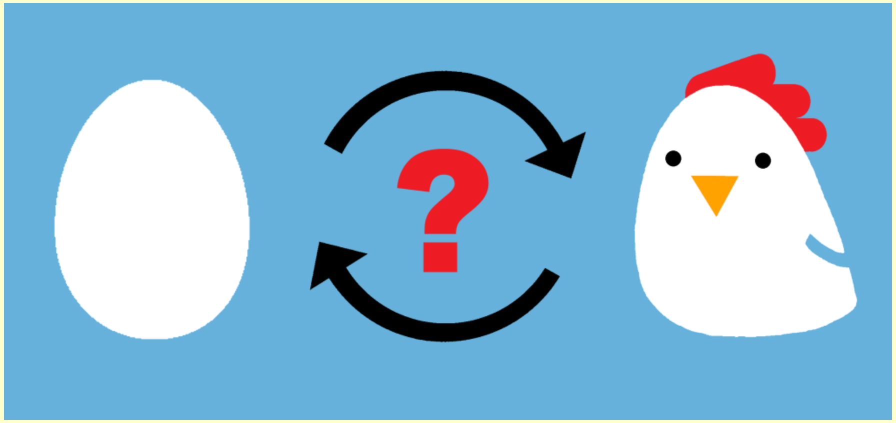
Cause-Effect Pairs Challenge



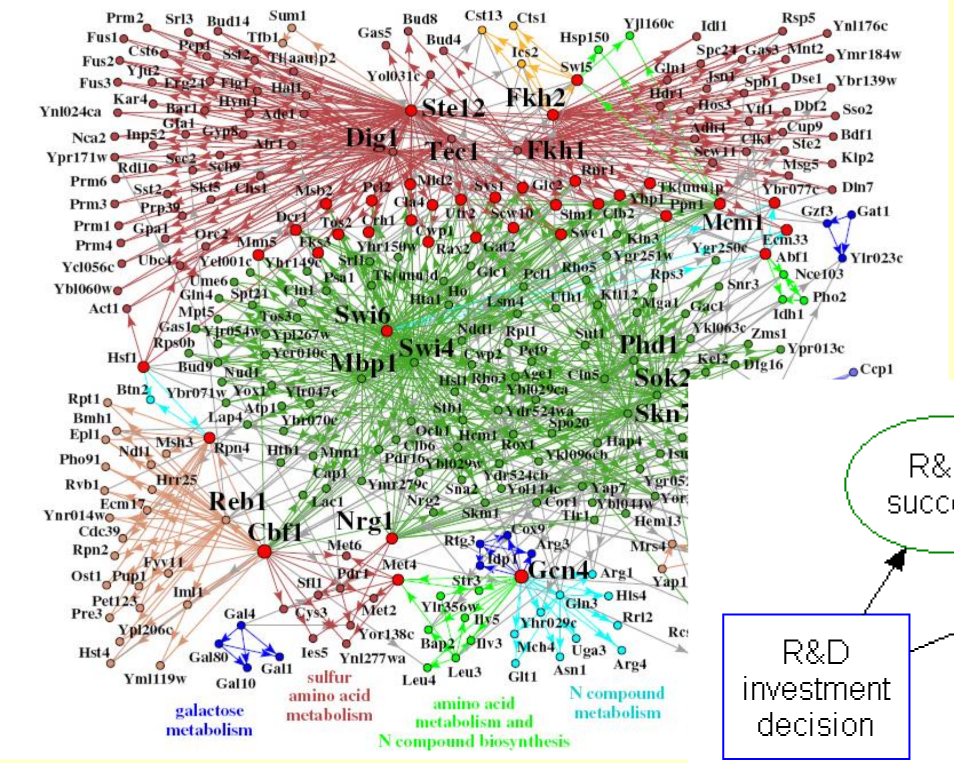
Isabelle Guyon

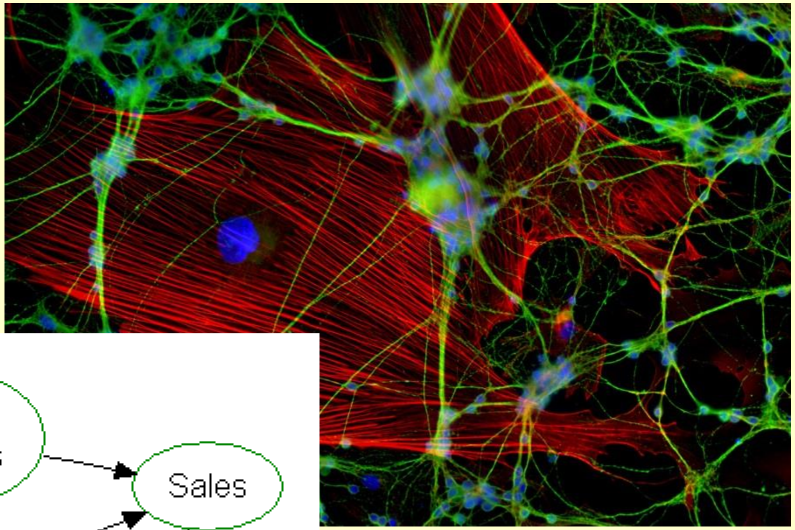
ChaLearn

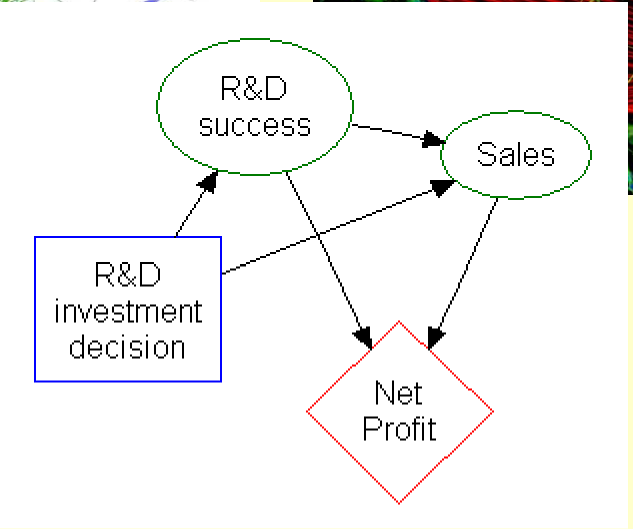
Thanks

Initial impulse: Joris Mooij, Dominik Janzing, and Bernhard Schölkopf, from the Max Planck.
Examples of algorithms and data: Povilas Daniušis, Arthur Gretton, Patrik O. Hoyer, Dominik Janzing, Antti Kerminen, Joris Mooij, Jonas Peters, Bernhard Schölkopf, Shohei Shimizu, Oliver Stegle, and Kun Zhang, Jakob Zscheischler.
Datasets and result analysis: Isabelle Guyon + Mehreen Saeed + {Mikael Henaff, Sisi Ma, and Alexander Statnikov}, from NYU.
Website and sample code: Isabelle Guyon + Ben Hamner (Kaggle).
Review, testing: Marc Boullé, Hugo Jair Escalant, Frederick Eberhardt, Seth Flaxman, Patrik Hoyer, Dominik Janzing, Richard Kennaway, Vincent Lemaire, Joris Mooij, Jonas Peters, Florin , Peter Spirtes, Ioannis Tsamardinos, Jianxin Yin, Kun Zhang.

