Applied Machine Learning for Design Optimization in Cosmology, Neuroscience, and Drug Discovery

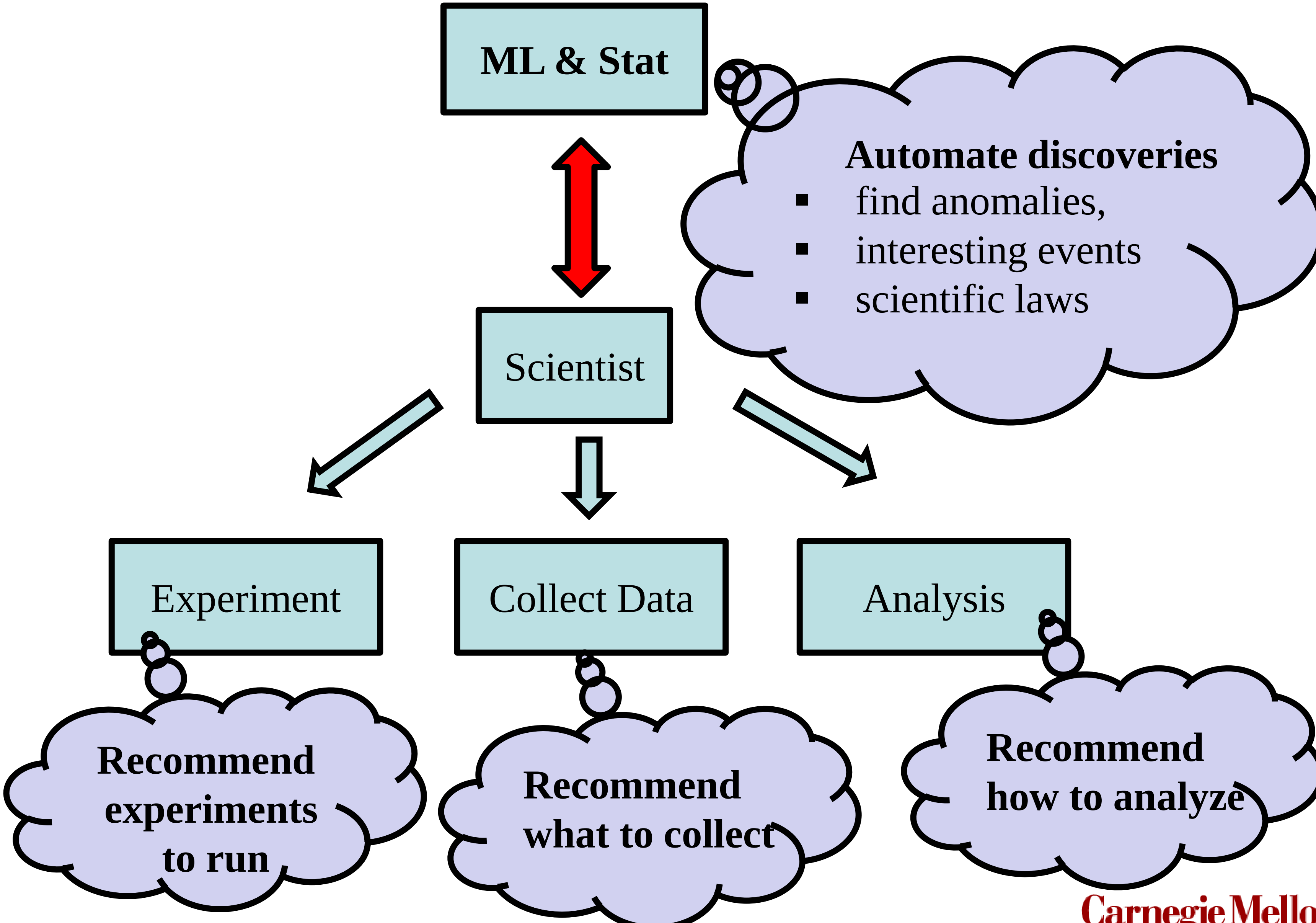
Barnabas Poczos
Machine Learning Department
Carnegie Mellon University

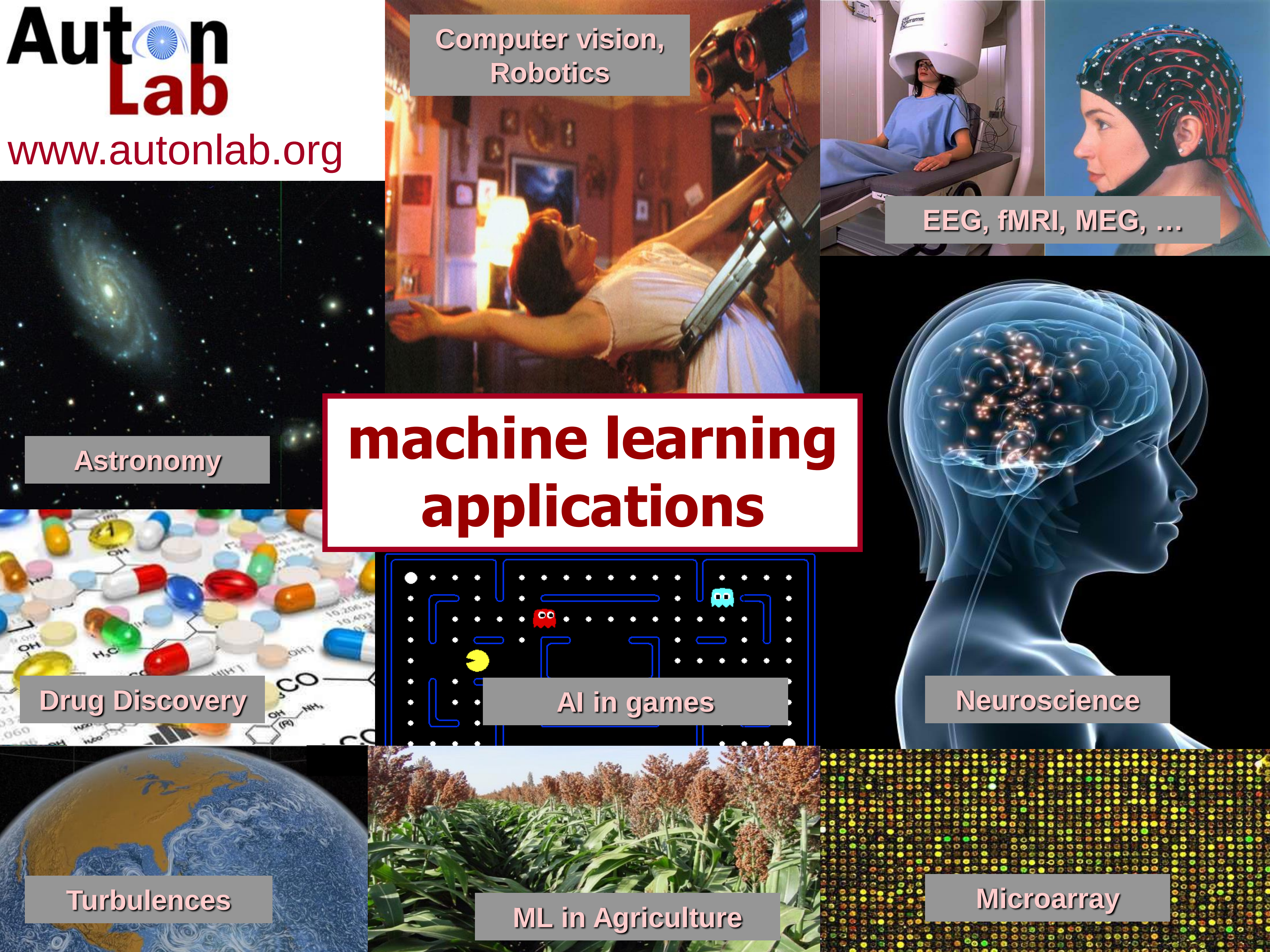
Machine Learning Technologies and their Applications for Scientific and Engineering Domains Workshop

NASA Langley Research Center

August 16, 2016

Goal: Create a Scientific Assistant



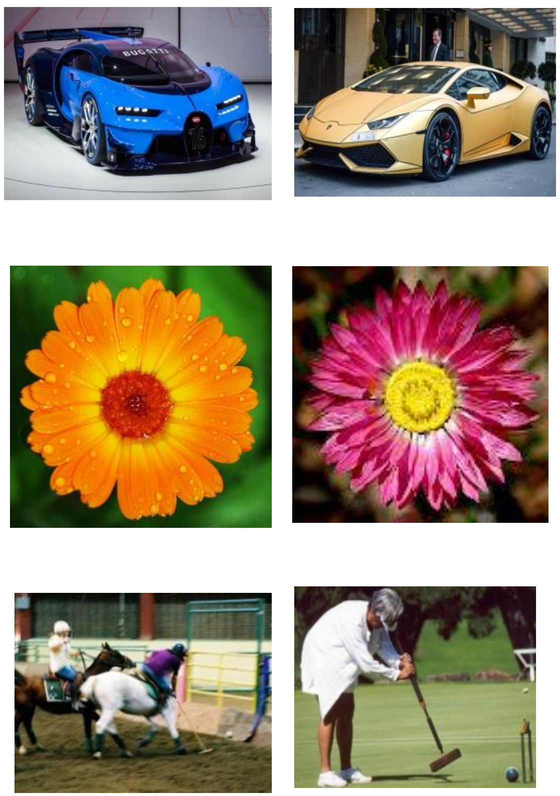


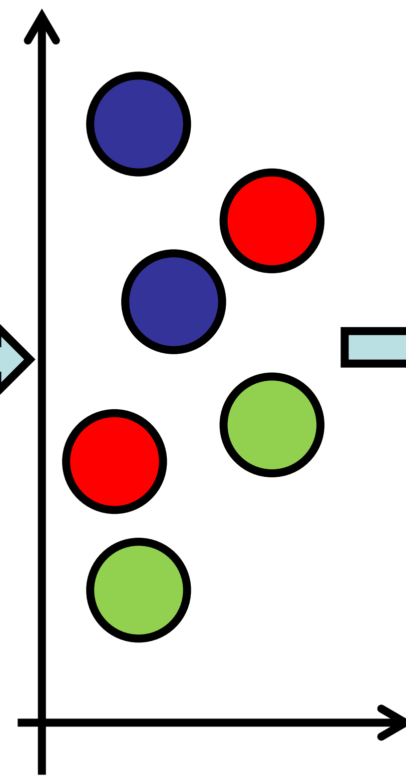
Auton Lab
www.autonlab.org
Computer vision, Robotics
EEG, fMRI, MEG, …
Astronomy
machine learning applications
Drug Discovery
AI in games
Neuroscience
Turbulences
ML in Agriculture
Microarray

