The Enigmatic *Kime*: Time Complexity in Data Science

Ivo D. Dinov

Statistics Online Computational Resource
Health Behavior & Biological Sciences
Computational Medicine & Bioinformatics
Michigan Institute for Data Science

University of Michigan

http://SOCR.umich.edu

joint work with Milen V. Velev (BTU)

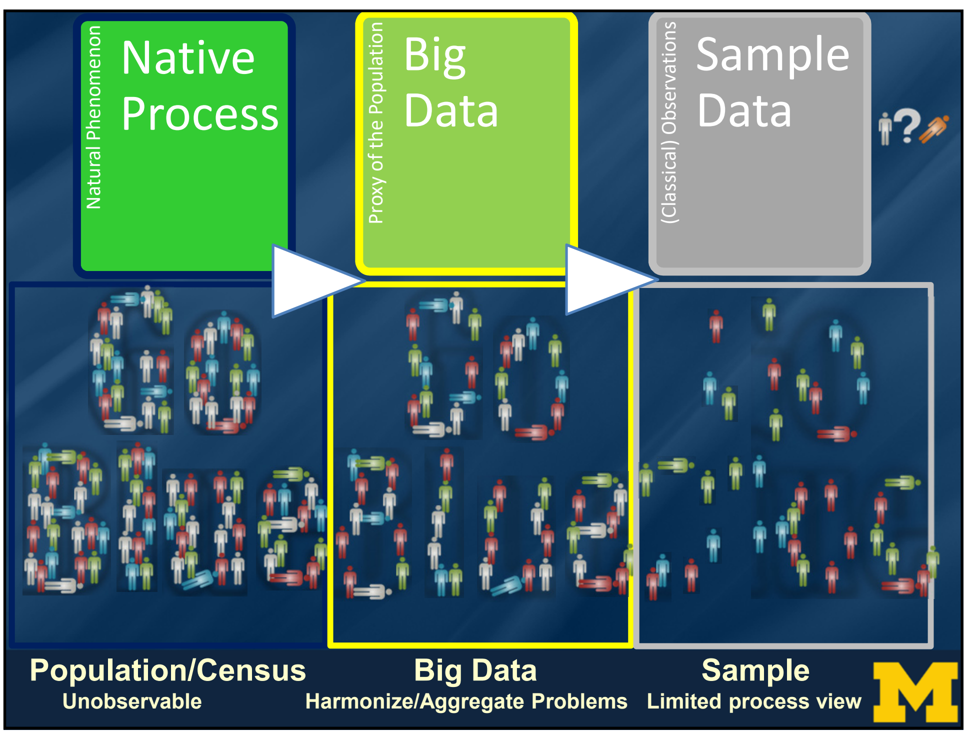
SCHOOL OF NURSING
STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR)
UNIVERSITY OF MICHIGAN

Slides Online:
“SOCR News”

Outline

- ❑ Big Biomedical/Health Data
- ❑ Data Analytic Challenges
- ❑ Mathematical Physics Background
- ❑ Time complexity (*kime*)
- ❑ Advanced predictive analytics & scientific inference
- ❑ Inferential Uncertainty

Big Biomedical/Health Data



Characteristics of Big Biomed Data

IBM Big Data 4V's: Volume, Variety, Velocity & Veracity

Big Bio Data Dimensions	Tools
Size	Harvesting and management of vast amounts of data
Complexity	Wranglers for dealing with heterogeneous data
Incongruency	Tools for data harmonization and aggregation
Multi-source	Transfer and joint modeling of disparate elements
Multi-scale	Macro to meso to micro scale observations
Time	Techniques accounting for longitudinal effects
Incomplete	Reliable management of missing data

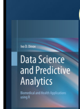
Example: analyzing observational data of 1,000's Parkinson's disease patients based on 10,000's signature biomarkers derived from multi-source imaging, genetics, clinical, physiologic, phenomics and demographic data elements

Software developments, student training, service platforms and methodological advances associated with the Big Data Discovery Science all present existing opportunities for learners, educators, researchers, practitioners and policy makers

Dinov, *GigaScience* (2016) PMID:26918190

Data Science & Predictive Analytics

- **Data Science**: an emerging extremely transdisciplinary field - bridging between the theoretical, computational, experimental, and applied areas. Deals with enormous amounts of complex, incongruent and dynamic data from multiple sources. Aims to develop algorithms, methods, tools, and services capable of ingesting such datasets and supplying semi-automated decision support systems
- **Predictive Analytics**: process utilizing advanced mathematical formulations, powerful statistical computing algorithms, efficient software tools, and distributed web-services to represent, interrogate, and interpret complex data. Aims to forecast trends, cluster patterns in the data, or prognosticate the process behavior either within the range or outside the range of the observed data (e.g., in the future, or at locations where data may not be available)



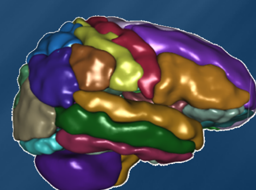
http://DSPA.predictive.space

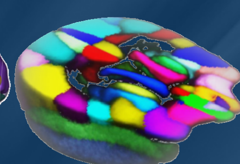
Dinov, Springer (2018)

Data Analytic Challenges

Driving Biomedical/Health Challenges

- **Neurodegeneration**: Structural Neuroimaging in Alzheimer's Disease illustrates the Big Data challenges in modeling complex neuroscientific data. 808 ADNI subjects, 3 groups: 200 subjects with Alzheimer's disease (AD), 383 subjects with mild cognitive impairment (MCI), and 225 asymptomatic normal controls (NC). The 80 neuroimaging biomarkers and 80 highly-associated SNPs.





A: Individual brain parcellation			B: LPBA40 atlas		
Index	Volume Intensity	ROI Name	Index	Volume Intensity	ROI Name
1	21	L superior frontal gyrus	29	65	L inferior occipital gyrus
2	24	R middle frontal gyrus	30	164	R putamen
3	50	R precuneus	31	61	L superior occipital gyrus
4	181	cerebellum	32	30	R middle orbitofrontal gyrus
5	47	L angular gyrus	33	42	R postcentral gyrus
6	122	R cingulate gyrus	34	27	L precentral gyrus
7	83	L middle temporal gyrus	35	32	R lateral orbitofrontal gyrus
8	90	R lingual gyrus	36	121	L cingulate gyrus
9	81	L superior temporal gyrus	37	31	L lateral orbitofrontal gyrus
10	91	L fusiform gyrus	38	92	R fusiform gyrus
11	44	R superior parietal gyrus	39	45	L supramarginal gyrus
12	66	R inferior occipital gyrus	40	88	R parahippocampal gyrus
13	87	L parahippocampal gyrus	41	22	R superior frontal gyrus
14	162	R caudate	42	29	L middle orbitofrontal gyrus
15	85	L inferior temporal gyrus	43	68	R cuneus
16	182	brainstem	44	62	R superior occipital gyrus
17	43	L superior parietal gyrus	45	33	L gyrus rectus
18	28	R precentral gyrus	46	48	R angular gyrus
19	23	L middle frontal gyrus	47	64	R middle occipital gyrus
20	89	L lingual gyrus	48	84	R middle temporal gyrus
21	41	L postcentral gyrus	49	49	L precuneus

http://DSPA.predictive.space
Moon, Dinov, *et al.* (2015)