Causal discovery without overfitting?





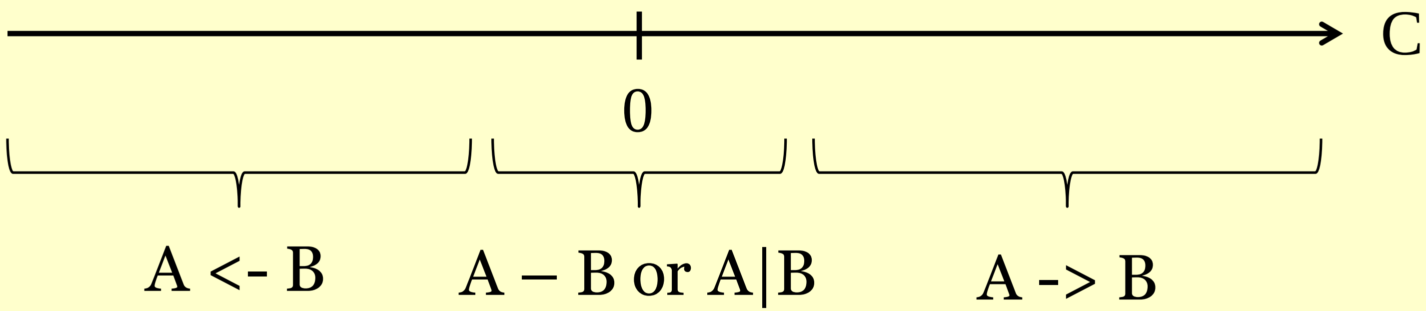


Neural networks
100 billion neurons

Gene networks
100,000 genes

Small networks: Influence diagrams

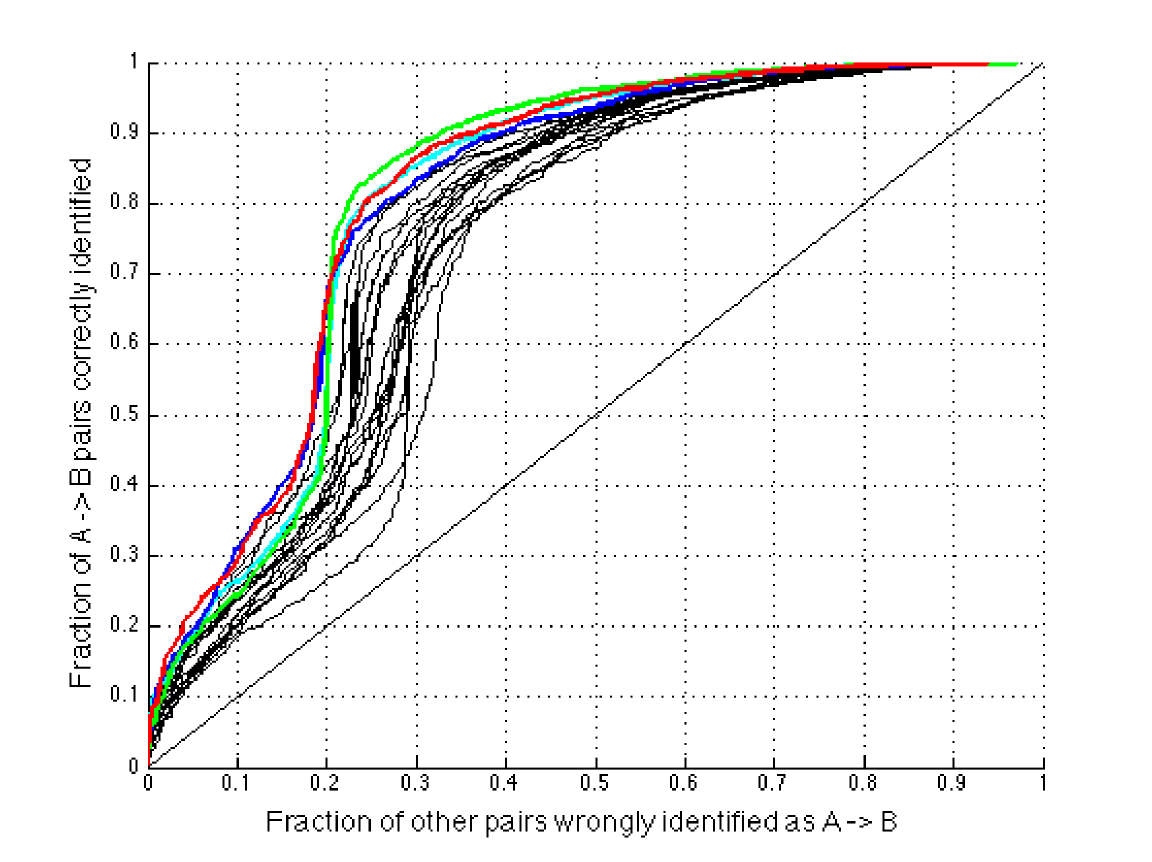
Causation coefficient



C can be used to

- RANK pairs of variables and prioritize experiments
- Orient edges in degenerate causal graphs

ROC curves for A->B



Winners

1. **ProtoML** (Rank 1): Diogo Moitinho de Almeida.
2. **Jarfo** (Rank 2): José Adrián Rodríguez Fonollosa.
3. **FirfID** (Rank 4): Spyridon Samothrakis.