Machine Learning on Complex Objects

Traditional Machine Learning

Observations → Feature vectors → Training data of feature vectors



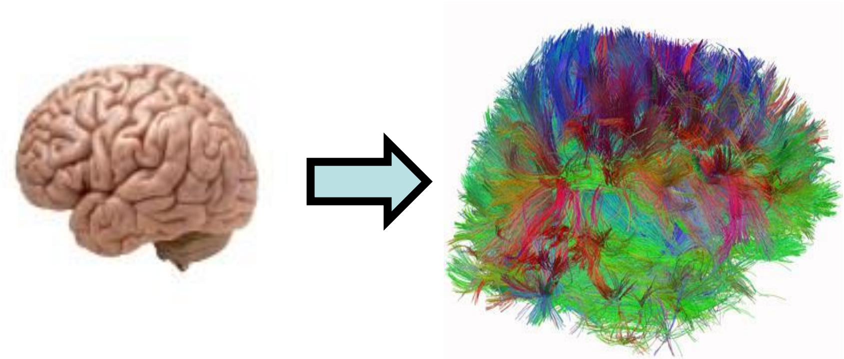


ML algorithm:
classification,
regression,
clustering, etc

Complex Data is Everywhere

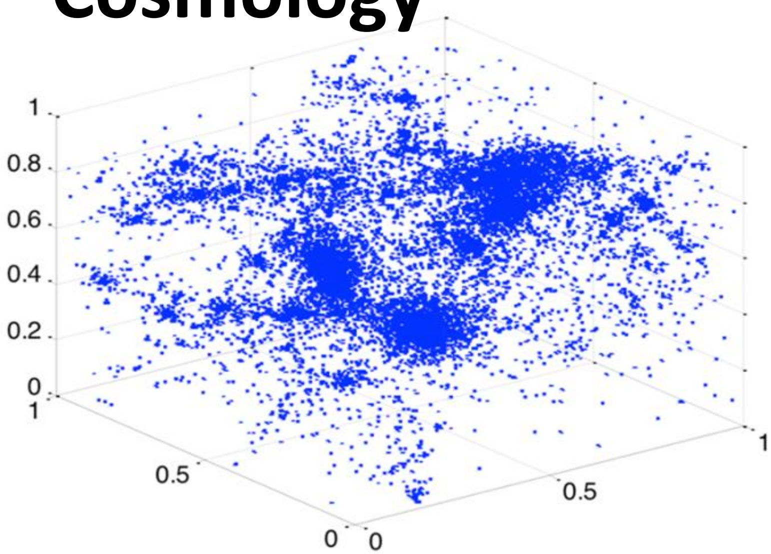
Finance

Neuroscience



Diffusion Weighted Imaging

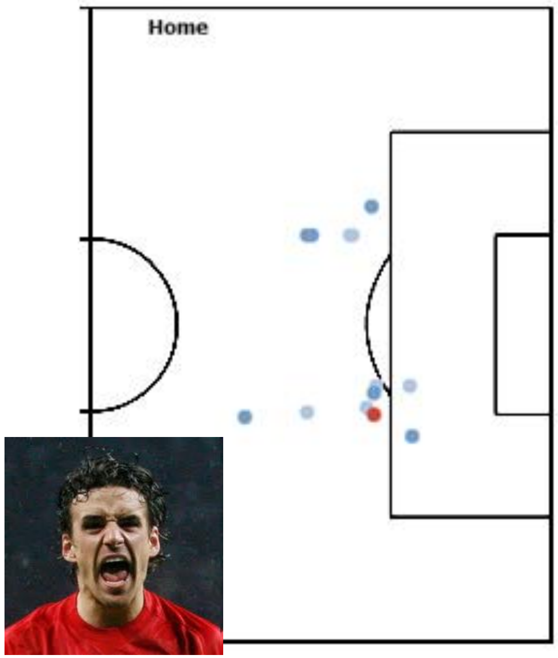
Cosmology



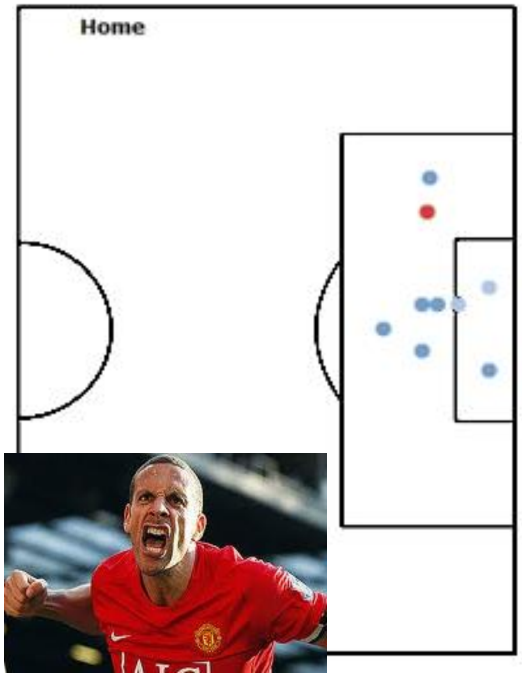
Images

Distributional Data

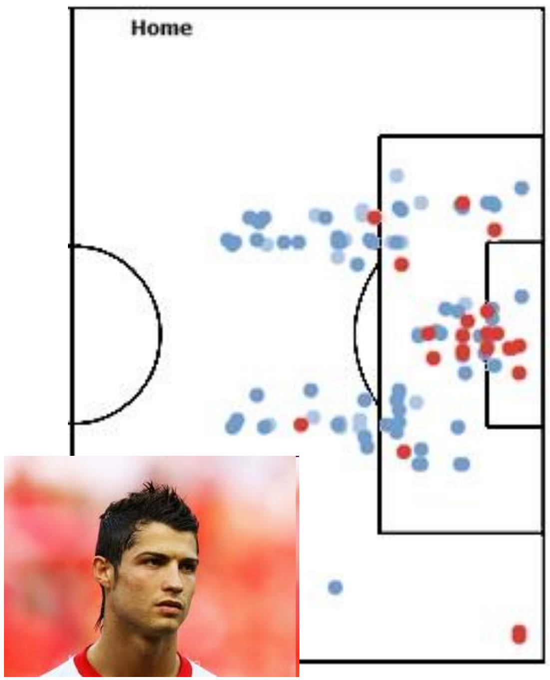
Manchester United 07/08



Owen Hargreaves



Rio Ferdinand



Cristiano Ronaldo

www.juhokim.com/projects.php

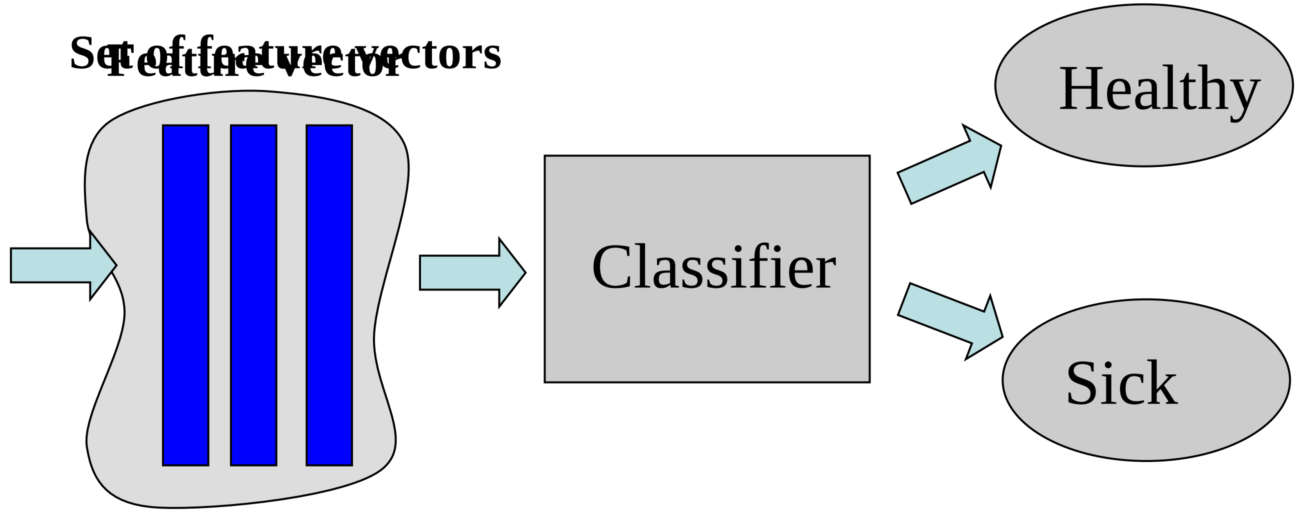
ML on Distributions

Standard machine learning

ML on sets/distributions

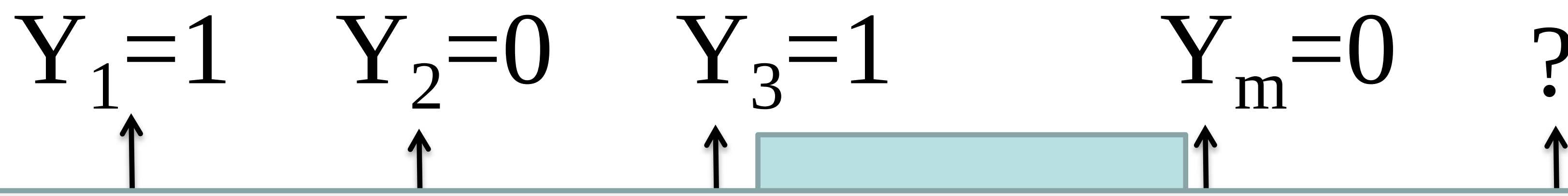
Medical tests:

blood pressure,
heart rate,
temperature,
blood sample
…



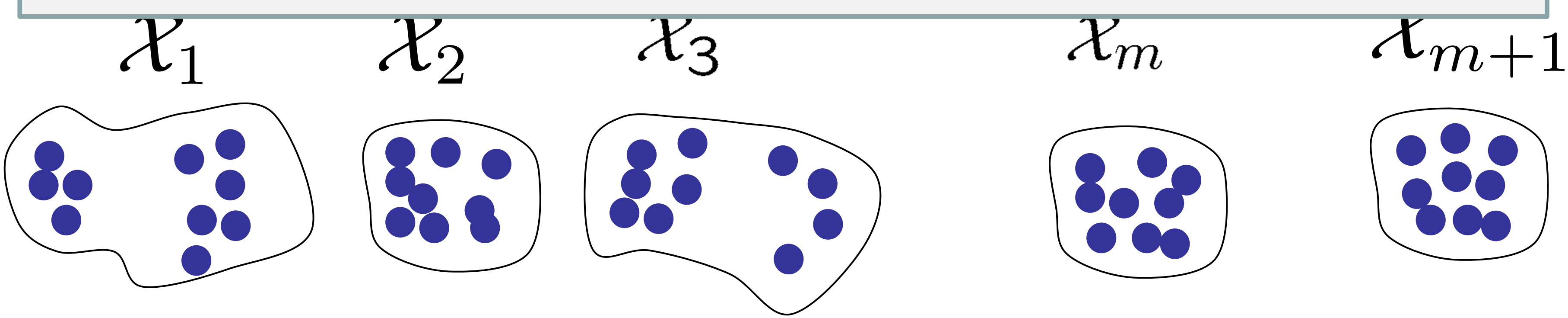
What happens if we repeat the medical tests?

Distribution Regression / Classification



Differences compared to standard methods on vectors

- The inputs are distributions, density functions (not vectors)
- We don't know these distributions, only sample sets are available (error in variables model)



Distribution Classification

We have T sample sets, $(\mathbf{X}_1, \ldots, \mathbf{X}_T)$. **[Training data]**
$\{X_{t,1}, \ldots, X_{t,m_t}\} = \mathbf{X_t} \sim p_t$. $\mathbf{X_t}$ has class $Y_t \in \{-1, +1\}$.

What is the class label Y of $\mathbf{X} = \{X_1, \ldots, X_m\} \sim p$?

Solution: Use RKHS based SVM!

Calculate the Gram matrix $K_{ij} \doteq \langle \phi(p_i), \phi(p_j) \rangle_K = K(p_i, p_j)$

$$\doteq \exp\left(-\frac{D(p_i, p_j)}{\sigma^2}\right)$$

Dual form of SVM:

$$\hat{\alpha} = \arg\max_{\alpha \in \mathbb{R}^T} \sum_{i=1}^{T} \alpha_i - \frac{1}{2} \sum_{i,j}^{T} \alpha_i \alpha_j y_i y_j K_{ij},$$

subject to $\sum_i \alpha_i y_i = 0$, $0 \leq \alpha_i \leq C$.

$$Y = \text{sign}\left(\sum_{i=1}^{T} \hat{\alpha}_i y_i K(p_i, p)\right) \in \{-1, +1\}$$

Problems: We do not know p_i, p, $K(p_i, p_j)$, or $K(p_i, p)$...

Distances / Divegences between Distributions

Euclidean: $D(p,q) = (\int (p(x) - q(x))^2 dx)^{1/2}$

Kullback-Leibler: $D(p,q) = KL(p,q) = \int p(x) \log \frac{p(x)}{q(x)} dx$

Renyi: $D(p,q) = R_\alpha(p\|q) = \frac{1}{\alpha-1} \log \int p^\alpha q^{1-\alpha}$

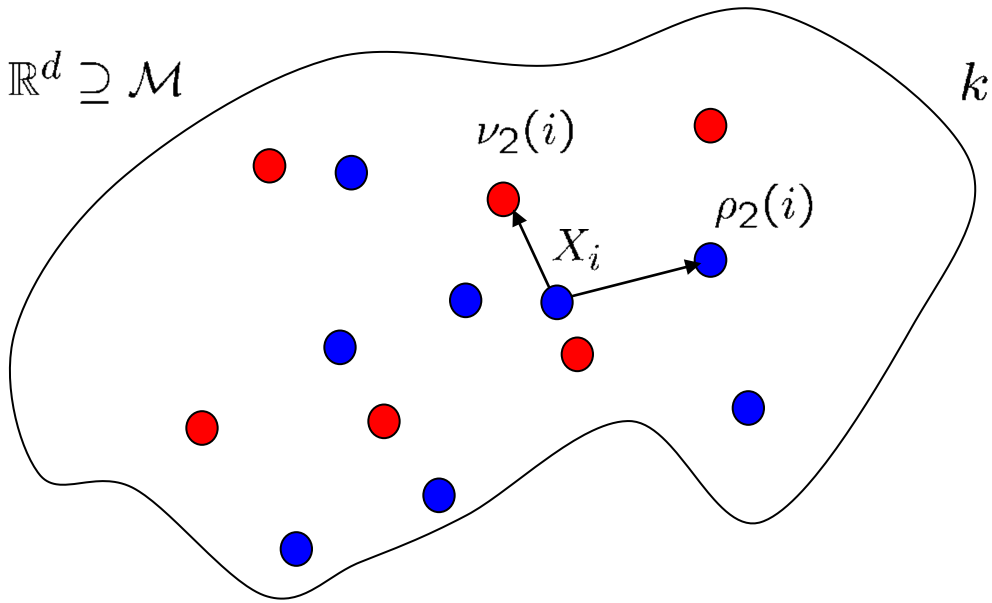
RÉNYI DIVERGENCE ESTIMATION

without density estimation

Using $X_{1:n} = \{X_1, \ldots, X_n\} \sim p$ $Y_{1:m} = \{Y_1, \ldots, Y_m\} \sim q$

Estimate divergence $R_\alpha(p\|q) \doteq \frac{1}{\alpha - 1} \log \int p^\alpha q^{1-\alpha}$

The Estimator



$k = 2.$

$k \geq 1$, fixed.

$\rho_k(i)$: the distance of the k-th nearest neighbor of X_i in $X_{1:n}$

$\nu_k(i)$: the distance of the k-th nearest neighbor of X_i in $Y_{1:m}$

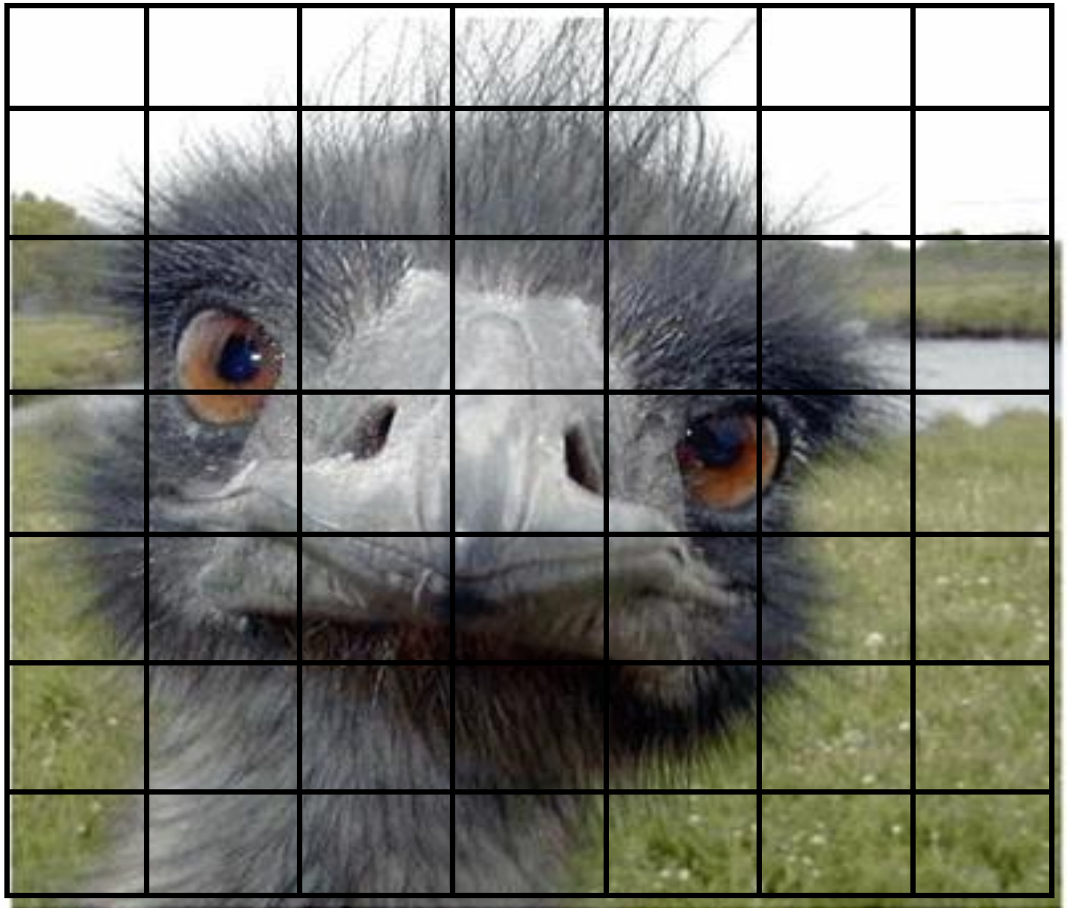
$$D_\alpha(p\|q) \doteq \int p^\alpha q^{1-\alpha}$$

$$\widehat{D}_\alpha(X_{1:n}\|Y_{1:m}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(n-1)\rho_k^d(i)}{m\nu_k^d(i)}\right)^{1-\alpha}\frac{\Gamma(k)^2}{\Gamma(k-\alpha+1)\Gamma(k+\alpha-1)}$$

Auton Lab

Carnegie Mellon

ML on Distributions



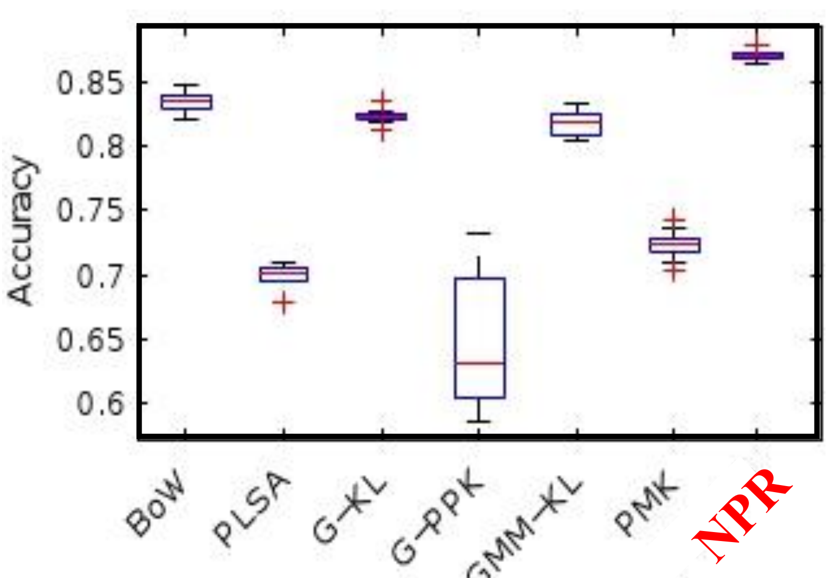
Dealing with complex objects

- break into smaller parts, represent the input as a **set** of smaller parts
- treat the set elements as sample points from some **unknown distribution**
- do ***ML on these unknown distributions*** represented by sets

Sport Events Classification
[Li and Fei Fei, 2007]

badminton bocce croquet polo climbing rowing sailing snowboard

8 categories, 1040 images, each represented by 295 to 1542 57 dim points.