Big Data	Information	Knowledge	Action
Raw Observations	Processed Data	Maps, Models	Actionable Decisions
Data Aggregation	Data Fusion	Causal Inference	Treatment Regimens
Data Scrubbing	Summary Stats	Networks, Analytics	Forecasts, Predictions
Semantic-Mapping	Derived Biomarkers	Linkages, Associations	Healthcare Outcomes

Dinov, et al. (2016) PMID:26918190

Longitudinal ALS Study

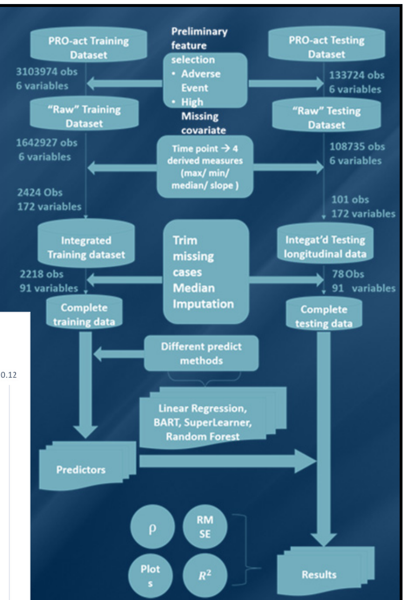
- ❑ Identify predictive classifiers to detect, track and prognosticate the progression of ALS (in terms of clinical outcomes like ALSFRS and muscle function)
- ❑ Provide a decision tree prediction of adverse events based on subject phenotype and 0-3 month clinical assessment changes

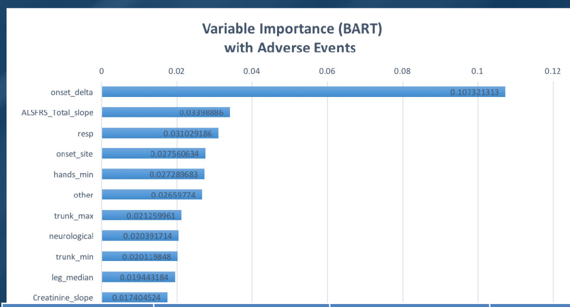
Data Source	Sample Size/Data Type	Summary
ProAct Archive	Over 100 variables are recorded for all subjects including: Demographics: age, race, medical history, sex; Clinical data: **Amyotrophic Lateral Sclerosis** Functional Rating Scale (ALSFRS), adverse events, onset_delta, onset_site, drugs use (riluzole) The PRO-ACT training dataset contains clinical and lab test information of 8,635 patients. Information of 2,424 study subjects with valid gold standard ALSFRS slopes used for processing, modeling and analysis	The time points for all longitudinally varying data elements are aggregated into signature vectors. This facilitates the modeling and prediction of ALSFRS slope changes over the first three months (baseline to month 3)

Tang, et al. (2018), Neuroinformatics

Case-Studies – ALS

- Detect, track, and prognosticate the progression of ALS
- Predict adverse events based on subject phenotype and 0-3 month clinical assessment changes





Methods	Linear Regression	Random Forest	BART	SuperLearner
R-squared	0.081	0.174	**0.225**	0.178
RMSE	0.619	0.587	**0.568**	0.585
Correlation	0.298	0.434	**0.485**	0.447

Math & Physics Background

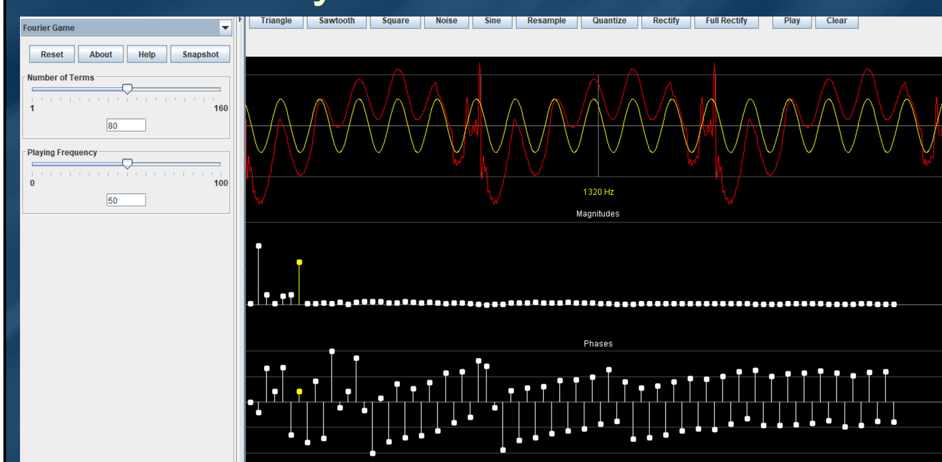
Math & Physics – Fourier Transform

By separability, the spacetime Fourier transform is just four Fourier transforms, one for each of the four spacetime dimensions, $(\boldsymbol{x}, t) = (x, y, z, t)$. Although the FT exp-sign may be chosen either way, traditionally the sign convention represents a wave with angular frequency ω that propagates in the wave number direction $\boldsymbol{k}$ (space frequency). Symbolically, for a 4D ($n = 4$) spacetime function f, the forward and inverse Fourier transforms are defined by:

$$FT(f) = \hat{f}(\boldsymbol{k}, \omega) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int f(\boldsymbol{x}, t) e^{i(\omega t - \boldsymbol{kx})} dt d^3\boldsymbol{x},$$

$$IFT(\hat{f}) = \hat{\hat{f}}(\boldsymbol{x}, t) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int \hat{f}(\boldsymbol{k}, \omega) e^{-i(\omega t - \boldsymbol{kx})} d\omega d^3\boldsymbol{k}.$$