#	Δ1w	Team Name ‡ model uploaded * in the money	Score	Entries
1	↑190	ProtoML ‡ *	0.81960	25
2	↑68	jarfo ‡	0.81052	123
3	↑157	HiDLoN ‡	0.80720	59
4	↑116	FirfiD ‡	0.79957	221
5	↓1	mouse ‡	0.78782	30
6	↑31	Domcastro & Sayani ‡	0.78133	324
7	↑222	nor ‡	0.77595	20
8	↓5	LucaToni ‡	0.77081	126
9	↑105	Rangel Dokov ‡	0.76780	32
10	↑16	liubenyuan & Abhishek ‡	0.76502	70
11	↑73	Saeh & Xing ‡	0.76181	33
12	↑133	Rahan ‡	0.75666	48

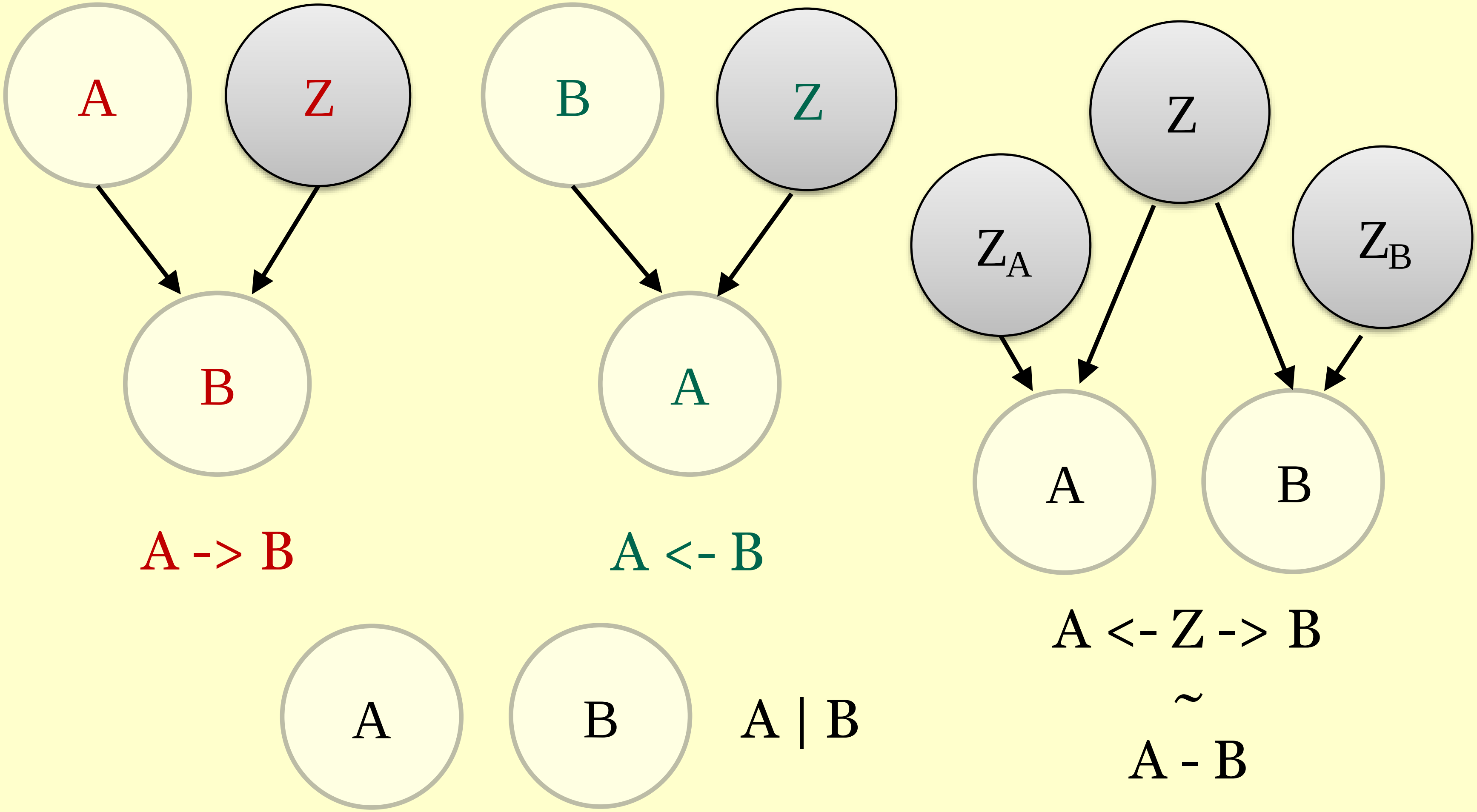
Data

Cause-effect pairs method

Test whether *A -> B* is a better explanation than *A <- B* comparing two hypotheses:

$$B = f(A, noise)$$
$$A = f(B, noise)$$

Setting of the challenge



Setting

- No feed-back loops.
- No explicit time information.
- A variable can be though of as an aggregate statistic, like life expectancy of a population, or a measurement like temperature.
- We consider pairs of variables {A, B} for which A−>B means B = f (A, noise).
- Pairs are independent of each other.

Data provided

SampleID	A	B
train1	-6348 4599 -9340 -13170 8456 -10079 -68 7957 -44	6 7 6 7 7 7 7 7 1 7 7 7 1 7 7 7 7 3 3 7 7 7 6 7 7 7 7 1 1 7 7
train2	-6462 7666 19406 -2299 -22045 -6262 24734 -8854	3 4 5 5 5 4 3 7 4 5 5 4 1 5 4 4 5 4 5 5 5 5 1 4 1 4 4 4 5 5
train3	12800 -6791 -539 -9092 9818 1646 13806 324 -6031	-7469 38799 -6292 2224 -11357 -10823 -8578 -5095 -
train4	14 580 -6627 10738 13938 -13793 -17467 -1269 -84	11883 -82 -3086 3150 -7775 -5290 1338 8765 5267 11
train5	1 1 0 0 1 0 1 1 1 0 1 0 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
train6	2 2 2 1 3 1 2 2 3 3 1 2 3 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 3	-1705 -772 4741 3431 -8435 -6487 -9733 7563 -3063 -
train7	4 4 4 4 7 4 5 4 5 7 2 2 4 4 4 4 2 4 4 7 4 4 2 4 4 4 4 7 4	3449 14747 -21631 -4777 -13084 4257 -3262 15538 -2
train8	-15214 -4766 6890 129 14304 -8366 -5836 2902 -47	10161 14740 2615 1491 -6601 24004 -1357 -2027 -951
train9	-8062 19980 -12068 -6059 -50 9965 -10065 5959 -16074 -10065 -50 -6059 -12068 3956 11968 -50 -2053 15	