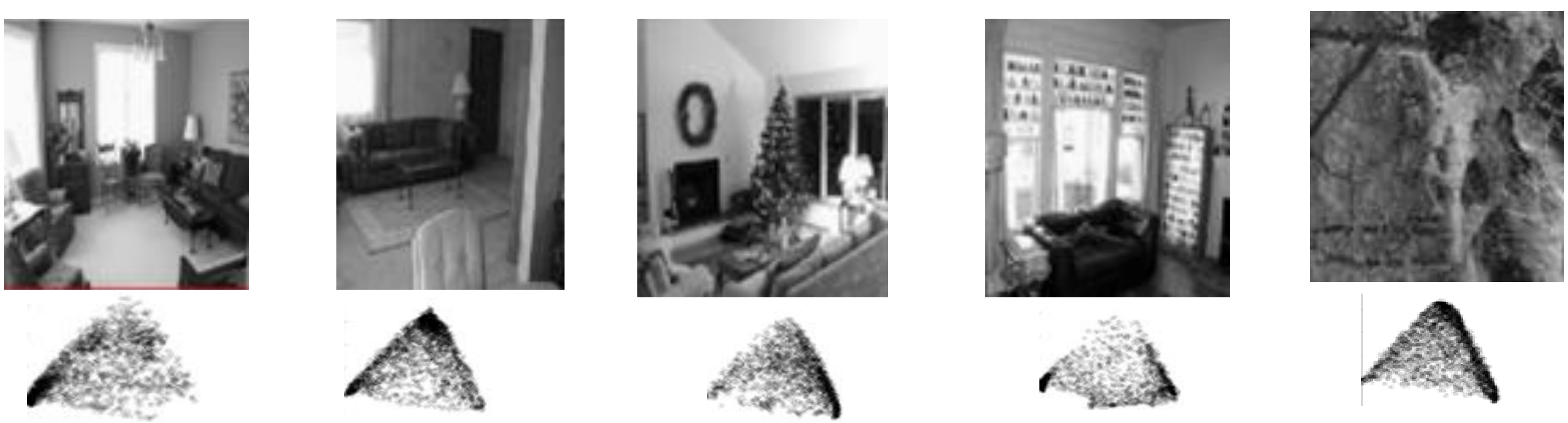
- Best published: **86.7%**
  (Zhang et al, CVPR 2011)
- NPR: **87.1%**

2 fold CV, 16 runs

Detecting Anomalous Images

50 highway images

5 anomalies

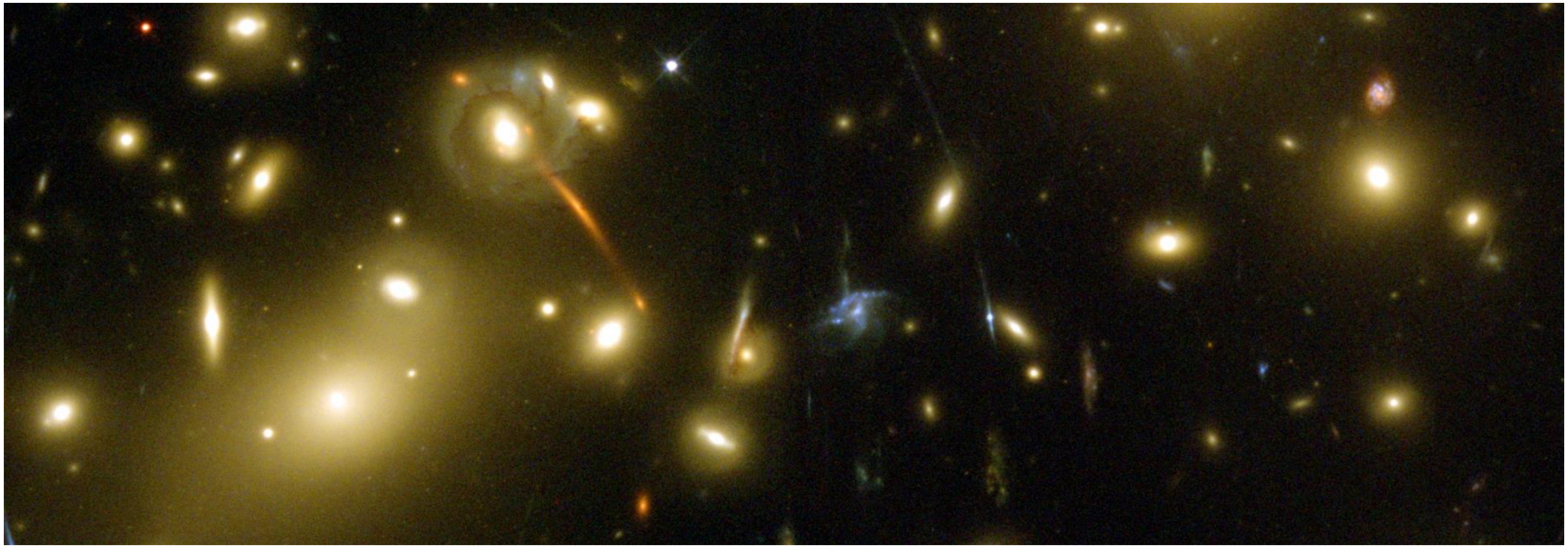


2-dimensional sample set representation of images (128 dim SIFT ⇒ 2 dim)

Anomaly score: divergences between the distributions of these sample sets

Cosmology Applications

Find new scientific laws in physics



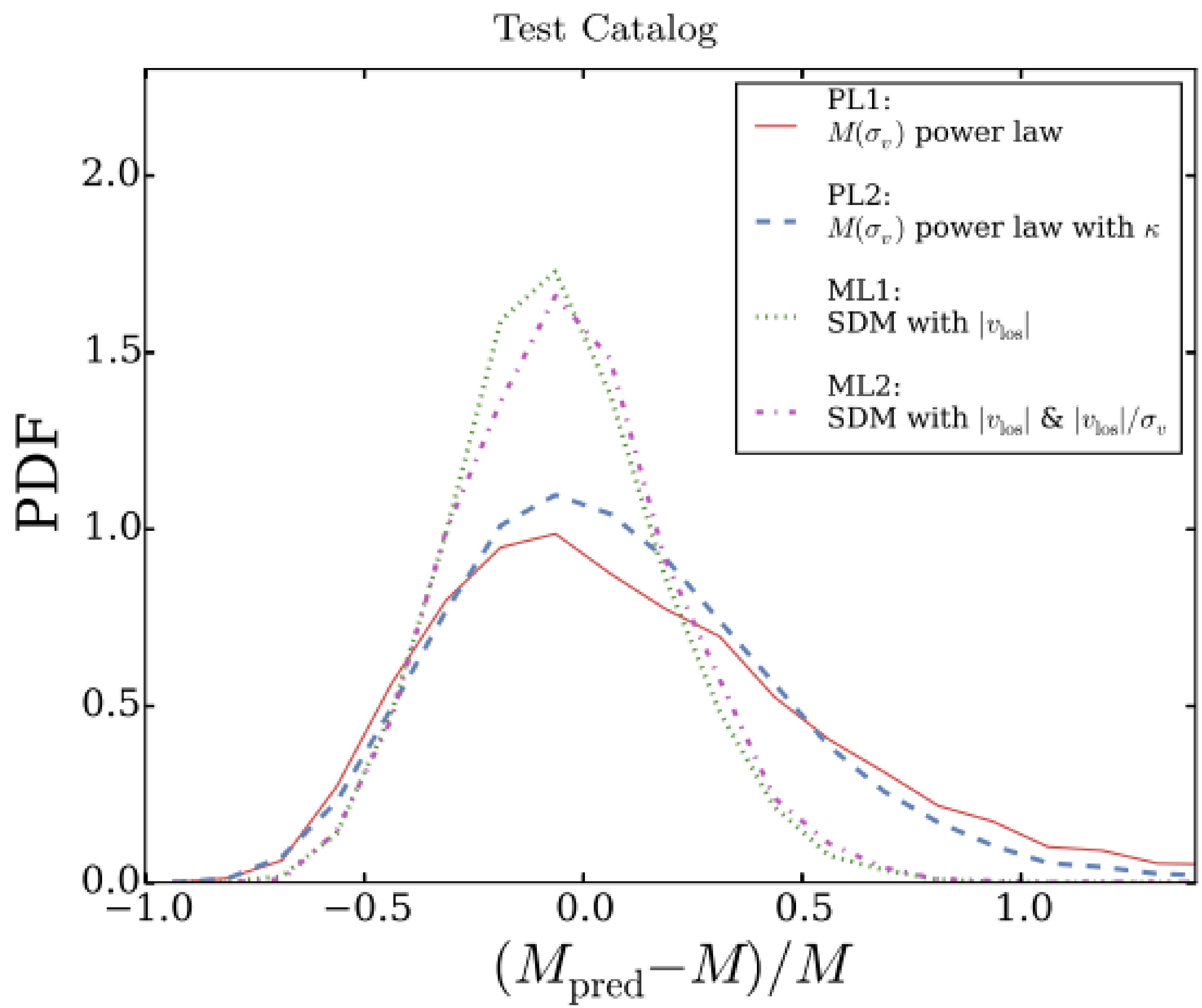
Goal: Estimate dynamical mass of galaxy clusters.

Importance: Galaxy clusters are being the largest gravitationally bound systems in the Universe. Dynamical mass measurements are important to understand the behavior of dark matter and normal matter.

Difficulty: We can only measure the velocity of galaxies not the mass of their cluster. Physicists estimate dynamical cluster mass from single velocity dispersion.

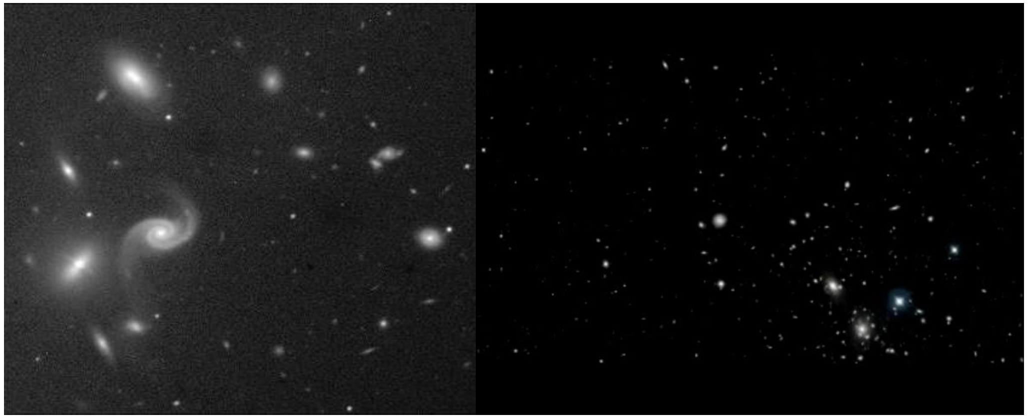
Our method: Estimate the cluster mass from the whole distribution of velocities rather than just a simple velocity distribution.

Find new scientific laws in physics



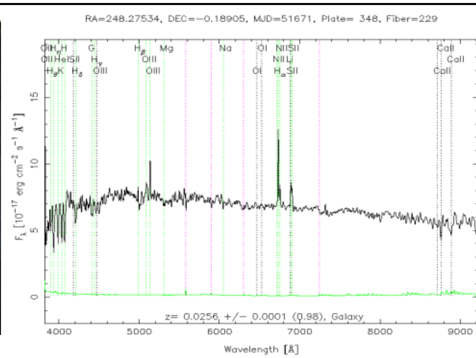
Michelle Ntampaka et al, A Machine Learning Approach for Dynamical Mass Measurements of Galaxy Clusters, APJ 2015

Find interesting Galaxy Clusters

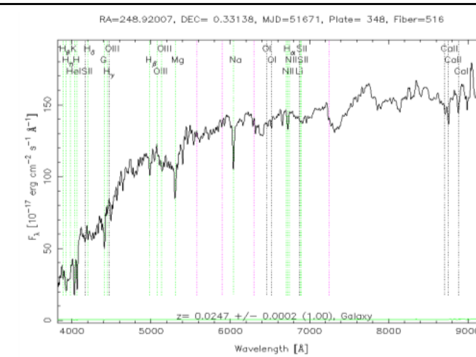


Sloan Digital Sky Survey (SDSS)

- ❑ continuum spectrum
- ❑505 galaxy clusters (10-50 galaxies in each)
- ❑7530 galaxies



Blue galaxy



Red galaxy

What are the most anomalous galaxy clusters?