Math & Physics – Fourier Transform



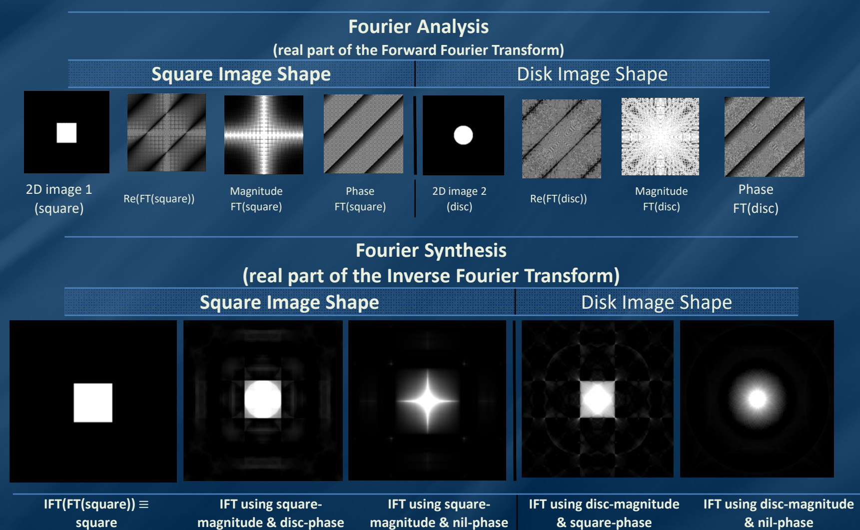
SOCR 1D Fourier / Wavelet signal decomposition into *magnitudes* and *phases (Java applet)*

Top-panel: original signal (image), white-color curve drawn manually by the user and the reconstructed synthesized (IFT) signal, red-color curve, computed using the user modified magnitudes and phases

Bottom-panels: the Fourier analyzed signal (FT) with its magnitudes and phases

http://www.socr.ucla.edu/htmls/game/Fourier_Game.html

Math & Physics – Fourier Transform



Math & Physics – Wavefunctions

- For simplicity, focus on 1D space + time:

  $\psi(x,t)$ represents the ℂ-valued wave (amplitude + phase)

  $$\frac{\partial^2\psi(x,t)}{\partial t^2} = c^2\frac{\partial^2}{\partial x^2}\psi(x,t), \qquad x \in R, t \in R^+, c = \text{speed of light } m/s$$

- Correspondence: observable $(x) \Leftrightarrow$ operator $(\hat{x})$ acting on ψ: $\hat{x}(\psi) = \hat{x}|\psi\rangle = |\psi'\rangle$

  measured observable instance, x, corresponds to operator expectation $\langle\hat{x}\rangle$

- Schrödinger picture (PDE based) – operators=constant, states=time-dependent

  $$\underbrace{i\hbar\frac{\partial}{\partial t}}_{\text{Energy}}|\psi\rangle = \underbrace{H}_{\text{Hamiltonian}}|\psi\rangle$$, general differential formulation w.r.t. states $|\psi\rangle$

- Heisenberg picture (Linear Algebraic) – operators=time-dependent, states=timeless

  $$\frac{d}{dt}A(t) = \frac{i}{\hbar}\underbrace{[H,A(t)]}_{HA-AH} + \left(\frac{\partial A}{\partial t}\right)_H$$, linear operator A corresponding to an observable

  $$\text{Expectation: } E(A) \equiv \langle A\rangle_t = \langle\psi(t)|A(t)|\psi(t)\rangle = \int_R \psi^*(x)A\psi(x)dx, \quad ^*\text{compex conjugate}$$

  $$\text{Uncertainty: } (\Delta A)^2 \equiv \langle A - \underbrace{\langle A\rangle_t}_{\text{constant}}\rangle_t = \langle\psi(t)|(A(t)-\langle A\rangle_t)^2|\psi(t)\rangle$$

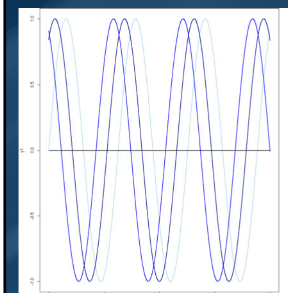
- Example position $(\hat{x} = x)$-momentum $(\hat{p} = -i\hbar\frac{\partial}{\partial x})$ commutator:

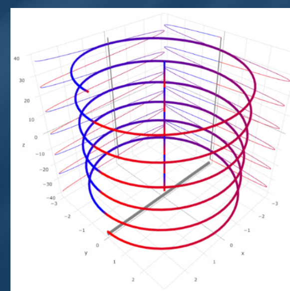
  $$x \to \hat{x} = x$$

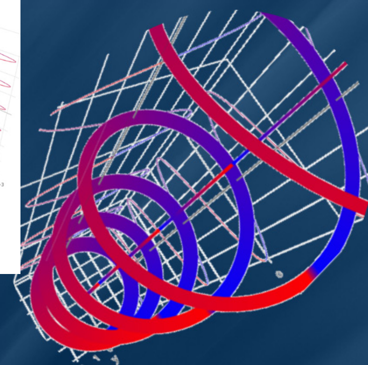
  $$p_x \to \hat{p} = -i\hbar\frac{\partial}{\partial x} \to [\hat{x},\hat{p}]\psi(x,t) = \left[x, -i\hbar\frac{\partial}{\partial x}\right]\psi(x,t) = \cdots = i\hbar\psi(x,t) \neq 0$$

  $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant

Math & Physics – Wavefunctions





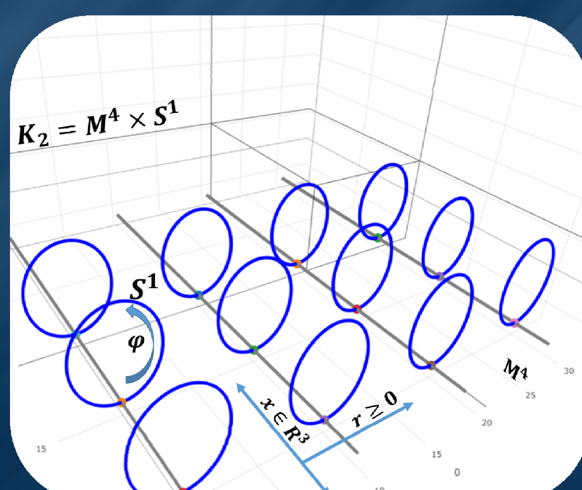


2D (oscillatory amplitude)

3D (fixed amplitude, complex phase)

Math & Physics – Kaluza-Klein Theory

- Theodor Kaluza developed (1921) an extension of the classical general relativity theory to 5D. This included the metric, the field equations, the equations of motion, the stress-energy tensor, and the cylinder condition. Oskar Klein (1926) interpreted Kaluza's 3D+2D theory in quantum mechanical space and proposed that the fifth dimension was curled up and microscopic.
- The topology of the 5D Kaluza-Klein spacetime is $K_2 \cong M^4 \times S^1$, where M^4 is a 4D Minkowski spacetime and S^1 is a circle (non-traversable).