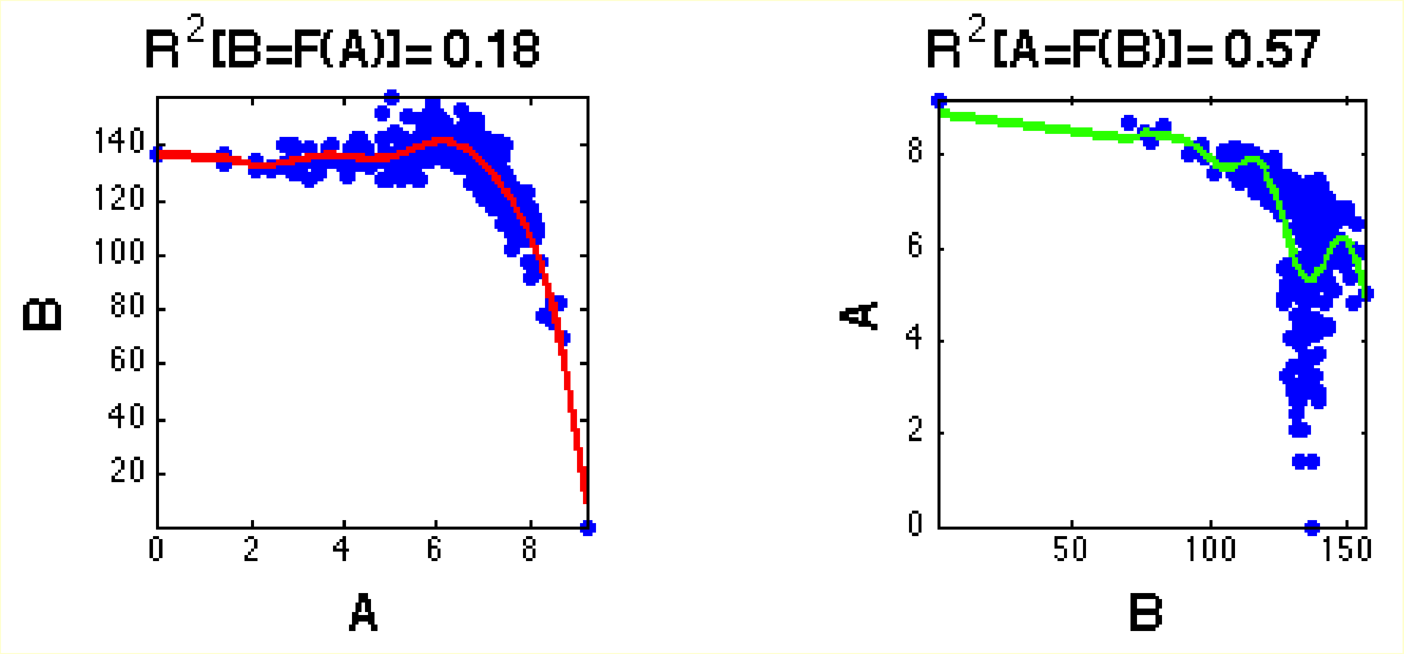
SampleID	Target [1 f	Details [1
train1	0	4
train2	1	1
train3	1	1
train4	1	1
train5	0	3
train6	1	1
train7	0	4
train8	0	4
train9	0	3

SampleID	Source	A name	B name	F1	N1	S/N 1	F2	N2	S/N 2	sample nu	A type	B type	RealData [
train1	INDEP	A	B	NA	NA	NA	NA	NA	NA	577	Numerica	Categorica	0
train2	ARTIFCE	A	B	atanh	preadd	0.25	NA	NA	NA	3017	Numerica	Categorica	0
train3	ARTIFCE	A	B	sine	postadd	1	NA	NA	NA	2728	Numerica	Numerica	0
train4	ARTIFCE	A	B	cubic	postmult	0.5	NA	NA	NA	802	Numerica	Numerica	0
train5	RELATED	A	B	tanh	postadd	1	sine	premult	1	611	Binary	Binary	0
train6	CAR	org	Weight	NA	NA	NA	NA	NA	NA	406	Categorica	Numerica	1
train7	INDEP	A	B	NA	NA	NA	NA	NA	NA	4056	Categorica	Numerica	0
train8	INDEP	A	B	NA	NA	NA	NA	NA	NA	2017	Numerica	Numerica	0
train9	RELATED	A	B	atanh	premult	2	line	postadd	1	7503	Numerica	Numerica	0

Example:
Best fit: A -> B

A -> B

A <- B



Large dataset

- Real data (18%):
  - Altitude -> Temperature
  - Age -> Wages
  - Car color -> Price
  - Country -> Infant mortality
- Artificial data (82%):

  B = f(A, noise)

~~Set~~	~~n pairs~~
~~TRAIN~~	~~31~~
~~VALID~~	~~264~~

Set	Num pairs
FINAL TRAIN	4050
FINAL VALID	4050
FINAL TEST	4050

Set	Num pairs
SUP1 [numerical]	5998
SUP2 [mixed]	5989
SUP3 [numerical+binary]	162

Real variables

Demographics:
Sex -> Height
Age -> Wages
Native country -> Education
Latitude -> Infant mortality

Ecology:
City elevation -> Temperature
Water level -> Algal frequency
Elevation -> Vegetation type
Distance to hydrology -> Fire