The most anomalous galaxy cluster contains mostly

- ❑ star forming blue galaxies
- ❑ irregular galaxies

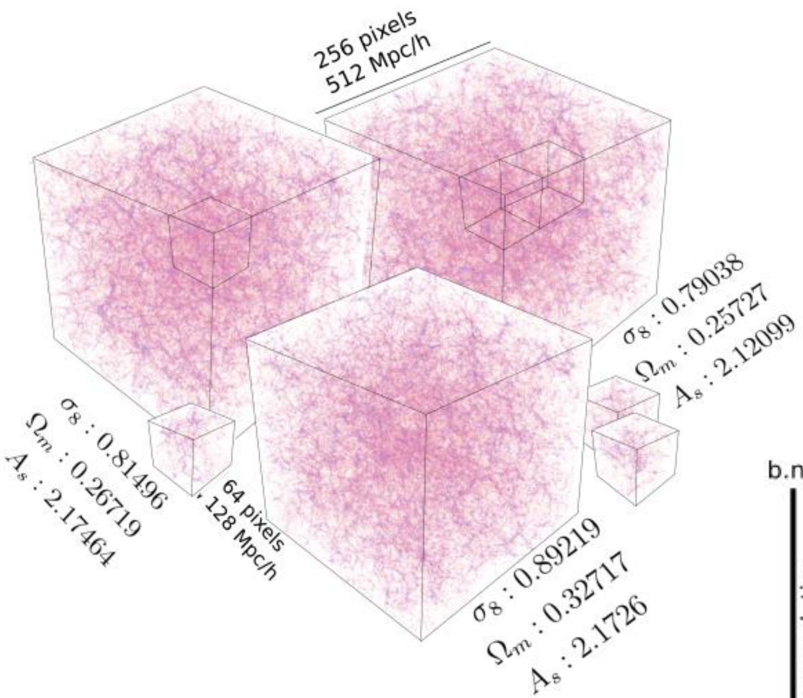
Auton Lab

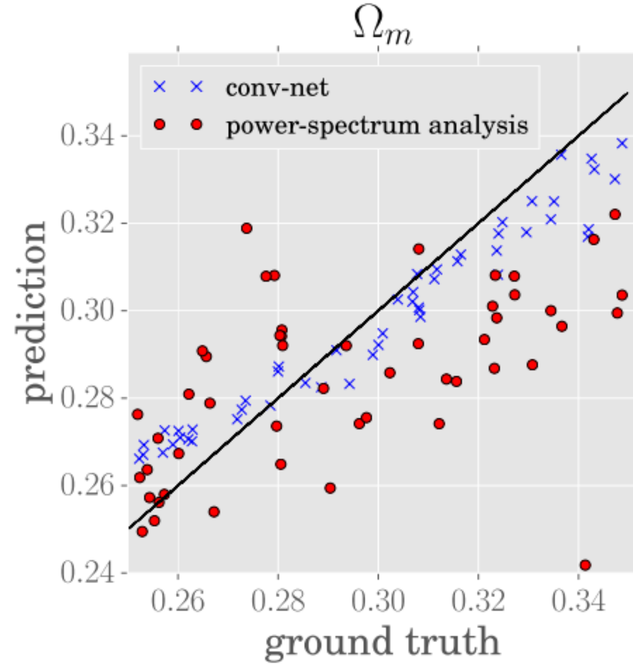
B. Póczos, L. Xiong & J. Schneider, UAI, 2011.

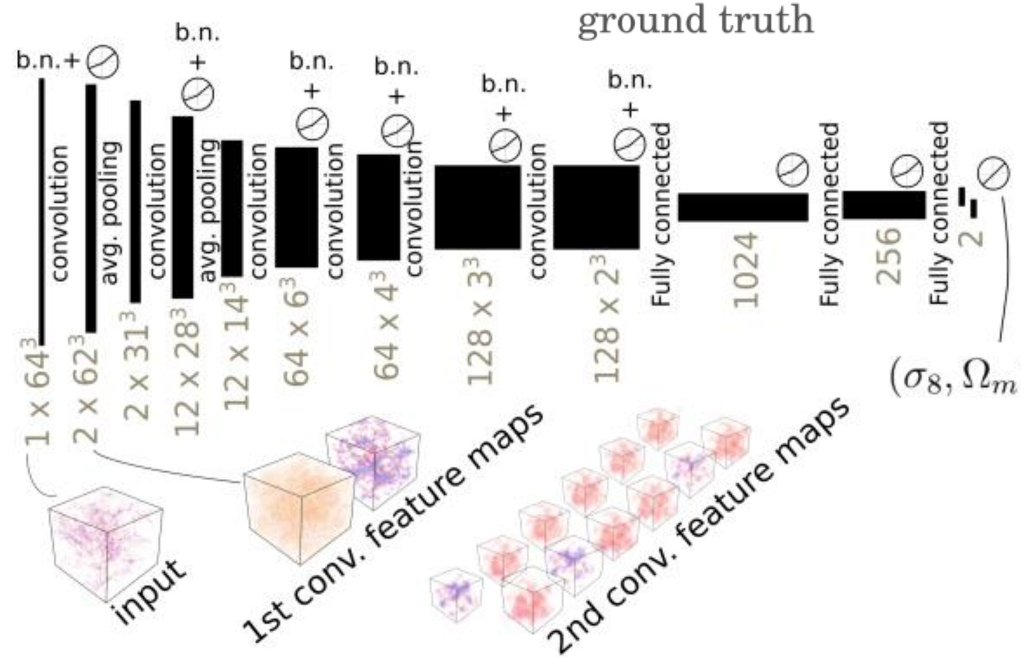
Credits: ESA, NASA Carnegie Mellon

Find the parameters of Universe

Given a distribution of particles, our goal is to predict the parameters of the simulated universe



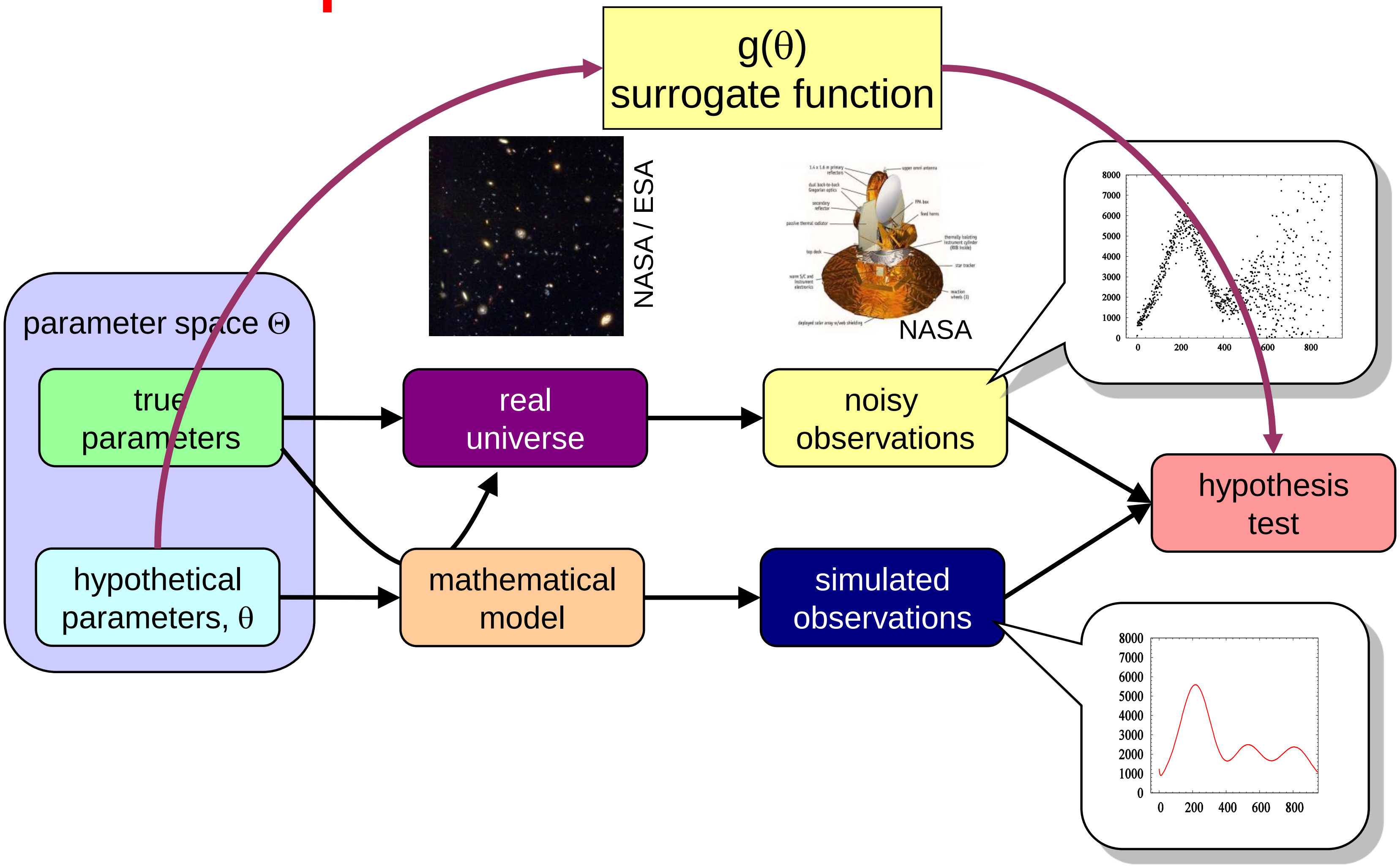




Auton Lab

Active Learning & Design Optimization

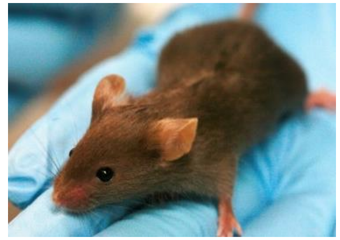
Recommend experiments to find the true parameters of the universe

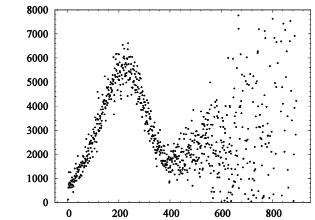


Computation problem: How to search parameter space

Solution: Learn a surrogate function and make experiment decisions using it

Recommend experiments for drug discovery





parameter space Θ

Parameters of Drugs:

- Compounds
- Quantities
- etc.

Drug effects on the Lab mouse

Expensive Observations ($, time)

- Real Observations
- Simulated Observations
- Expert predictions

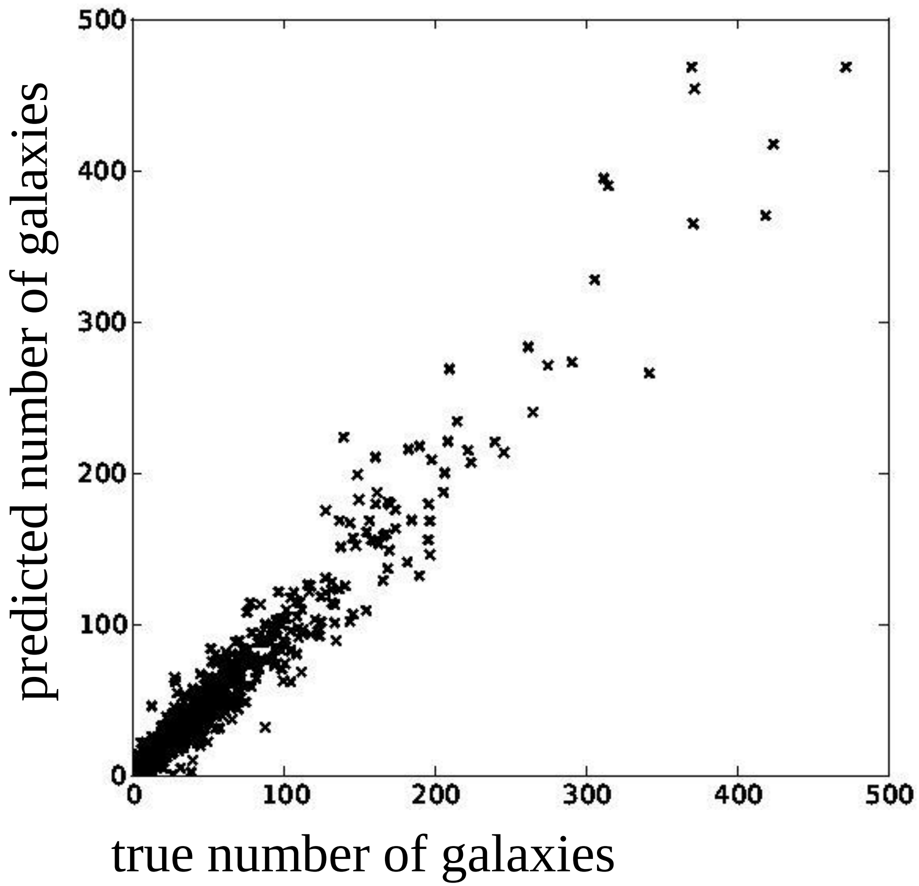
(Bias, Variance, Fidelity, Cost)

Observations:

- Blood samples
- Camera images
- EEG,
- etc.

g(θ) surrogate function to optimize

Learning Relationships from Simulations



[Xiaoying Xu, 2012]

Goal: predict the number of galaxies in a halo from a half dozen dark matter halo parameters

(#particles in a halo, velocity dispersion, max circular velocity, half mass radius,…)

