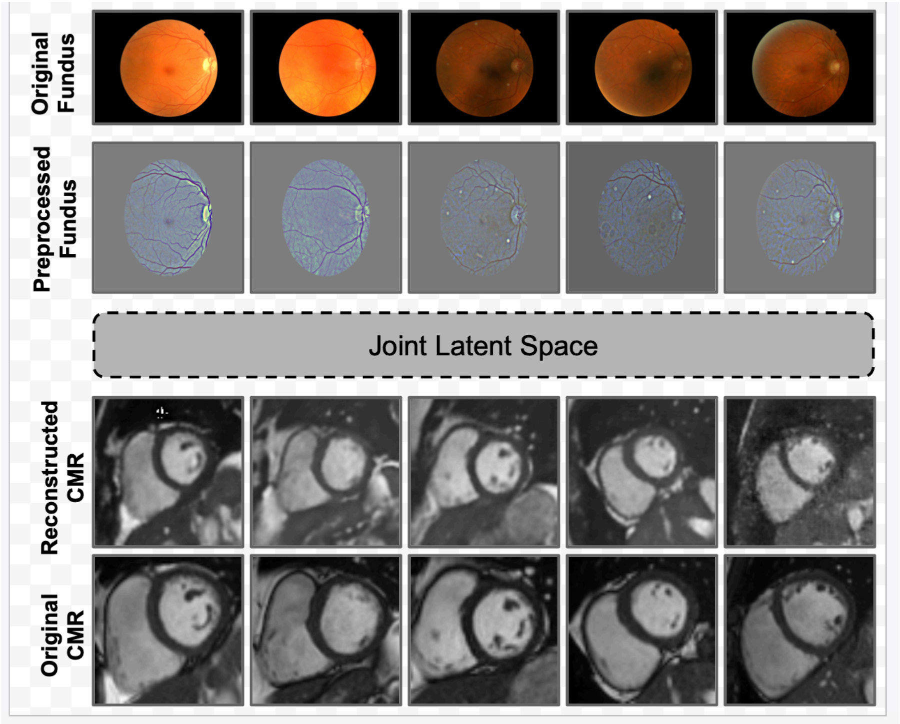






Courtesy of Lazaridis et al.
work in progress, UCL-Inria collaboration

From fundus to cardiac images



Courtesy of Diaz-Pinto et al.
work in progress at CISTIB, University of Leeds, UK

Thank you