Time complexity (*kime*)

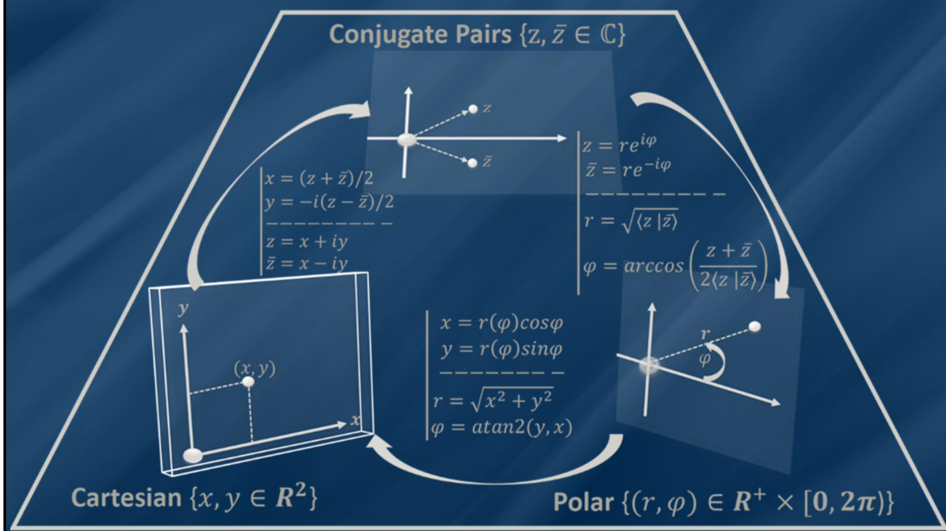
Time complexity (*kime*)

- **Spacekime:** $(\boldsymbol{x}, \boldsymbol{k}) = \left(\underbrace{x^1, x^2, x^3}_{\text{space}}, \underbrace{c\kappa_1 = x^4, c\kappa_2 = x^5}_{\text{kime}} \right) \in X$
- **Kevents** are points (or states) in the spacekime manifold X. Each kevent is defined by where ($\boldsymbol{x} = (x, y, z)$) it occurs in space, what is its *causal longitudinal order* $\left(r = \sqrt{(x^4)^2 + (x^5)^2} \right)$, and in what *kime-direction* ($\varphi = \operatorname{atan2}(x^5, x^4)$) it takes place.
- The general Minkowski 5×5 metric tensor $(\lambda_{ij})_{i=1,j=1}^{5,5}$ characterizes the geometry of the *curved spacekime*.

$$ds^2 = \sum_{i=1}^{5} \sum_{j=1}^{5} \lambda_{ij} dx^i dx^j = \lambda_{ij} dx^i dx^j$$

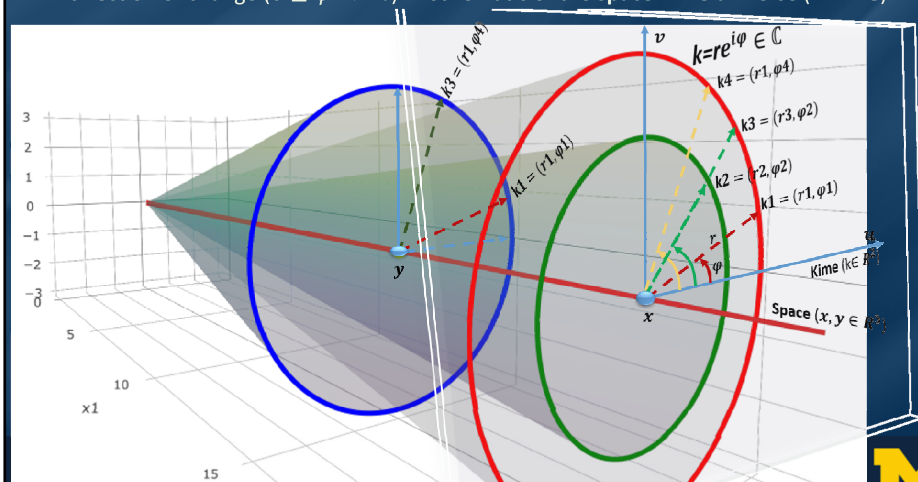
- *Euclidean (flat) spacekime* metric corresponds to the tensor: $(\lambda_{ij}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$
  - *Spacelike* intervals correspond to $ds^2 > 0$, where an inertial frame can be found such that two kevents $a, b \in X$ are simultaneous. An object can't be present at two kevents which are separated by a spacelike interval.
  - *Lightlike* intervals correspond to $ds^2 = 0$. If two events are on the line of a photon, then they are separated by a lightlike interval and a ray of light could travel between the two events.
  - *Kimelike* intervals correspond to $ds^2 < 0$. An object can be present at two different kevents, which are separated by a kimelike interval.

Kime Parameterization



Time complexity (*kime*)

- **Spacekime:** complex time (***kime***) represents the dual nature of *events order* ($r > 0$), characterizing the longitudinal displacement in time, and *event phase*, reflecting the direction of change ($0 \leq \varphi < 2\pi$). A schematic of the **space-kime** universe ($R^3 \times \mathbb{C}$).



Kime Math Generalizations

- **Spacekime generalization of Lorentz transform between two reference frames:**

$$\begin{pmatrix} x' \\ y' \\ z' \\ {k_1}' \\ {k_2}' \end{pmatrix} = \begin{pmatrix} \zeta & 0 & 0 & -\frac{c^2}{v_1}\beta^2\zeta & -\frac{c^2}{v_2}\beta^2\zeta \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -\frac{1}{v_1}\beta^2\zeta & 0 & 0 & 1+(\zeta-1)\frac{c^2}{(v_1)^2}\beta^2 & (\zeta-1)\frac{c^2}{v_1 v_2}\beta^2 \\ -\frac{1}{v_2}\beta^2\zeta & 0 & 0 & (\zeta-1)\frac{c^2}{v_1 v_2}\beta^2 & 1+(\zeta-1)\frac{c^2}{(v_2)^2}\beta^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ k_1 \\ k_2 \end{pmatrix}$$

- **Wirtinger derivative & kime-acceleration** (second order kime-derivative at $\boldsymbol{k} = (r, \varphi)$):

$$f''(\boldsymbol{k}) = \frac{1}{2}\left(\begin{array}{c} \underbrace{cos(2\varphi)\frac{\partial^2 f}{\partial r^2} - \frac{2}{r}sin2\varphi\frac{\partial^2 f}{\partial r \partial \varphi} - \frac{1}{r^2}cos(2\varphi)\frac{\partial^2 f}{\partial \varphi^2}}_{real} \\ \underbrace{-i\left(sin(2\varphi)\frac{\partial^2 f}{\partial r^2} + \frac{2}{r}cos(2\varphi)\frac{\partial^2 f}{\partial r \partial \varphi} - \frac{1}{r^2}sin(2\varphi)\frac{\partial^2 f}{\partial \varphi^2}\right)}_{imaginary} \end{array} \right)$$

- **Many others, including law of addition of velocities, energy-momentum conservation law, stability conditions for particles moving in space-kime, conditions for nonzero rest particle mass, and causal structure of space-kime …**

Where is the Data Science?