data: Millenium simulation 395,832 halos

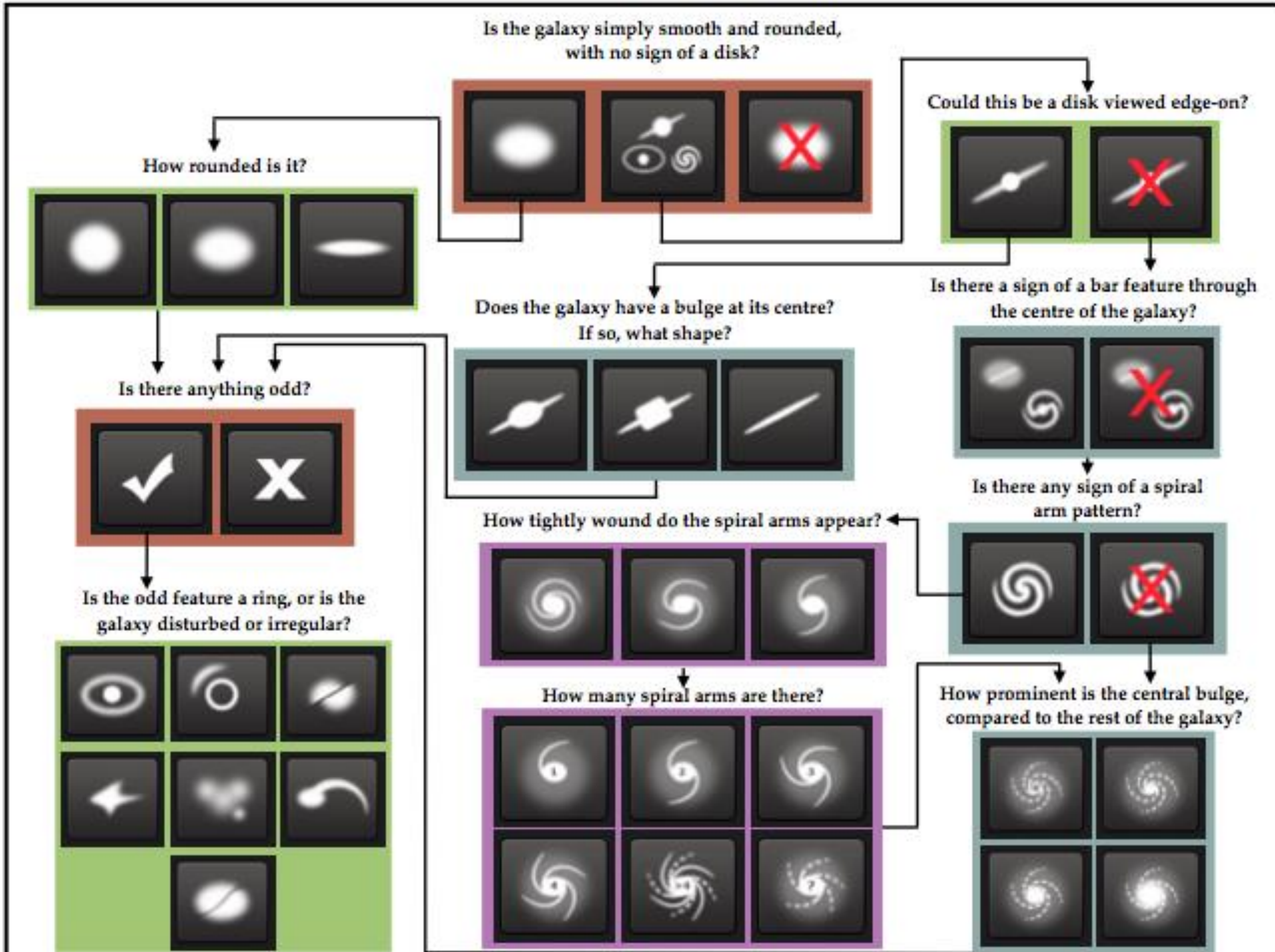
method: support vector regression

The Galaxy Zoo challenge

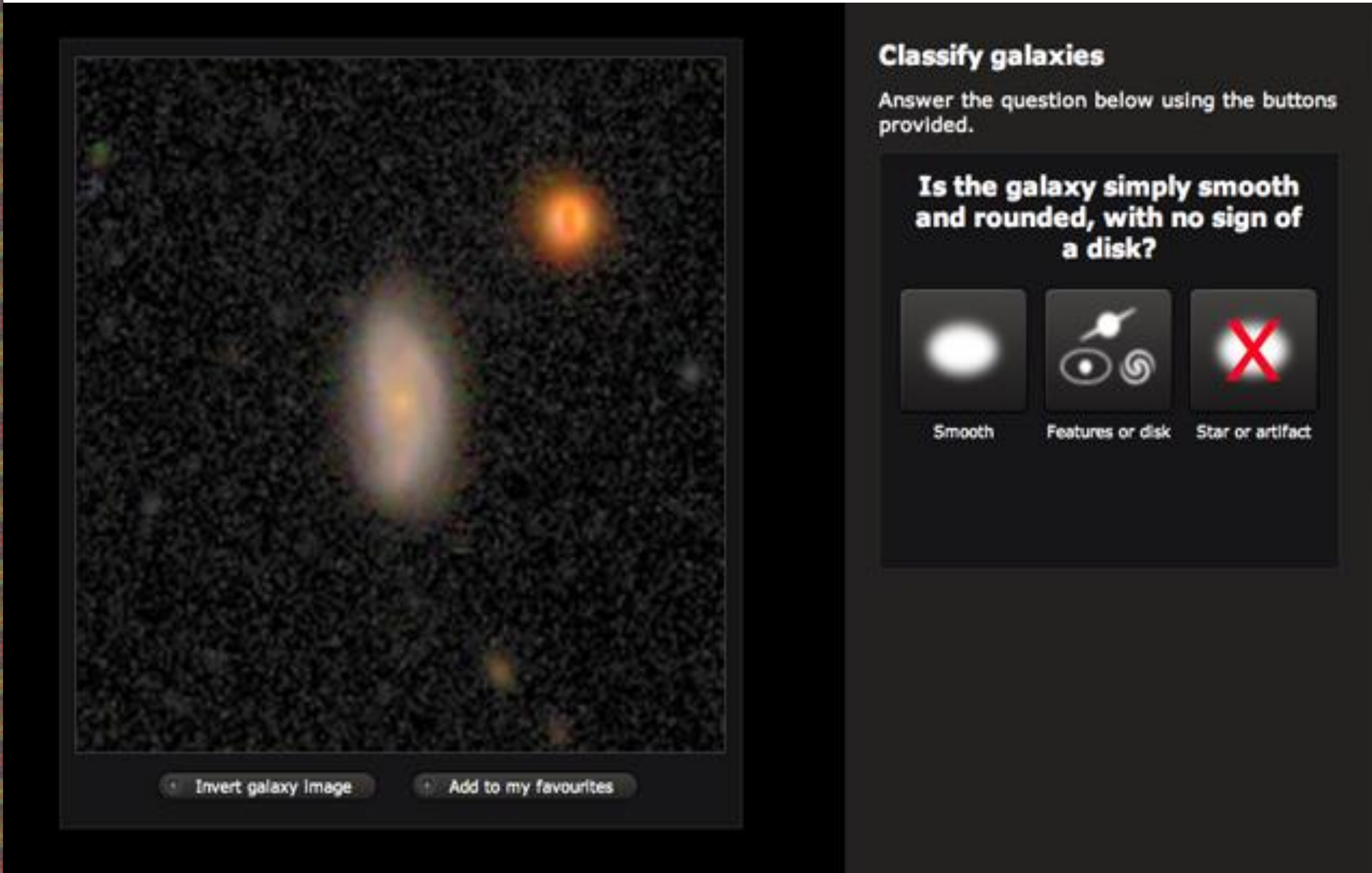
- Crowdsourcing project

- Users are asked to describe the morphology of galaxies based on images.
- They are asked questions such as "How rounded is the galaxy" and "Does it have a central bulge"…
- 37 different categories in a decision tree

- Training set: JPG images of 61578 galaxies.
- Test set: JPG images of 79975 galaxies
- Image resolution: 424x424 color JPEG images



Willett et al. 2013.


Classify galaxies
Answer the question below using the buttons provided.
Is the galaxy simply smooth and rounded, with no sign of a disk?
Smooth
Features or disk
Star or artifact
Invert galaxy image
Add to my favourites

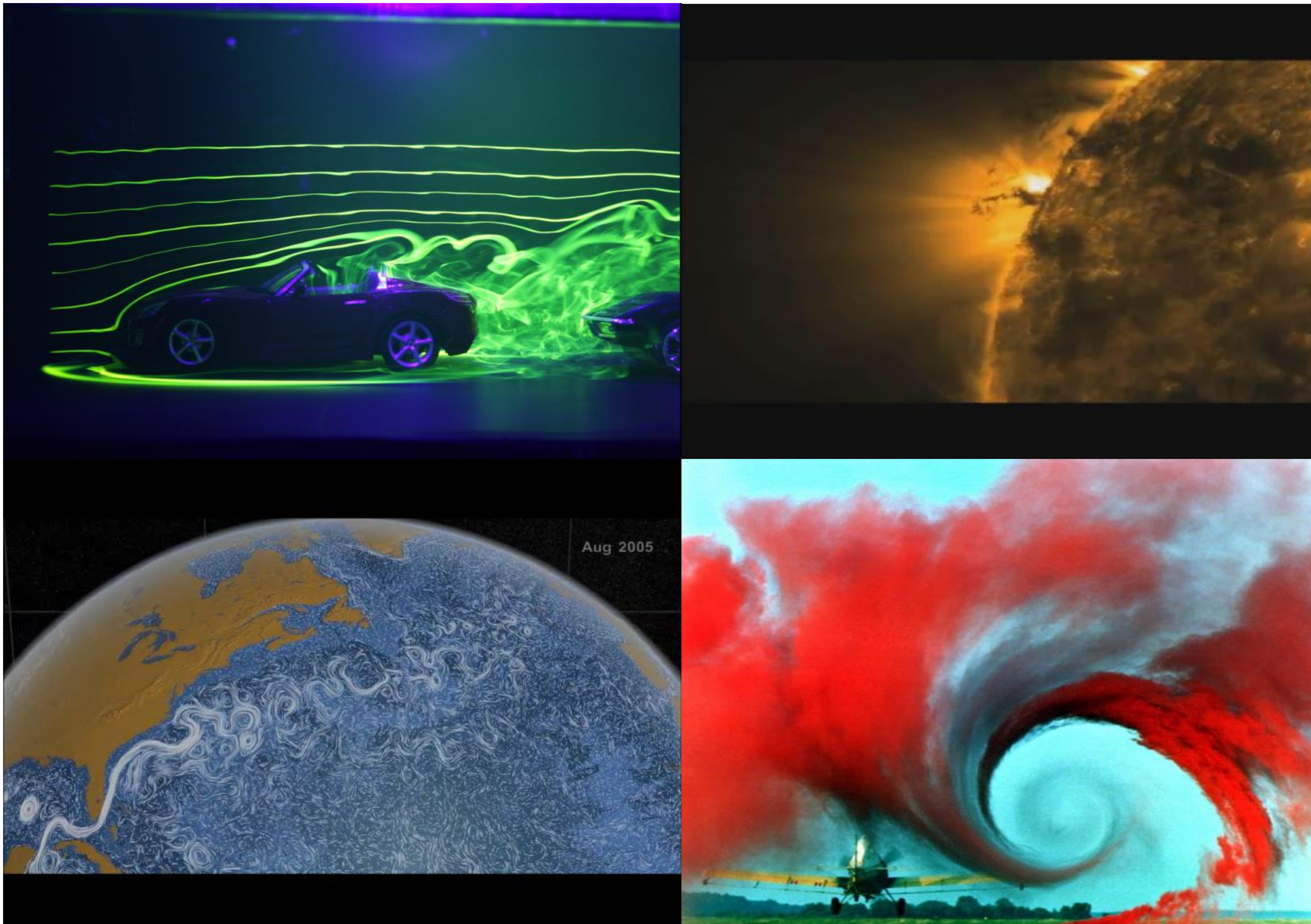
The Large Synoptic Survey Telescope

Big data questions

- 15 Terabytes of data … every night

Other Examples in Physics

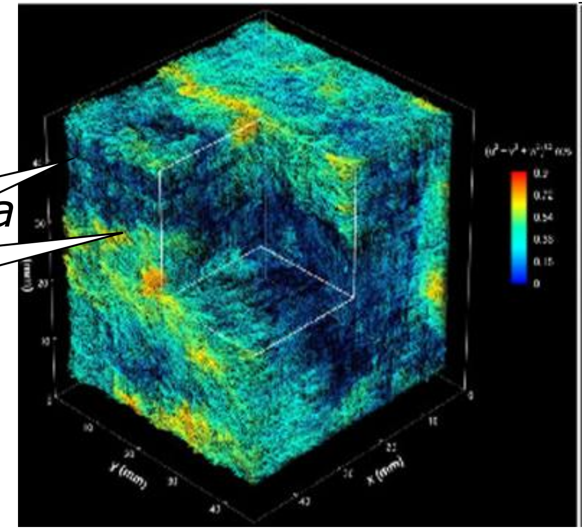
ML to Help Understanding Turbulences



Turbulence Data Classification

Simulated fluid flow through time

(JHU Turbulence Research Group, Alex Szalay)



Goal: *find 'interesting' events, ... ena*

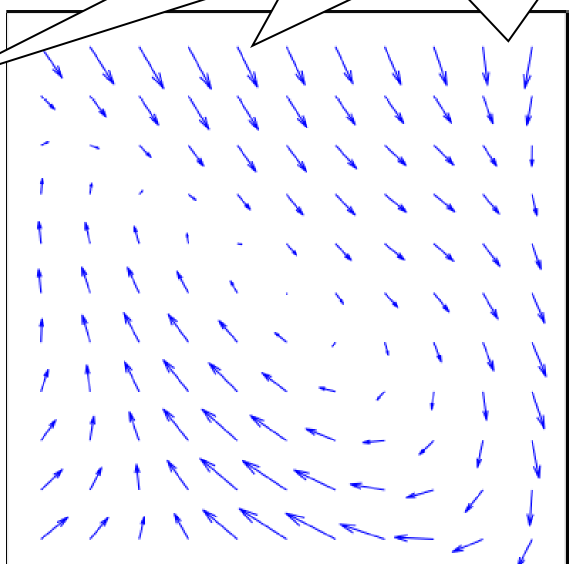
find vortices

Something interesting happened?

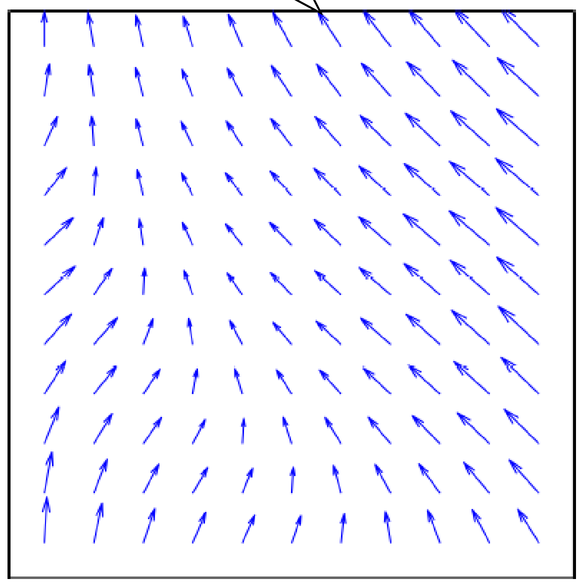
•11 positi...

•**Results:** Leave one out cross-validation : 97%

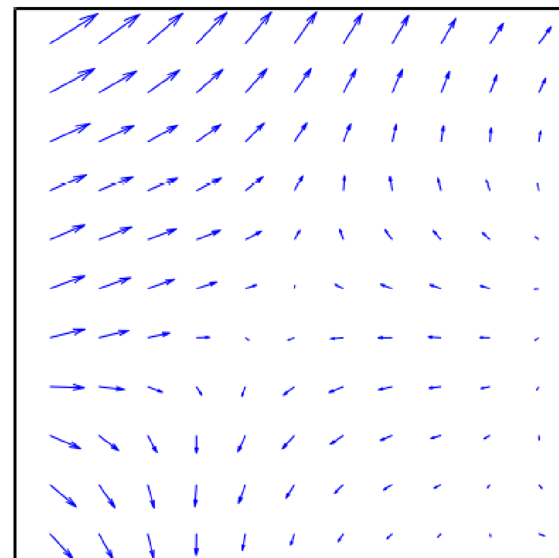
...ons



Positive (vortex)