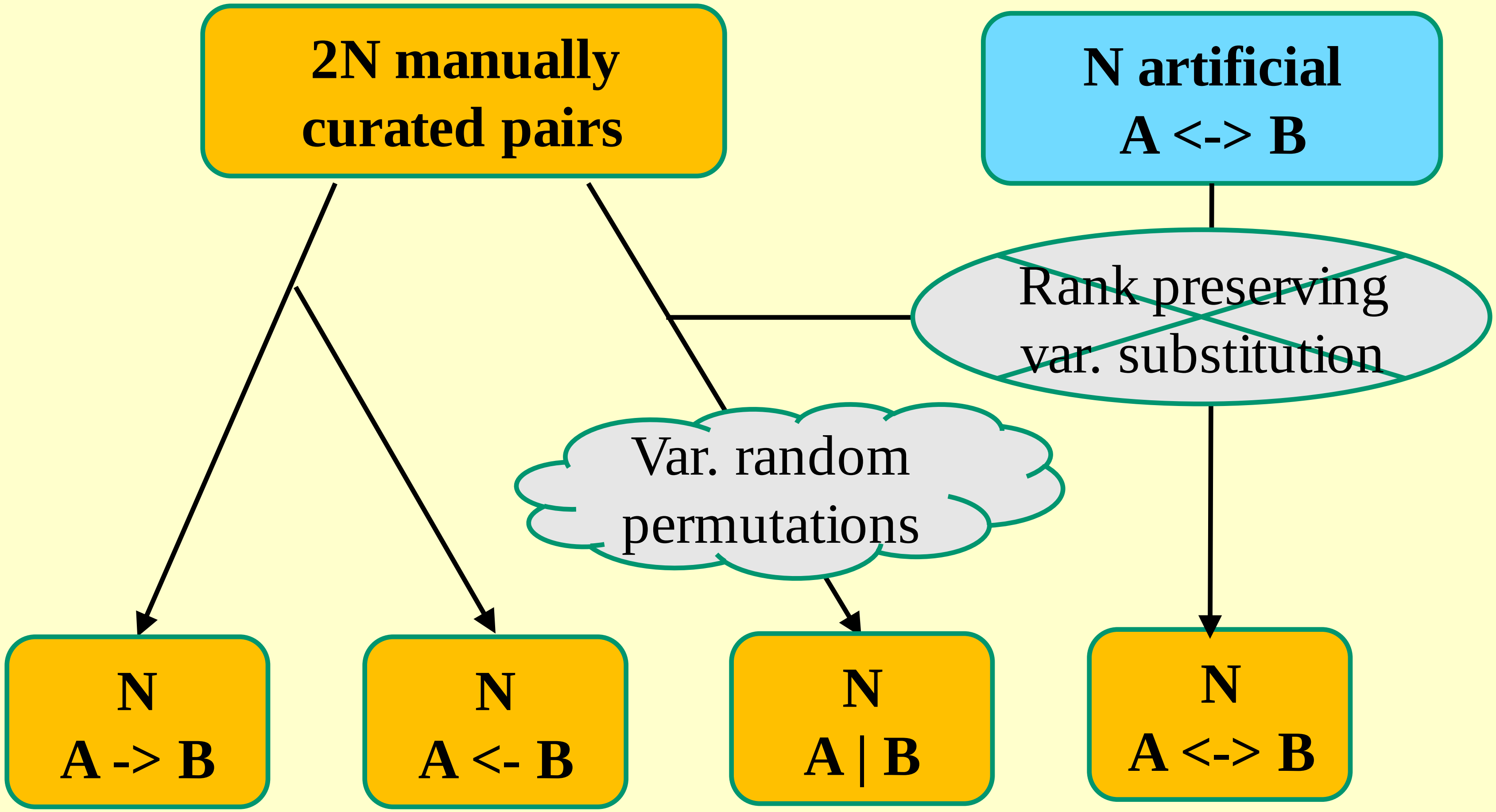
Econometrics:
Mileage -> Car resell price
Number of rooms -> House price
Trace price last day -> Trade price

Medicine:
Cancer volume -> Recurrence
Metastasis -> Prognosis
Age -> Blood pressure

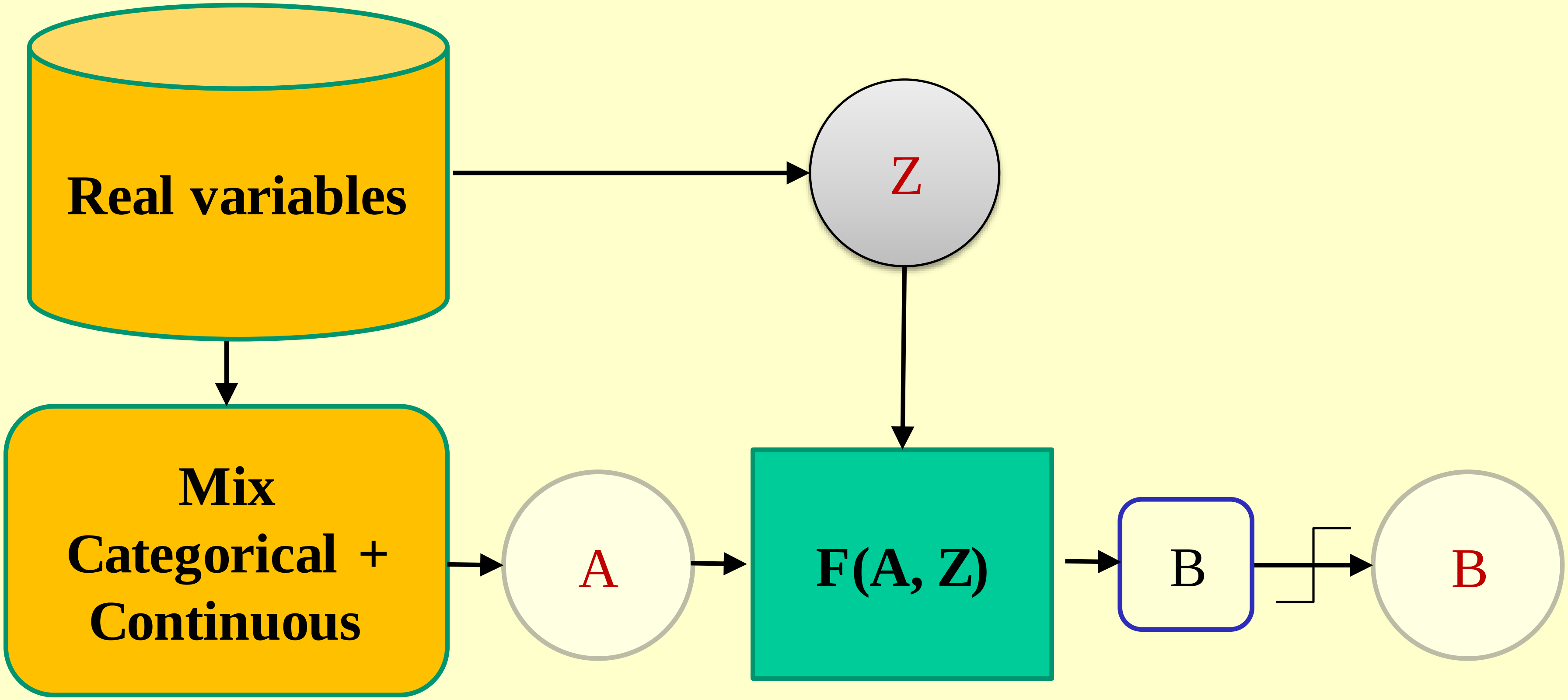
Genomics (mRNA level):
transcription factor -> protein induced

Engineering:
Car model year -> Horsepower
Number of cylinders -> MPG
Cache memory -> Compute power
Roof area -> Heating load
Cement used -> Compressive strength