Math-Physics ⟹ Data Science

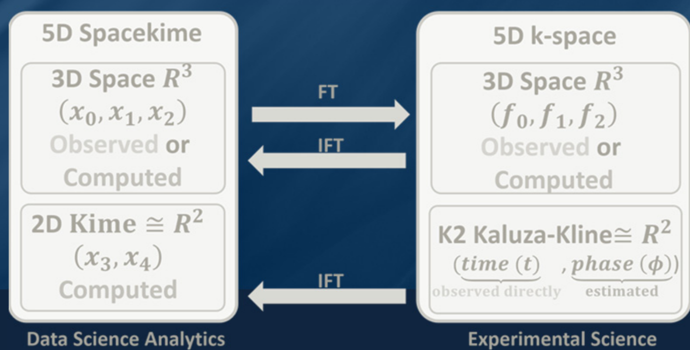
Math-Physics	Data Science
A **particle** is a small localized object that permits observations and characterization of its physical or chemical properties	An **object** is something that exists by itself, actually or potentially, concretely or abstractly, physically or incorporeal (e.g., person, subject, etc.)
An **observable** a dynamic variable about particles that can be measured	A **feature** is a dynamic variable or an attribute about an object that can be measured
Particle **state** is an observable particle characteristic (e.g., position, momentum)	**Datum** is an observed quantitative or qualitative value, an instantiation, of a feature
Particle **system** is a collection of as a collection of independent objects, particles, in a closed system	**Problem**, aka Data System, is a collection of independent objects, without necessarily associated with some a priori hypotheses
Wave-function	Inference-function
Reference-Frame Transforms (e.g., Lorentz)	Data transformations (e.g., wrangling, log-transform)
State of the system is an observed measurement of all particles ~ wavefunction	Dataset (data) is an observed instance of a set of datum elements about the problem system, $\boldsymbol{O} = \{X, Y\}$.
A **particle system is computable** if (1) the entire system is logical, consistent, complete and (2) the unknown internal states of the system don't influence the computation (wavefunction, intervals, probabilities, etc.)	**Computable data object** is a very special representation of a dataset which allows direct application of computational processing, modeling, analytics, or inference based on the observed dataset

Math-Physics ⟹ Data Science

Math-Physics	Data Science
Wavefunction $\psi(x,t) = Ae^{i(kx-wt)}$ represents a traveling wave. Note that: $\frac{\partial^2}{\partial x^2}\psi = -k^2\psi.$	*Inference function* - describing of a solution to a specific data analytic system (problem). For example, • A <u>linear (GLM) model</u> represents a solution of a prediction inference problem where the inference function quantifies the effects for all independent features (X) on the dependent outcome (Y), $\boldsymbol{O} = \{X, Y\}$: $\psi(\boldsymbol{O}) = \psi(X,Y) = \hat{\beta} = \hat{\beta}^{OLS} = \langle X\|X\rangle^{-1}\langle X\|Y\rangle = \left(X^T X\right)^{-1} X^T Y.$ • A non-parametric, <u>non-linear</u>, alternative inference is SVM classification. If $\psi_x \in \boldsymbol{H}$, is the lifting function $\psi: \boldsymbol{R}^\eta \to \boldsymbol{R}^d$ ($\psi: x \in \boldsymbol{R}^\eta \to \tilde{x} = \psi_x \in \boldsymbol{H}$), where $\eta \ll d$ and the kernel $\psi_x(y) = \langle x\|y\rangle: \boldsymbol{O} \times \boldsymbol{O} \to \boldsymbol{R}$, the observed data $\boldsymbol{O}_i = \{x_i, y_i\}$ are lifted to $\psi_{\boldsymbol{O}_i}$. Then, the SVM prediction operator is the weighted sum of the kernel functions at $\psi_{\boldsymbol{O}_i}$ (where β^* is a solution to the SVM regularized optimization): $\langle \psi_O \| \beta^* \rangle_H = \sum_{i=1}^{n} p_i^* \langle \psi_O \| \psi_{O_i} \rangle_H .$ The linear coefficients, p_i^*, are the dual weights that are multiplied by the label corresponding to each training instance, $\{y_i\}$. The inference always depends on the (input) data, however, it does not have 1-1 and onto bijective correspondence with the data, as the inference function quantifies the predictions in a probabilistic sense.

Space-*kime* Analytics

- Often, we can't directly observe (record data) in 5D spacekime.
- Yet, we can measure quite accurately the kime-magnitudes (r) as event orders, "times".
- To reconstruct the 2D spatial structure of kime, borrow tricks used by crystallographers [1] to resolve the structure of atomic particles by only observing the magnitudes of the diffraction pattern in k-space. This approach heavily relies on (1) prior information about the kime directional orientation, which may be obtained from using similar datasets and phase-aggregator analytical strategies, or (2) experimental reproducibility by repeated confirmations of the data analytic results using longitudinal datasets.