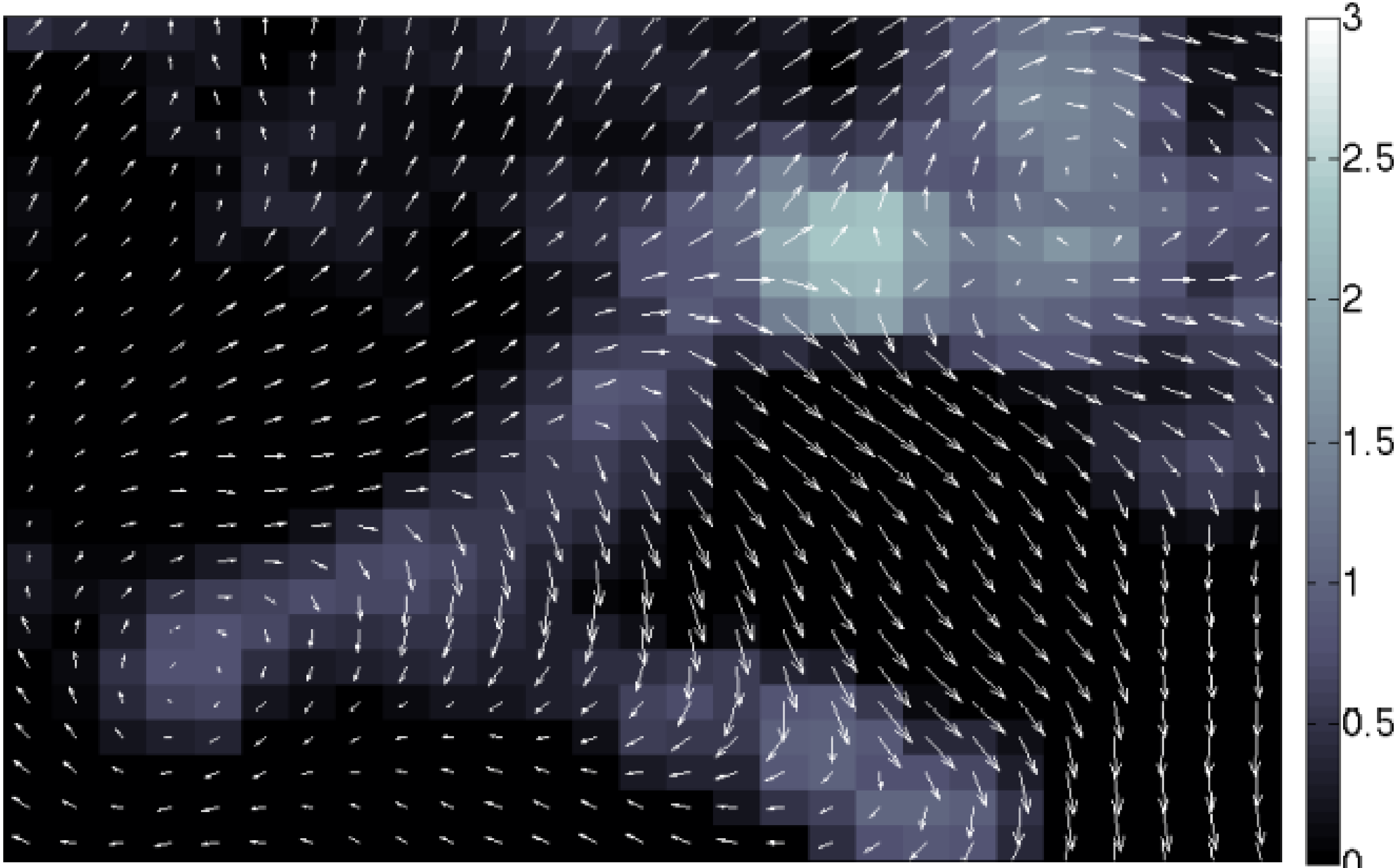
Negative



Negative

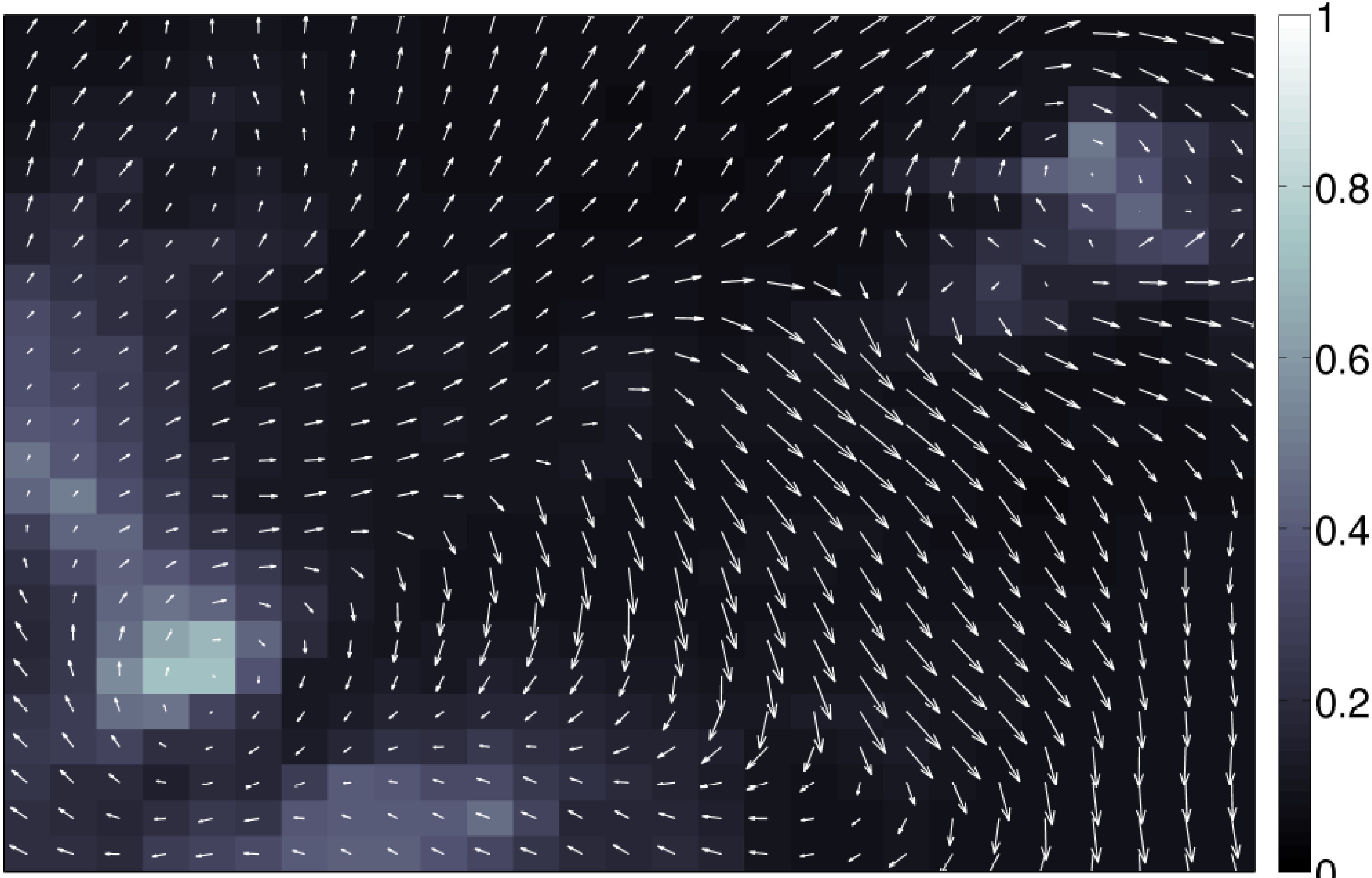
Find Interesting Phenomena in Turbulence Data

Anomaly detection



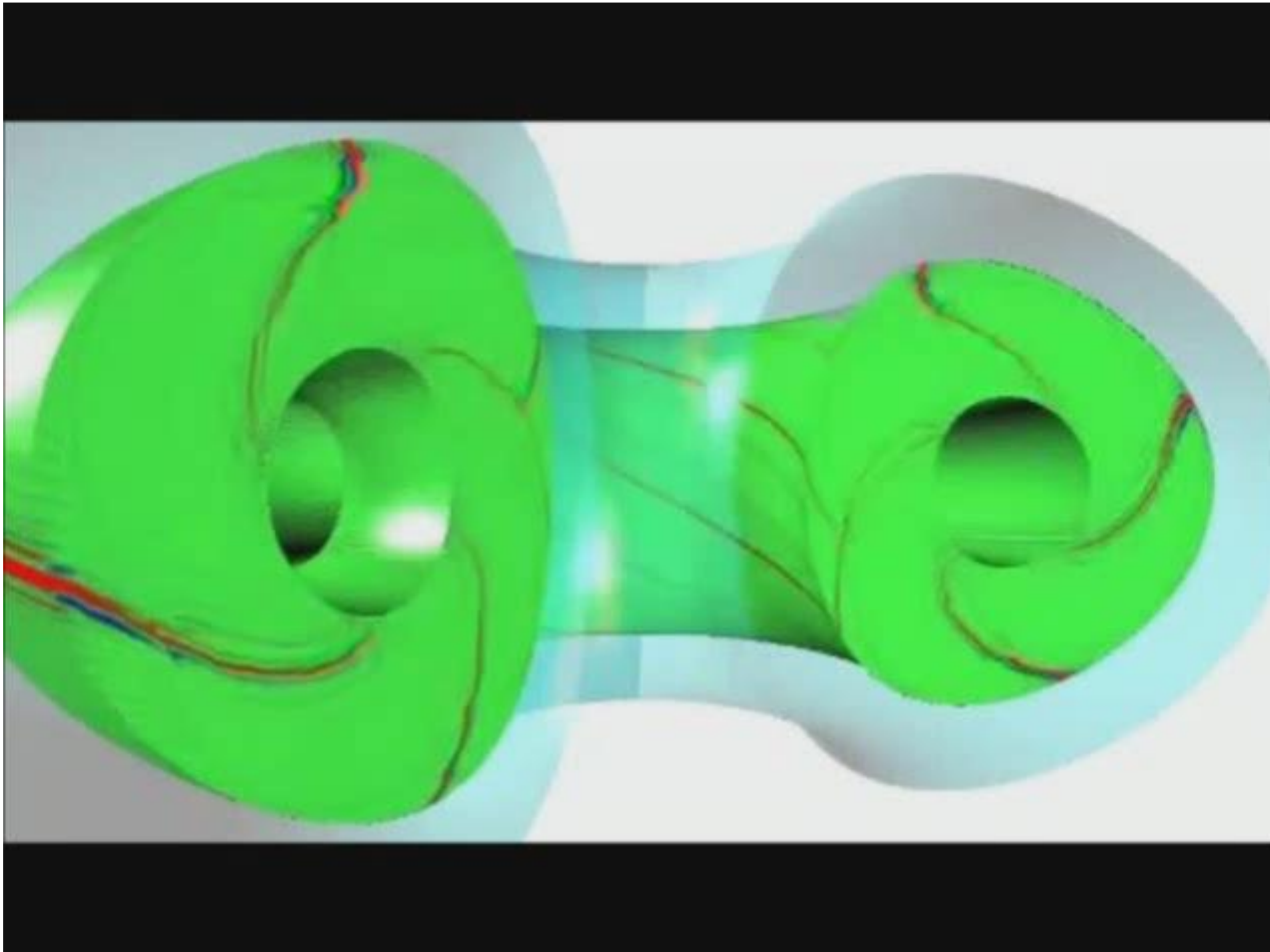
Anomaly scores

Finding Vortices



Classification probabilities

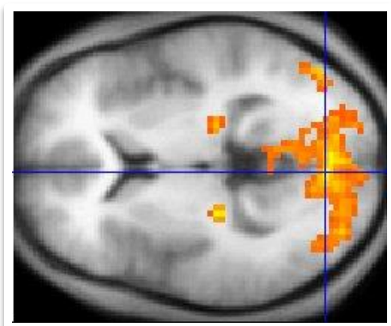
Fusion power plants



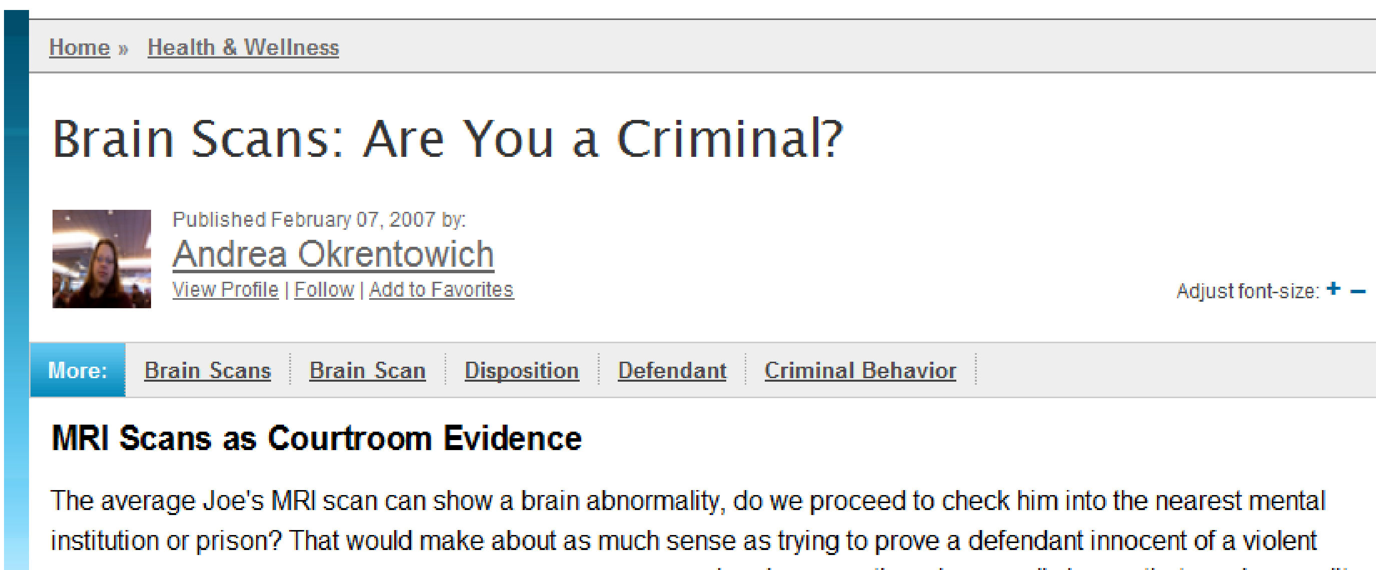
Neuroscience

ML in Action: Neuroimaging

- MEG/ fMRI mind reading contest
- MRI lie detector
- Decoding thoughts from brain scans



Rob a bank …



Home » Health & Wellness

Brain Scans: Are You a Criminal?

Published February 07, 2007 by:
Andrea Okrentowich
View Profile | Follow | Add to Favorites

Adjust font-size: + –

More: Brain Scans | Brain Scan | Disposition | Defendant | Criminal Behavior

MRI Scans as Courtroom Evidence

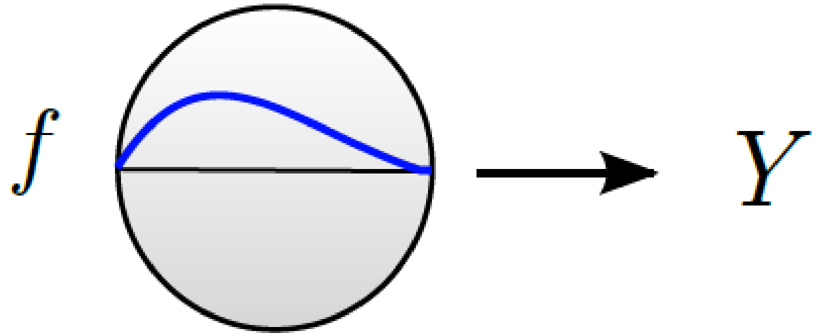
The average Joe's MRI scan can show a brain abnormality, do we proceed to check him into the nearest mental institution or prison? That would make about as much sense as trying to prove a defendant innocent of a violent

Carnegie Mellon

FuSSO = Functional Shrinkage and Selection Operator
(Functional Lasso)