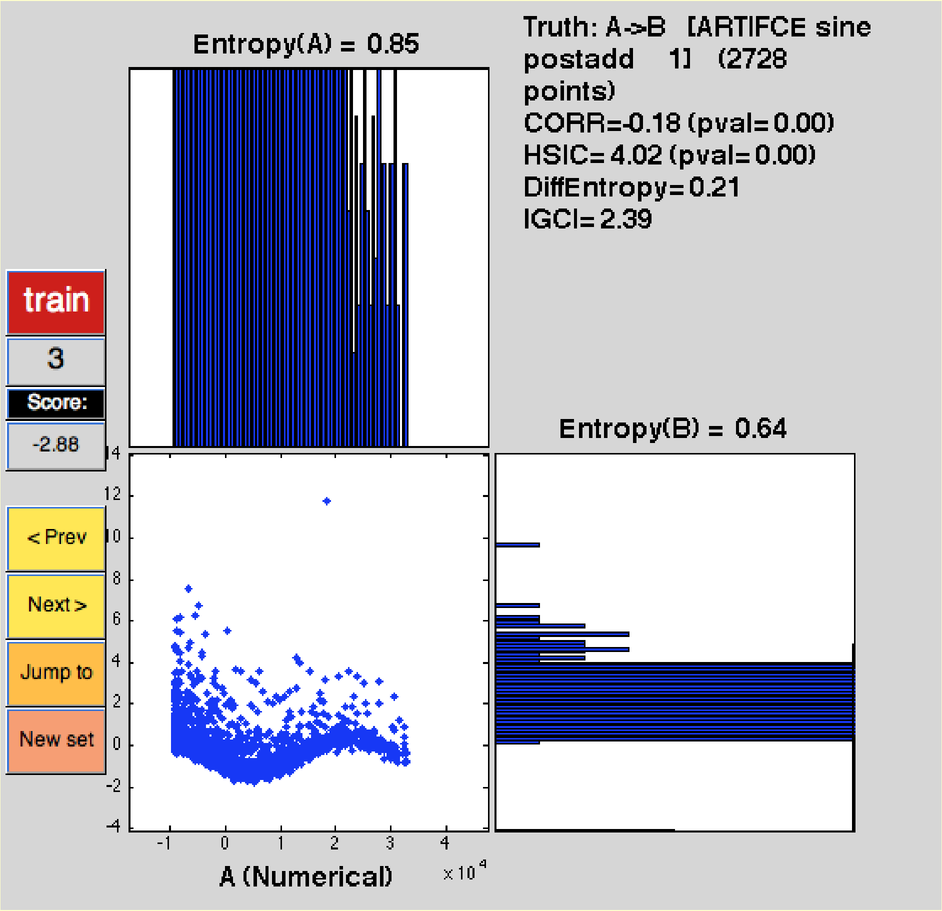
Real variables



Artificial data



Data browser and sample code



Result analysis

Model-based methods

- Additive Noise Model (ANM): Best fit, compare independence of input and residual.
- Latent variable models (LINGAM): Enforce independence of input and residual, compare model weights.
- Complexity-based models: Select simplest explanation of the data (GPI and IGCI).

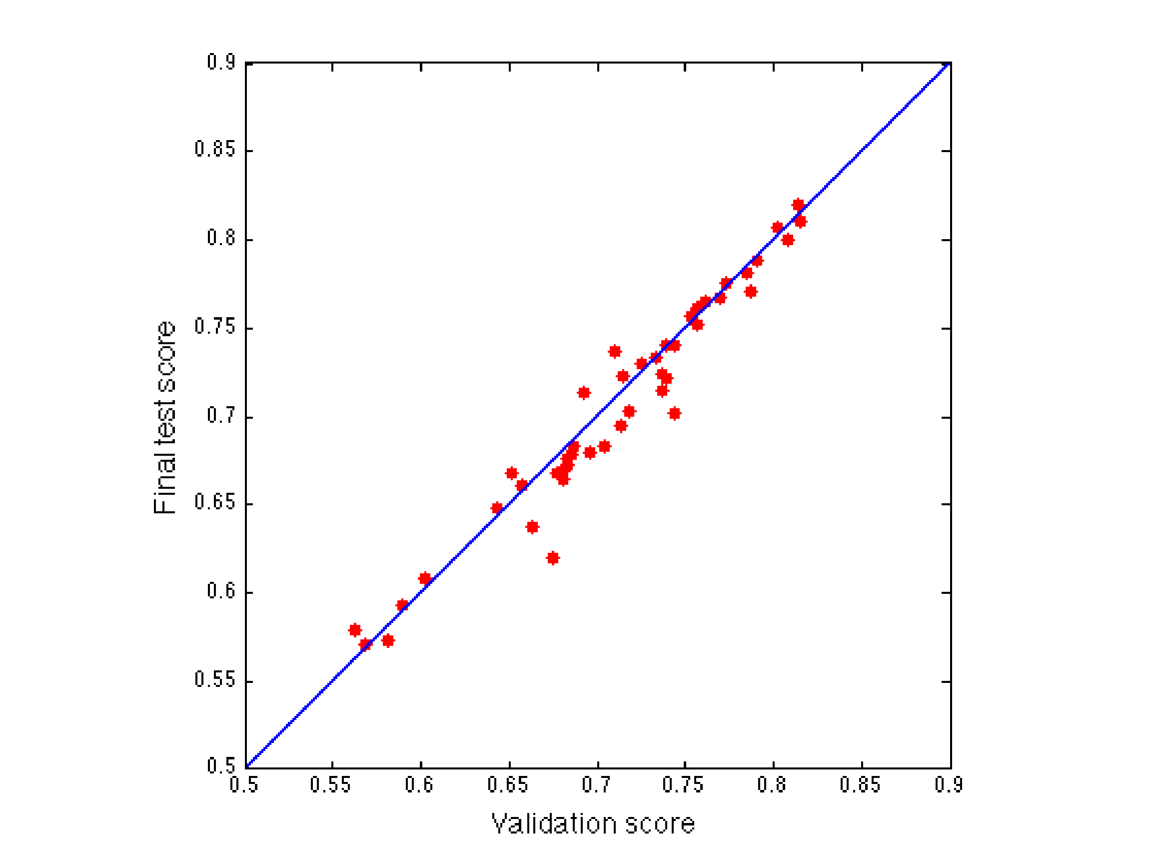
http://webdav.tuebingen.mpg.de/causality/

Empirical methods

- 267 teams and 4578 entries.
- All baseline methods outperformed!
- Code of 3 winners available.