[1] Rodriguez, Ivanova, Nature 2015

Space-*kime* Analytics: Example

- 3D isosurface Reconstruction of (2,1) fMRI signal

Reconstruction using trivial phase-angle; kime=time=(magnitude, 0)

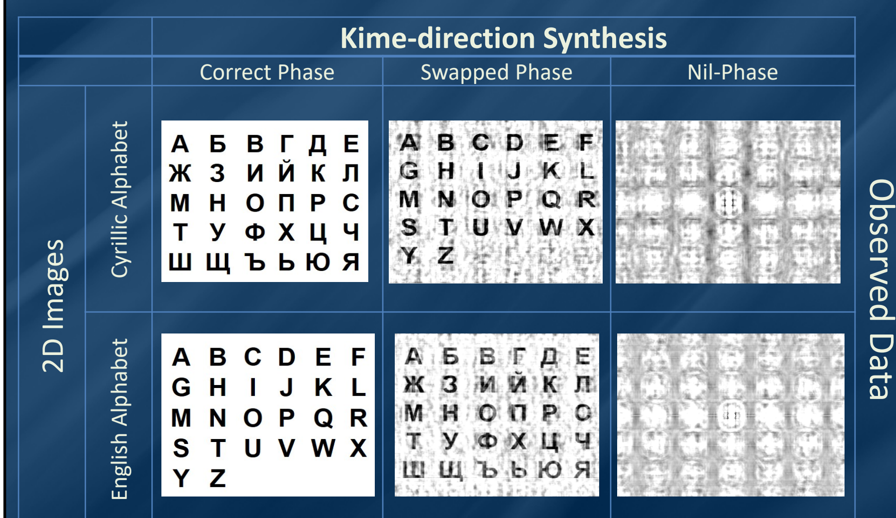
Reconstruction using correct kime=(magnitude, phase)

3D pseudo-spacetime reconstruction:

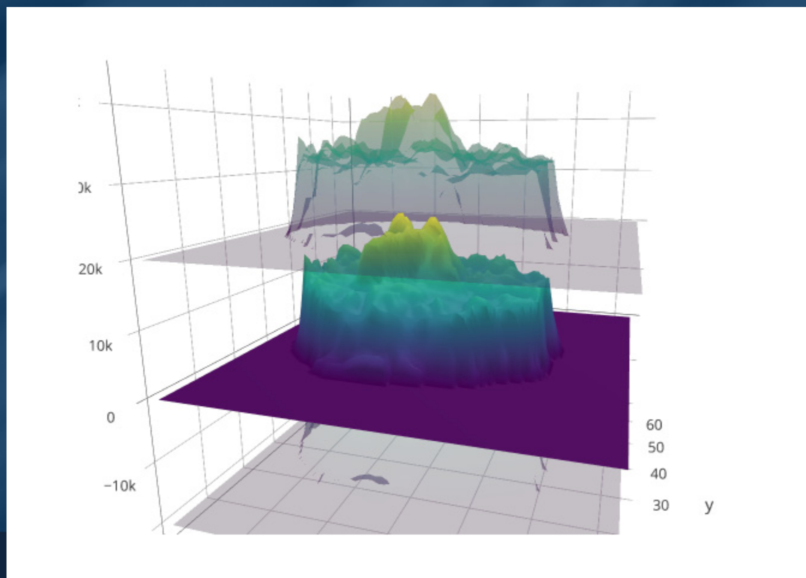
$$f = \hat{h}(\underbrace{x_1, x_2}_{space}, \underbrace{t}_{time})$$

Advanced predictive analytics & scientific inference

2D Image Analysis / Character Recognition



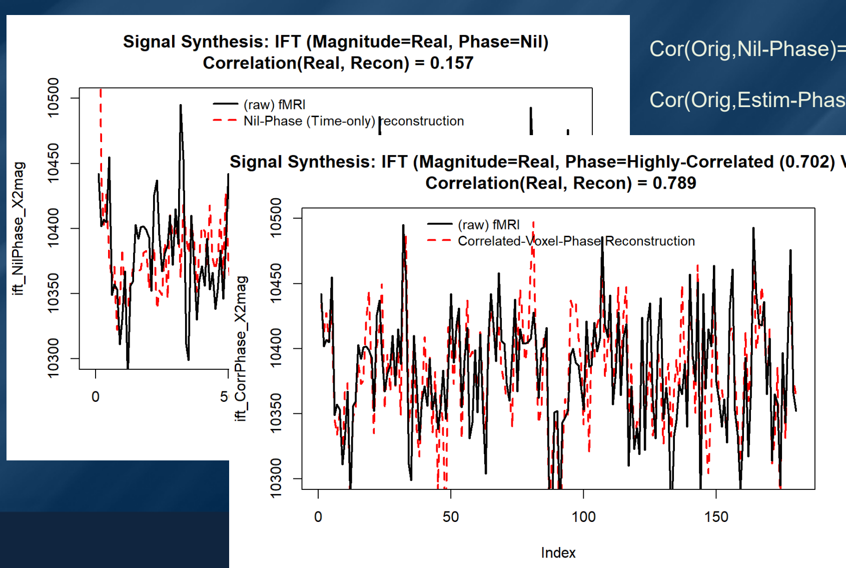
Back to fMRI (4D spacetime data)



3D rendering of 3 time cross-sections of the fMRI series over a 2D spatial domain

Back to fMRI (4D spacetime data)

Reconstruction of the fMRI timeseries at one spatial voxel location

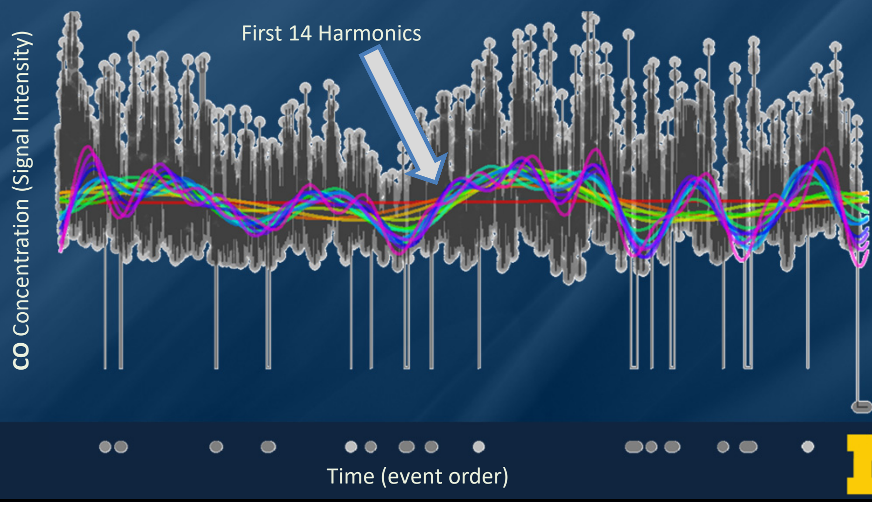


Cor(Orig,Nil-Phase)=0.16

Cor(Orig,Estim-Phase)=0.79

Exogenous Feature Time-series Analysis

ARIMAX modeling of UCI ML Air Quality Dataset (9,358 hourly-averaged CO responses from an array of sensors). Demonstrate the effect of kime-direction on the analysis of the longitudinal data.