37

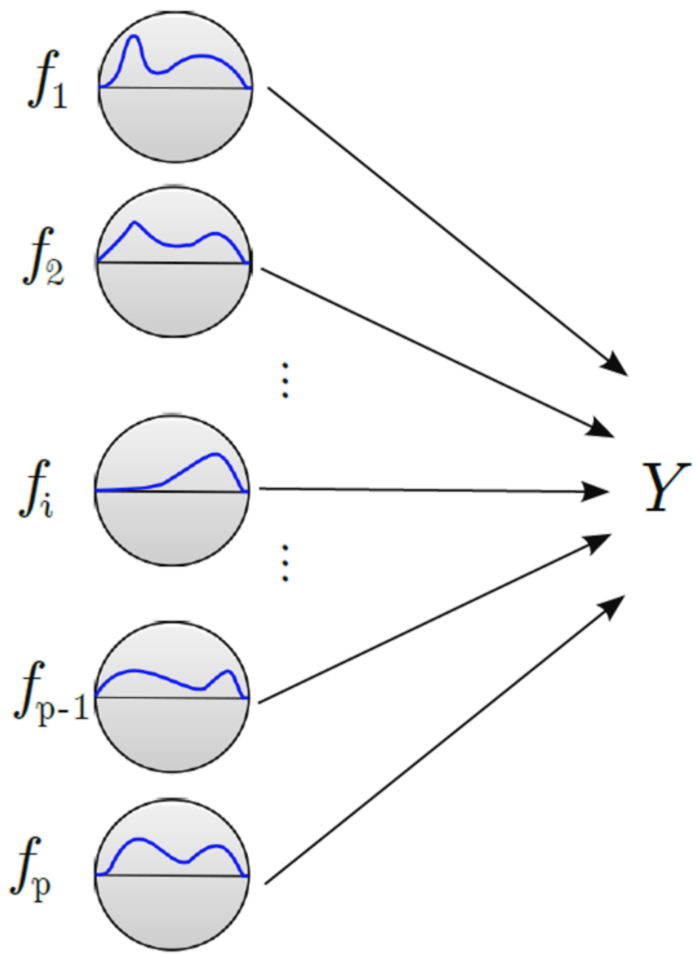
Function-to-Real regression



38

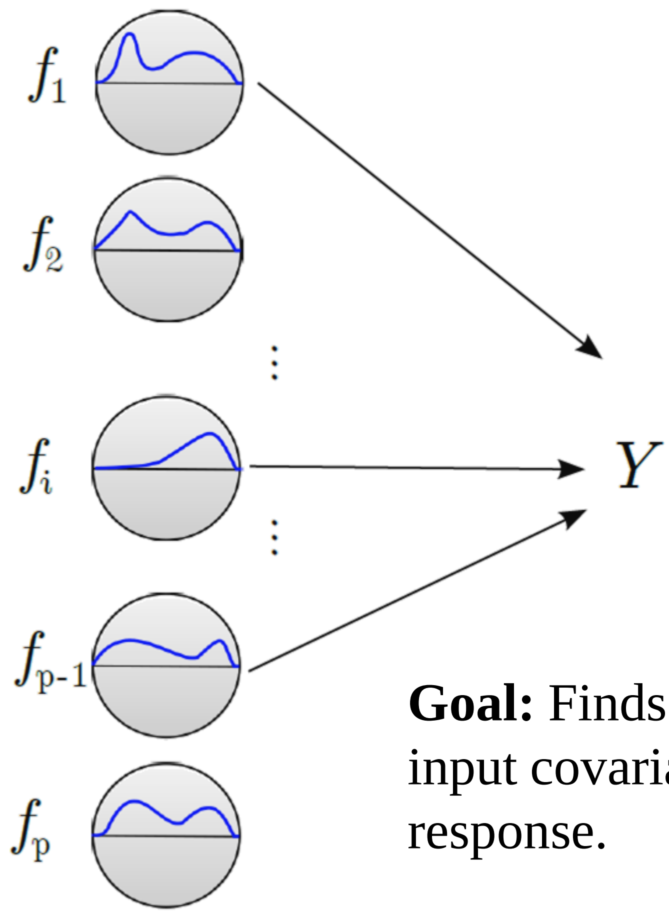
Many Functions-to-Real regression

Similarly, one may consider a mapping that takes in multiple functions:



Sparse Functions-to-Real regression

When the number of functional input covariates may be very large, a sparse model that depends only on a few of the functional covariates may be preferred:



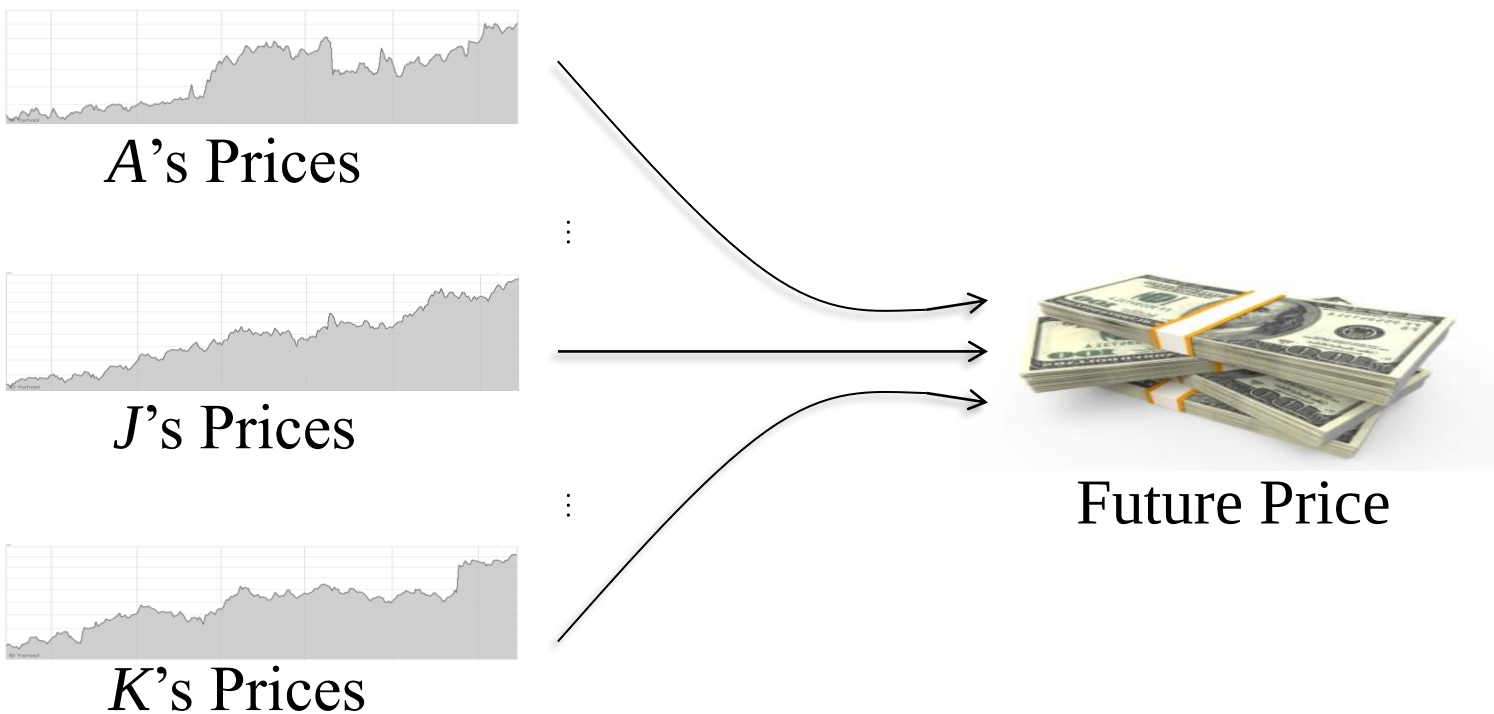
Goal: Finds a sparse set of functional input covariates to predict a real-valued response.

FuSSO Example Applications

Finance:

Inputs: Time-series of several product prices in the past

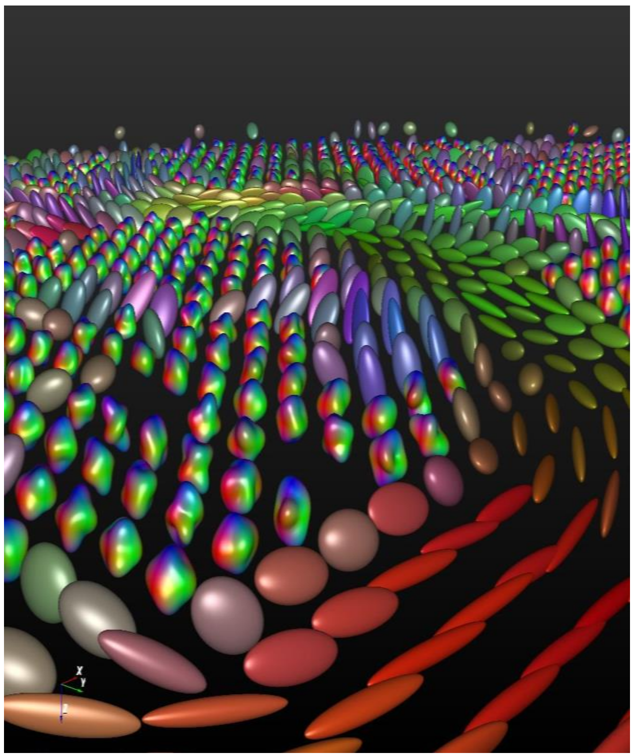
Output: Price of a particular product in the nearby future



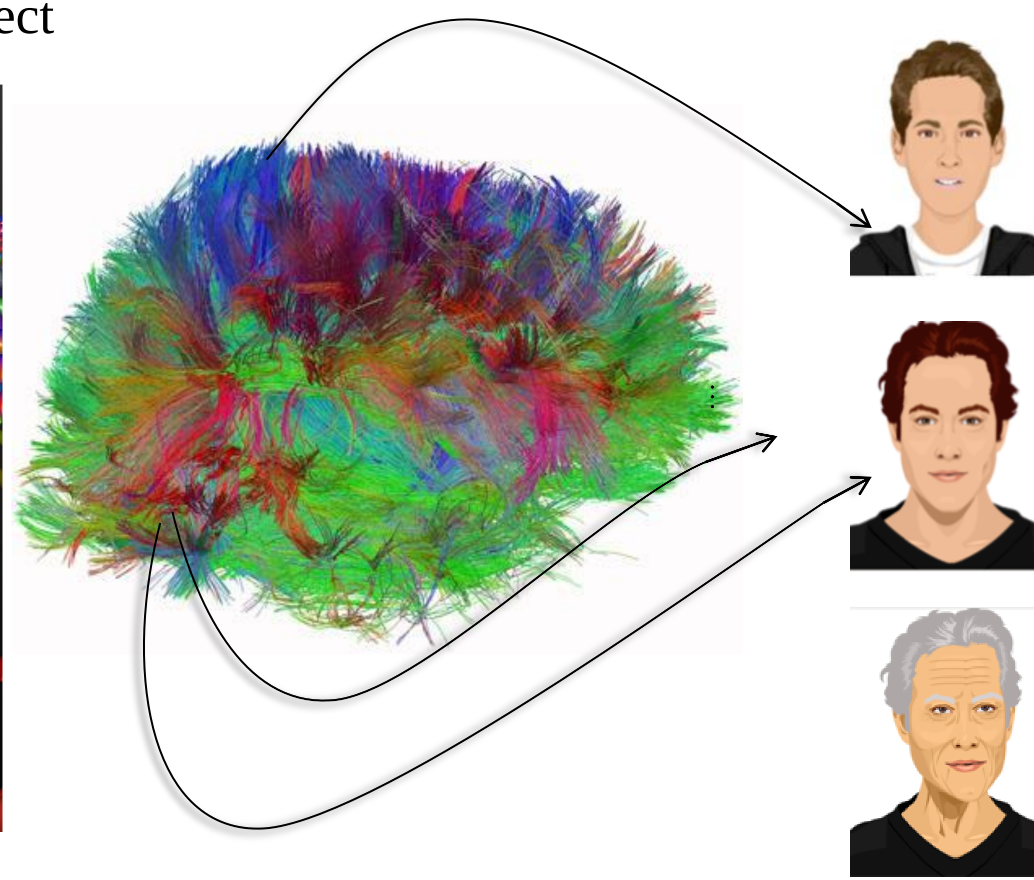
FuSSO Applications in Neuroimaging

Inputs: Functions at each voxel (e.g. orientation distribution functions)

Output: The age of the subject



Voxels' ODFs



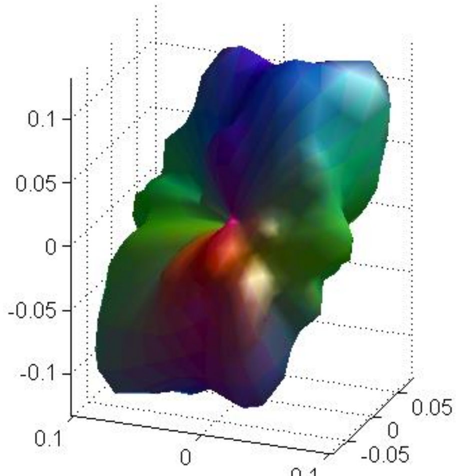
Age

Image credit: http://bmia.bmt.tue.nl/software/viste/

Results: Neuroimaging dataset

- Dataset with over 25K functions per subject for 89 total subjects (18 to 60 years old)
- Orientation distribution functions (ODF) at white matter voxels
- **Goal**: Predict the subject's age, given ODFs
- We compared to LASSO with peak ODF (quantitative anisotropy, QA) values. Finite dim non-functional data set.

Example Voxel ODF



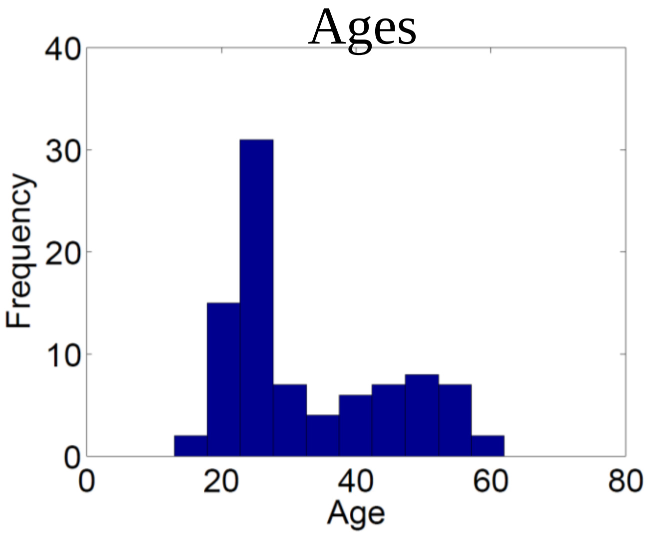


Image Sources: http://www.aging2.com/wp-content/uploads/2013/05/Screen-Shot-2013-05-28-at-9.48.49-PM.png; http://media.salon.com/2013/02/money1.jpg; http://3278as3udzze1hdk0f2th5nf18c1.wpengine.netdna-cdn.com/wp-content/uploads/2010/10/connectome-brain-diffusion-spectrum-imaging.jpg