#	Δ1w	Team Name ‡ model uploaded * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑8	ProtoML ‡ *	0.81960	25	Tue, 27 Aug 2013 13:33:43
2	↑7	jarfo ‡	0.81052	123	Tue, 27 Aug 2013 10:40:37
3	↑6	HiDLoN ‡	0.80720	59	Mon, 02 Sep 2013 05:44:45
4	↑5	FirfiD ‡	0.79957	221	Tue, 27 Aug 2013 13:28:46
5	↓1	mouse ‡	0.78782	30	Wed, 28 Aug 2013 20:21:42
6	↑3	Domcastro & Sayani ‡	0.78133	324	Wed, 28 Aug 2013 15:18:27

No overfitting

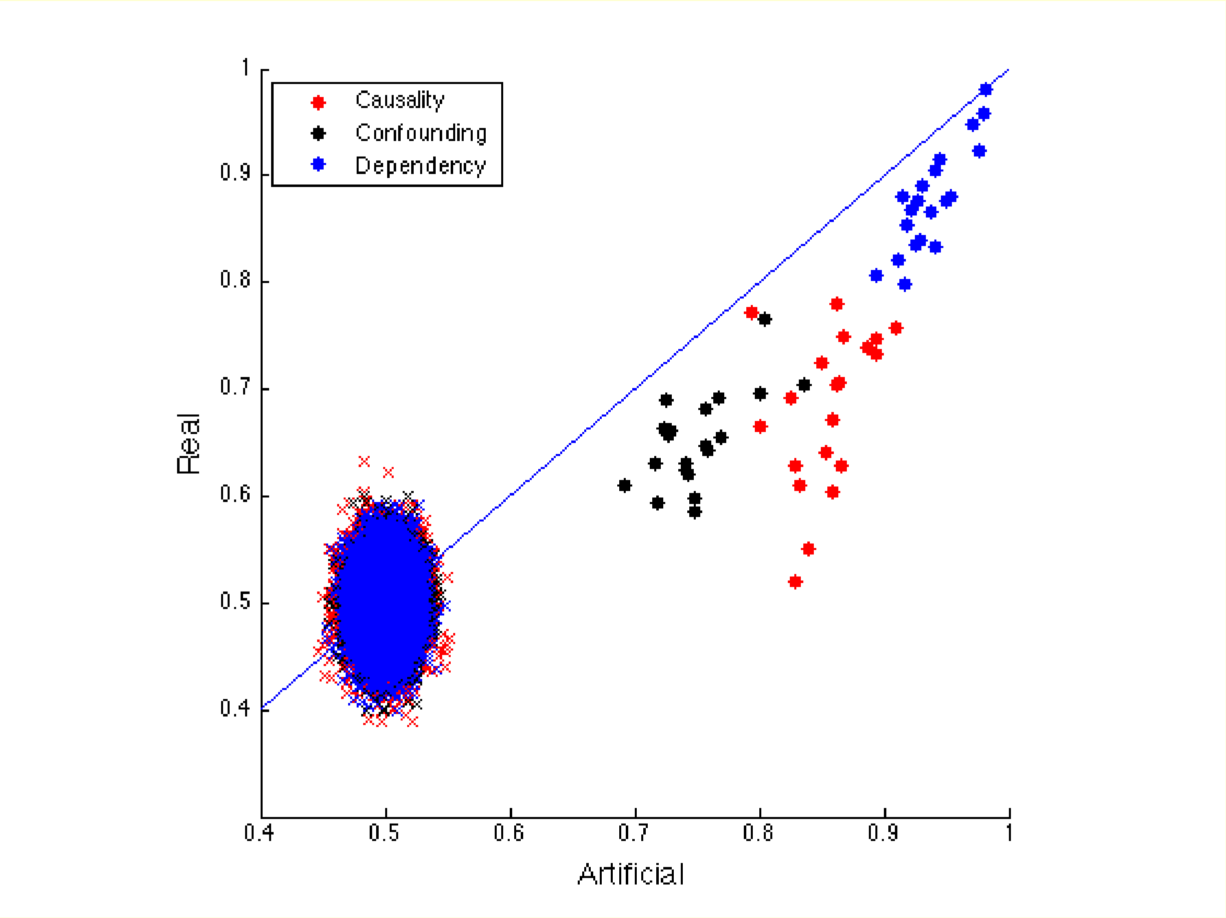


Result comparison

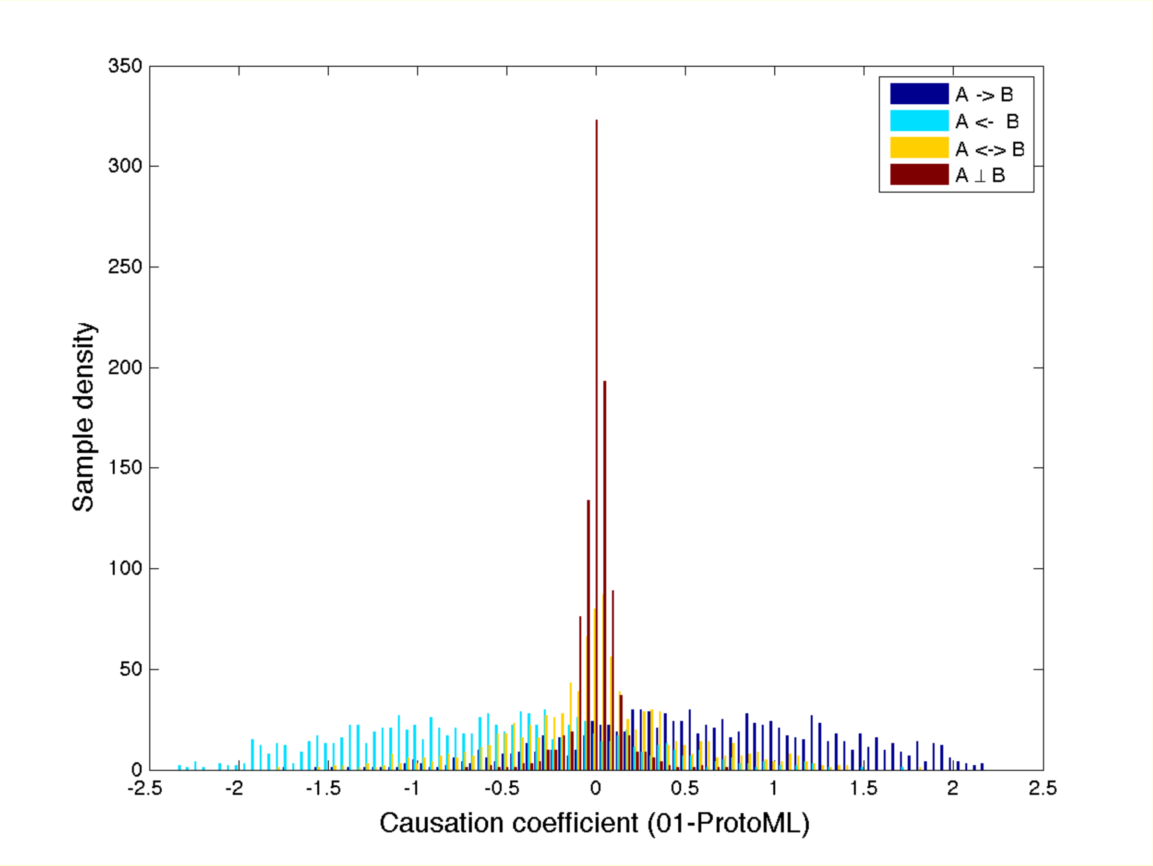
Artificial data					
Rank	Team	Dependency	Confounding	Causality	Score
1	ProtoML	0.95372	0.76944	0.90946	0.84206
2	jarfo	0.98063	0.83663	0.89425	0.83499
3	HiDloN	0.94416	0.76777	0.89466	0.82883
4	FirfiD	0.97644	0.80086	0.88644	0.82249
5	mouse	0.94966	0.75831	0.86722	0.80620
6	Domcasto & Sayani	0.91789	0.72655	0.86299	0.79507