Exogenous Feature Time-series Analysis

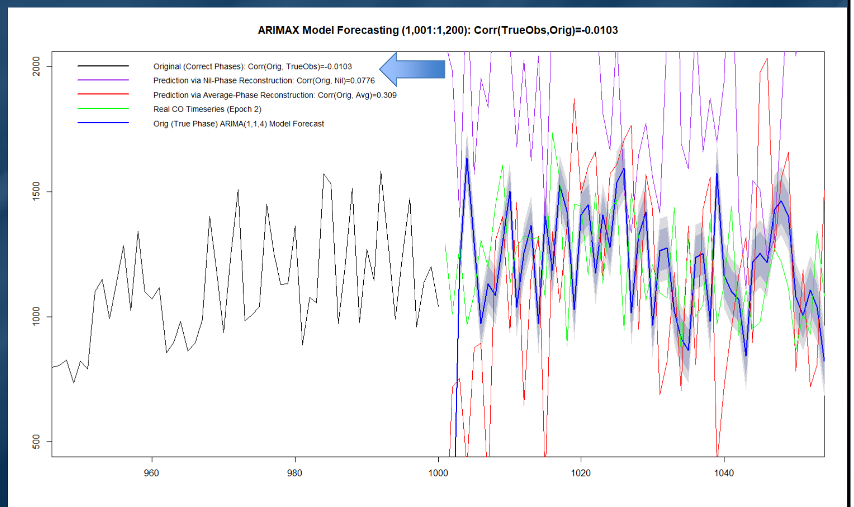
CO ARIMAX models derived on 3 different signal reconstructions based on alternative kime-direction estimates

Phase	Nil	Average	True=original
Model Estimate	ARIMA(2,0,1)	ARIMA(2,0,3)	ARIMA(1,1,4)
AIC	**13179**	**14183**	**10581**
ar1	**1.11406562**	**0.329482302**	**0.2765312**
ar2	-0.14565048	0.238363531	.
ma1	-0.78919188	0.267291585	-0.88913497
ma2	**.**	**-0.006079386**	**0.12679494**
ma3	**.**	**0.15726556**	**0.03043726**
ma4	.	.	-0.17655728
intercept	503.3455144	742.800113	.
xreg1	-0.40283891	0.58379483	0.08035744
xreg2	**0.13656613**	**0.280936931**	**6.14947902**
xreg3	-0.51457636	-0.649722755	0.09859223
xreg4	1.09611981	1.239910298	0.01634736
xreg5	**1.21946209**	**-0.026110332**	**-0.04816591**
xreg6	**1.30628469**	**1.081777956**	**-0.01104142**
xreg7	**1.20868397**	**0.254018471**	**0.1832854**
xreg8	**1.14905809**	**0.306524131**	**0.17648482**
xreg9	**-0.48233756**	**-0.405204908**	**6.53739782**
xreg10	**0.03145281**	**0.351063312**	**1.79388326**
xreg11	-0.46395772	-0.457689796	-12.06965578

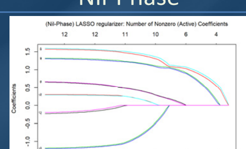
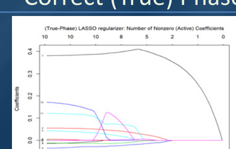
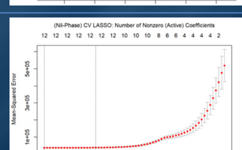
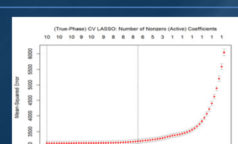
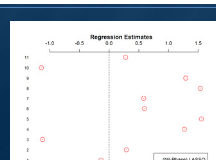
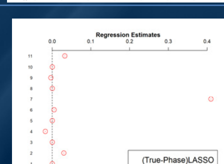
ARIMAX (*p*,*d*,*q*)
p = order (# of time lags) of the AR part
d = differencing (# of past values subtractions)
q = order of MA part

Exogenous Feature Time-series Forecasting

CO ARIMAX models derived on 3 different signal reconstructions based on alternative kime-direction estimates



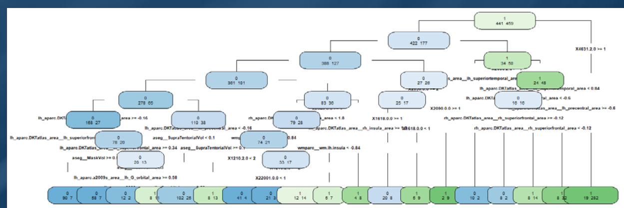
Exogenous Feature Time-series Analysis

	Synthesis Approach	
	Nil-Phase	Correct (True) Phase
Number of Nonzero (Active) LASSO Coefficients		
LASSO Mean Square Error CV Error of Model Coefficients		
LASSO Regression Model Coefficients		

Results of **regularized linear modeling** of CO-concentration using LASSO penalty

Big Data Analytics Study – UKBB

- 9,914 UKBB participants; 7,614 clinical measurements, phenotypic features, and derived neuroimaging biomarkers
- Supervised Decision Tree (binary Dx) Classification – **Correct Kime-Phase Estimates**

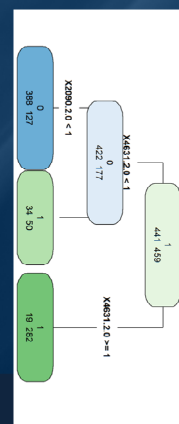


Raw Decision Tree

Prediction 0 1
0 362 60
1 79 399
Accuracy : 0.8456
95% CI : (0.82, 0.87)
No Information Rate : 0.51
P-Value [Acc > NIR] : <2e-16
Kappa : 0.6907
Mcnemar's Test P-Value : 0.1268
Sensitivity : 0.8209
Specificity : 0.8693
Detection Rate : 0.4022
Detection Prevalence : 0.4689
Balanced Accuracy : 0.8451

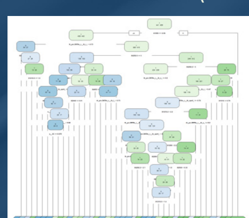
Pruned Decision Tree

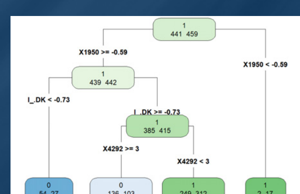
Prediction 0 1
0 388 127
1 53 332
Accuracy : 0.8
95% CI : (0.77, 0.83)
No Information Rate : 0.51
P-Value [Acc > NIR] : < 2.2e-16
Kappa : 0.6012
Mcnemar's Test P-Value : 5.295e-08
Sensitivity : 0.8798
Specificity : 0.7233
Detection Prevalence : 0.5722
Balanced Accuracy : 0.8016