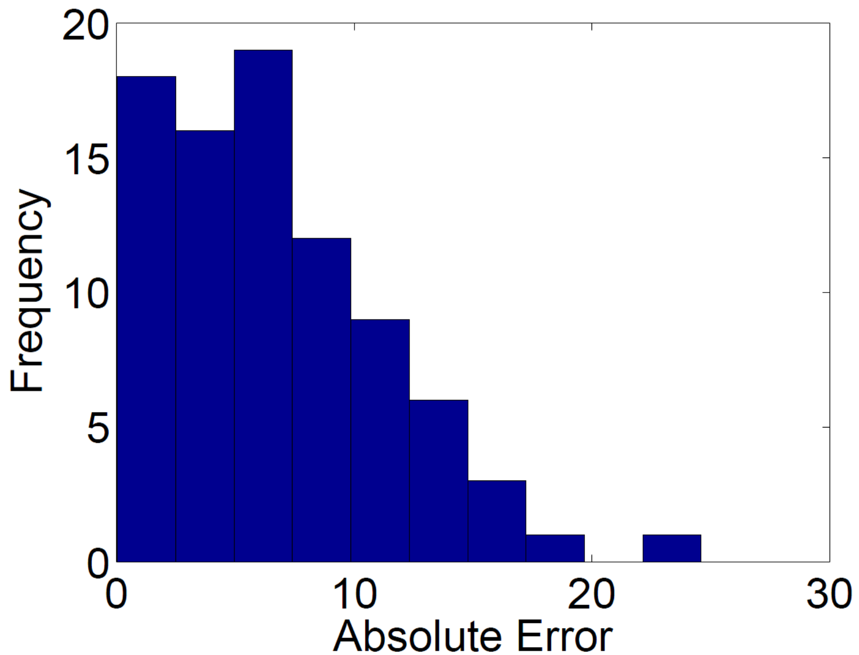
43

Results: Neuroimaging dataset

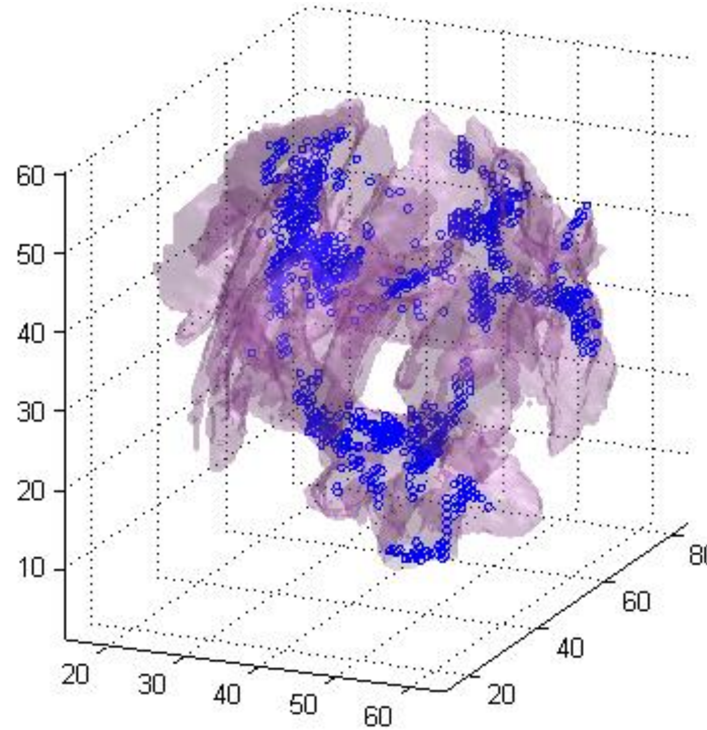
Results:

Method:	FuSSO (ODFs)	LASSO (QAs)	Mean Predict
MSE:	70.85	77.13	156.43

Absolute Errors per Subject



Selected Voxels

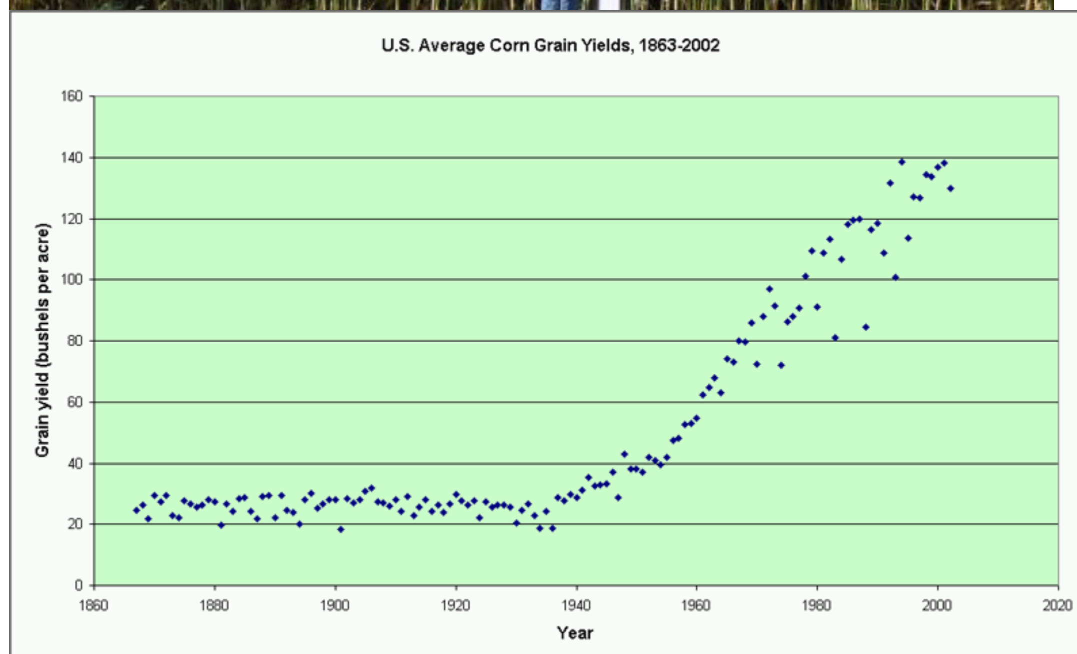


Mean error: 8.3 years, Naïve approach error: 12.5 years

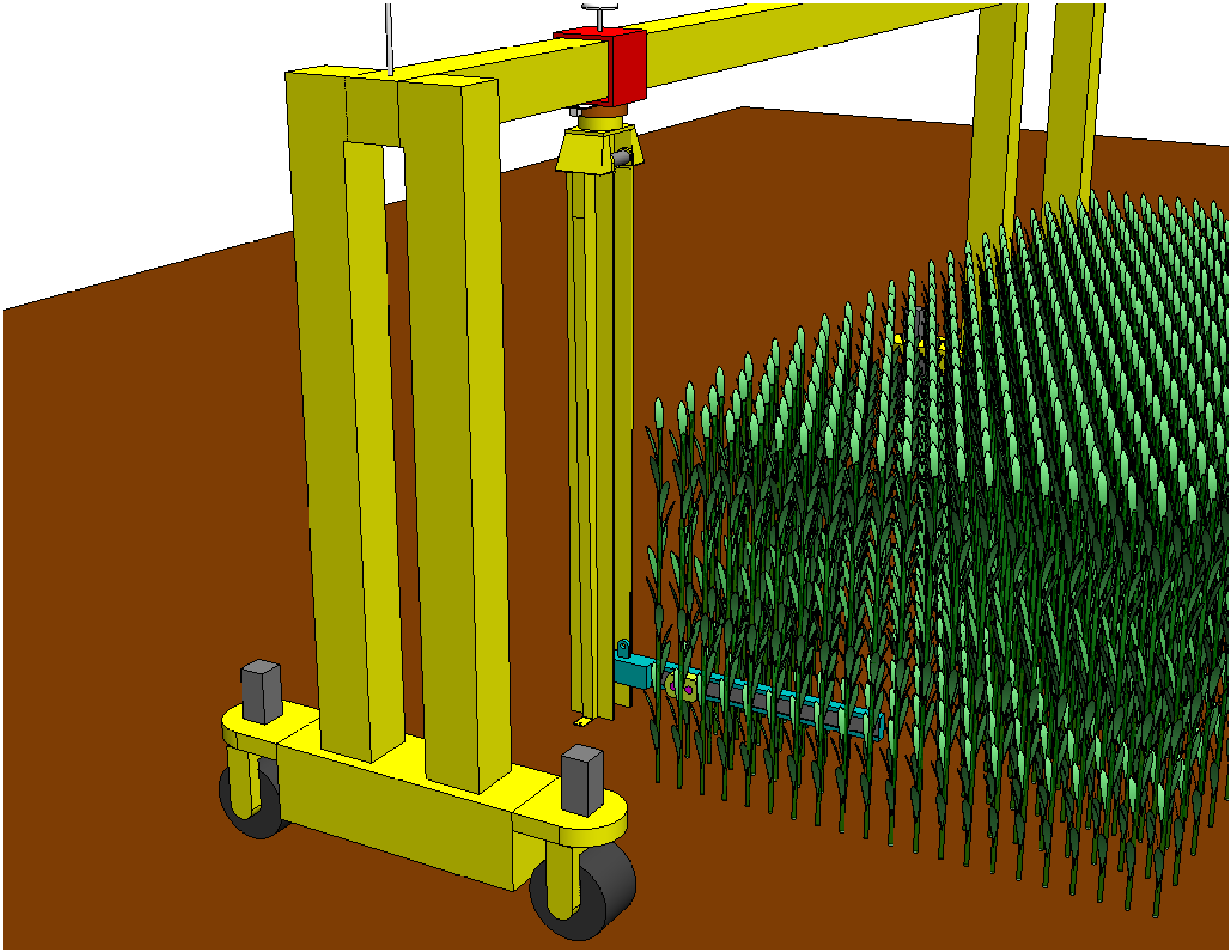
Agriculture

Agriculture

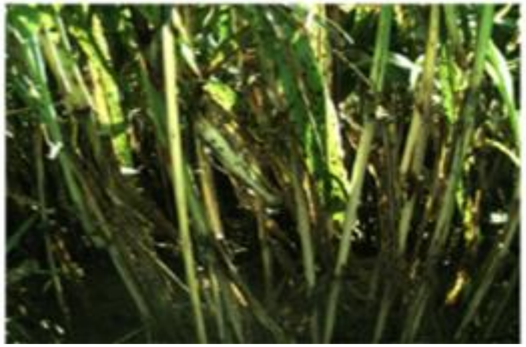
Recommend experiments (which plants to cross) to sorghum breeders.

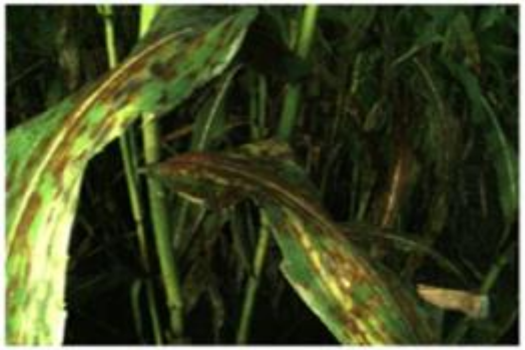


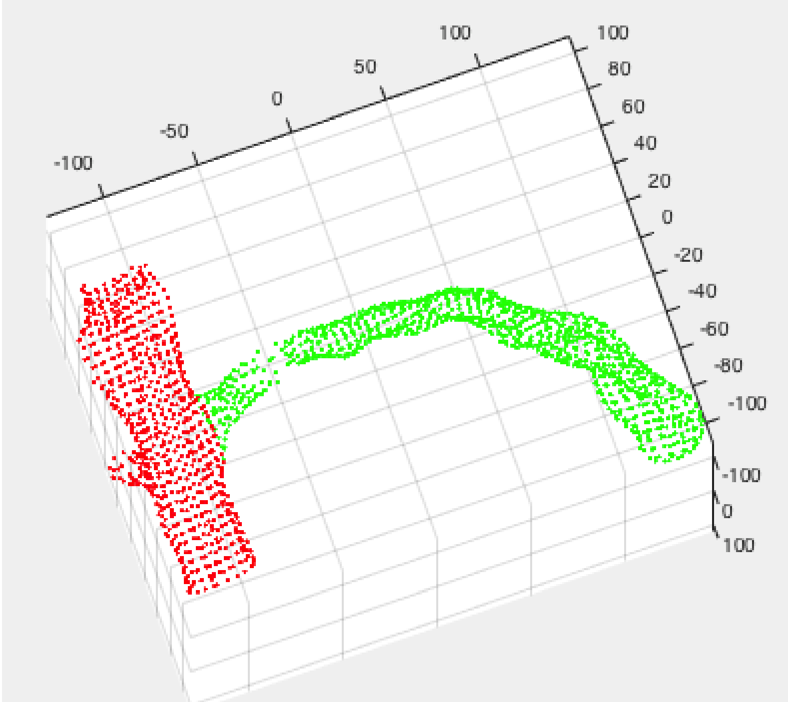
CMU Robot

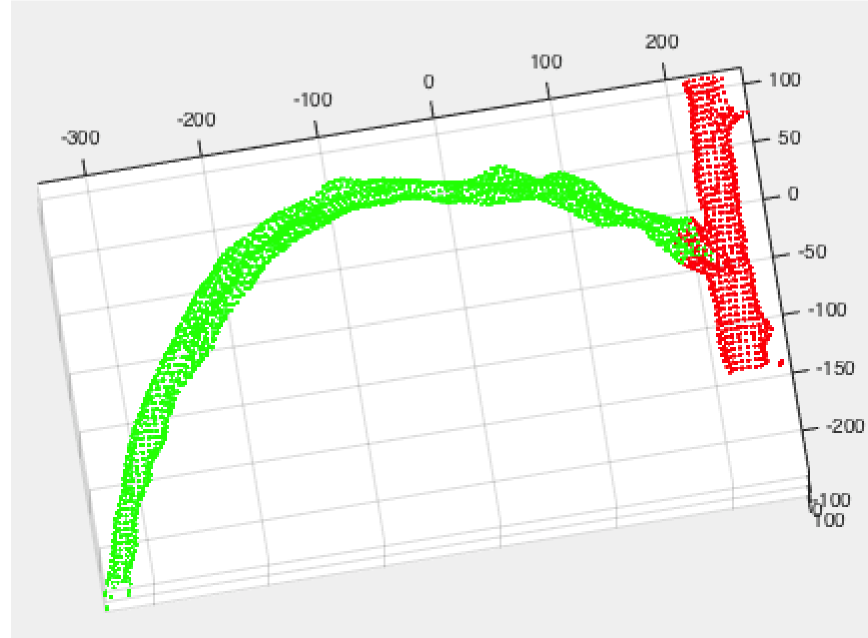


- 1 PB data …









Name	Range	RMSE error
Leaf angle*	75.94	3.30 (4.35%)
Leaf radiation angle*	120.66	4.34 (3.60%)
Leaf length*	35.00	0.87 (2.49%)
Leaf width [max]	3.61	0.27 (7.48%)
Leaf width [average]	2.99	0.21 (7.02%)
Leaf area*	133.45	8.11 (6.08%)

Grapes datasets

Take me home!

- ❑ **ML on Complex Objects**
  - o ML on distributions
  - o Lasso on functions

- ❑ **Active learning and design optimization**

- ❑ **Applications:**
  - ❑ Cosmology
  - ❑ Drug Design
  - ❑ Agriculture
  - ❑ Neuroscience

Thanks for your attention! ☺

If interested, please contact me! ☺

bapoczos@cs.cmu.edu, GHC-8231

Linear Functional Regression

Functional analogues to finite dim linear regression models:
for $Y_i \in \mathbb{R}$, $\epsilon_i \sim \mathcal{N}(0, \sigma)$, and $\Psi \subseteq \mathbb{R}^k$, a compact set:

One Real Vector vs. Functional Covariate:

Real Vector Covariate	**Functional Covariate**
$Y_i = \langle X_i, w \rangle + \epsilon_i$	$Y_i = \langle f^{(i)}, g \rangle + \epsilon_i$

where

$X_i, w \in \mathbb{R}^d$ and	$f^{(i)}, g \in L_2(\Psi)$ and
$\langle X_i, w \rangle = \sum_{j=1}^{d} X_{ij} w_j$	$\langle f^{(i)}, g \rangle = \int_\Psi f^{(i)}(t) g(t) \mathrm{d}t$