Big Data Analytics Study – UKBB

- 9,914 UKBB participants; 7,614 clinical measurements, phenotypic features, and derived neuroimaging biomarkers (11 epochs of 900 cases)
- Supervised Decision Tree (binary) Classification – **Epoch-average Kime-Phases**





Raw Decision Tree

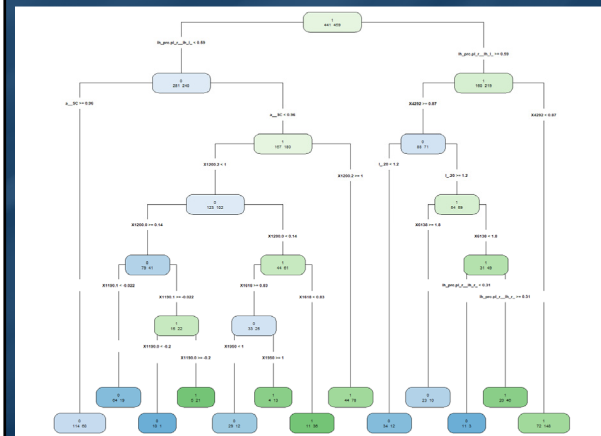
Reference
Prediction 0 1
0 354 85
1 87 374
Accuracy : 0.8089
95% CI : (0.78, 0.83)
No Information Rate : 0.51
P-Value [Acc > NIR] : <2e-16
Kappa : 0.6176
Mcnemar's Test P-Value : 0.9392
Sensitivity : 0.8027
Specificity : 0.8148
Detection Rate : 0.3933
Detection Prevalence : 0.4878
Balanced Accuracy : 0.8088

Pruned Decision Tree

Reference
Prediction 0 1
0 190 130
1 251 329
Accuracy : 0.5767
95% CI : (0.54, 0.61)
No Information Rate : 0.51
P-Value [Acc > NIR] : 3.501e-05
Kappa : 0.1484
Mcnemar's Test P-Value : 7.857e-10
Sensitivity : 0.4308
Specificity : 0.7168
Detection Rate : 0.2111
Detection Prevalence : 0.3556
Balanced Accuracy : 0.5738

Big Data Analytics Study – UKBB

- 9,914 UKBB participants; 7,614 clinical measurements, phenotypic features, and derived neuroimaging biomarkers
- Supervised Decision Tree (binary) Classification – **Nil-average Kime-Phases**



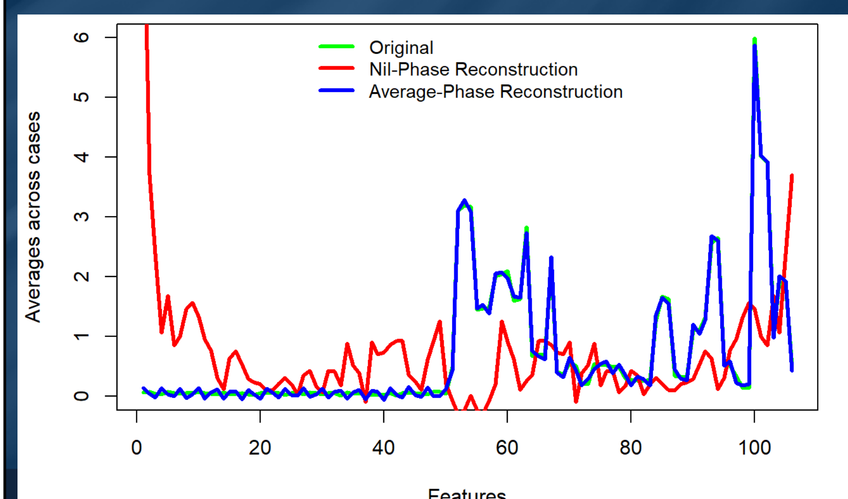
Raw Decision Tree

```
##          Reference
## Prediction   0   1
##        0 341  86
##        1 100 373
##        Accuracy : 0.7933
##          95% CI : (0.77, 0.82)
##    No Information Rate : 0.51
##    P-Value [Acc > NIR] : <2e-16
##               Kappa : 0.5862
## Mcnemar's Test P-Value : 0.3405
##          Sensitivity : 0.7732
##          Specificity : 0.8126
##       Detection Rate : 0.3789
##   Detection Prevalence : 0.4744
##    Balanced Accuracy : 0.7929
```

Pruned Decision Tree was degenerate

Big Data Analytics Study – UKBB

- 9,914 UKBB participants; 7,614 clinical measurements, phenotypic features, and derived neuroimaging biomarkers. Supervised Decision Tree (binary) Classification

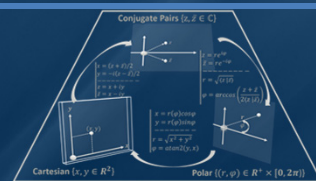


Overall feature averages across cases for the 3 complementary kime-reconstruction analytic strategies

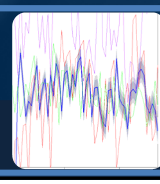
Summary

- There are substantial Big Biomed/Health Data Challenges

- Mathematical-physics models help with representation and analysis of complex (temporal) data



- *Spacekime representation is useful for* advanced predictive analytics



Acknowledgments

Funding

NIH: P20 NR015331, U54 EB020406, P50 NS091856, P30 DK089503, P30AG053760, UL1TR002240

NSF: 1734853, 1636840, 1416953, 0716055, 1023115

The Elsie Andresen Fiske Research Fund

http://SOCR.umich.edu

Collaborators

- SOCR: Milen Velev, Alexandr Kalinin, Selvam Palanimalai, Syed Husain, Matt Leventhal, Ashwini Khare, Rami Elkest, Abhishek Chowdhury, Patrick Tan, Gary Chan, Andy Foglia, Pratyush Pati, Brian Zhang, Juana Sanchez, Dennis Pearl, Kyle Siegrist, Rob Gould, Jingshu Xu, Nellie Ponarul, Ming Tang, Asiyah Lin, Nicolas Christou, Hanbo Sun, Tuo Wang. Simeone Marino
- LONI/INI: Arthur Toga, Roger Woods, Jack Van Horn, Zhuowen Tu, Yonggang Shi, David Shattuck, Elizabeth Sowell, Katherine Narr, Anand Joshi, Shantanu Joshi, Paul Thompson, Luminita Vese, Stan Osher, Stefano Soatto, Seok Moon, Junning Li, Young Sung, Carl Kesselman, Fabio Macciardi, Federica Torri
- UMich MIDAS/MNORC/AD/PD Centers: Cathie Spino, Chuck Burant, Ben Hampstead, Stephen Goutman, Stephen Strobbe, Hiroko Dodge, Hank Paulson, Bill Dauer, Brian Athey