Real data					
Rank	Team	Dependency	Confounding	Causality	Score
1	ProtoML	0.88057	0.65432	0.75756	0.70420
2	jarfo	0.95721	0.70386	0.73312	0.68642
3	HiDloN	0.91476	0.69209	0.74774	0.69669
4	FirfiD	0.92352	0.69547	0.73960	0.68274
5	mouse	0.87689	0.64211	0.75008	0.69259
6	Domcasto & Sayani	0.85339	0.65786	0.78075	0.71355

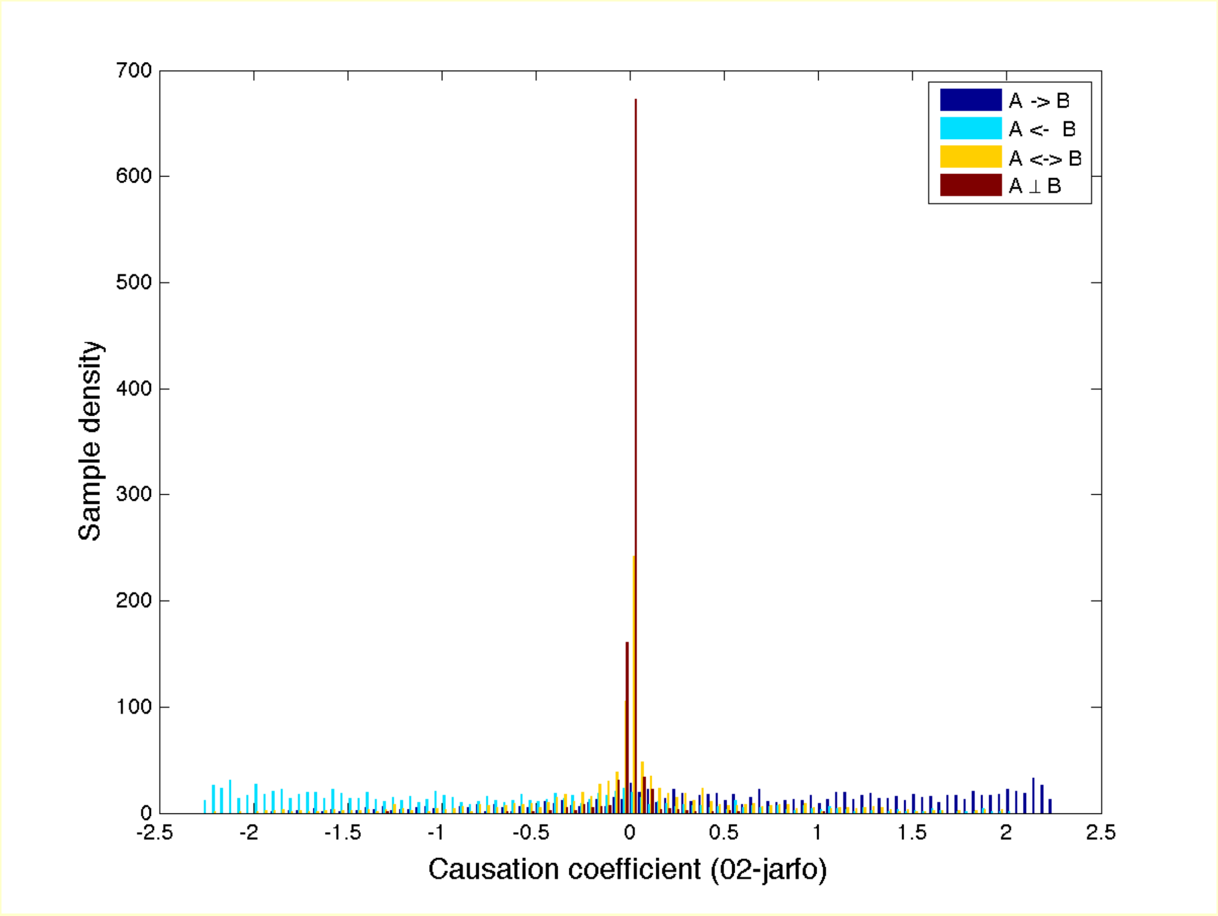
Statistical significance



Causation coefficient distribution



Causation coefficient distribution



Post-challenge verifications

3648 cause effect pairs from GeneNetWeaver 3.0 (http://gnw.sourceforge.net/) based on E. Coli transcriptional regulatory network.

Experiment 1: no retraining

Experiment 2: train ½, test ½.

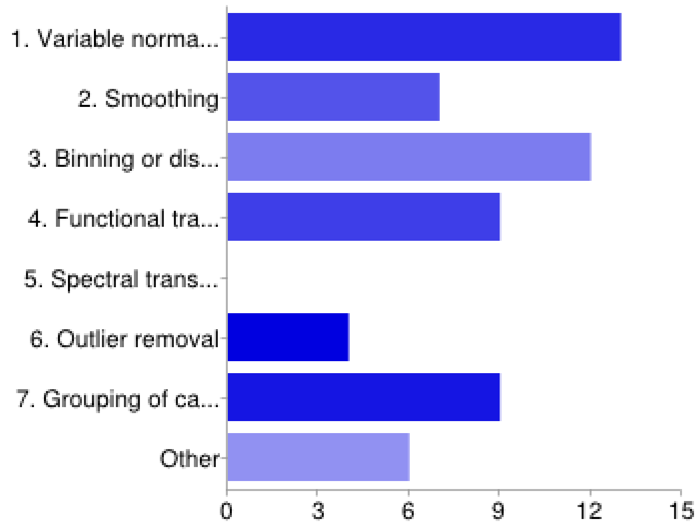
		Experiment 1	Experiment 2
AUC	Jarfo	0.873	0.9972
	FirfiD	0.5963	0.9845
	ProtoML	0.8085*	0.9908
Time	Jarfo	~ 5 hrs	~ 5 hrs
	FirfiD	~ 7 hrs	~ 8 hrs
	ProtoML	~ 10 hrs*	~ 12 hrs

Alexander Statnikov and Sisi Ma

Survey (27 responses)

Preprocessing

Preprocessing of A and B variables

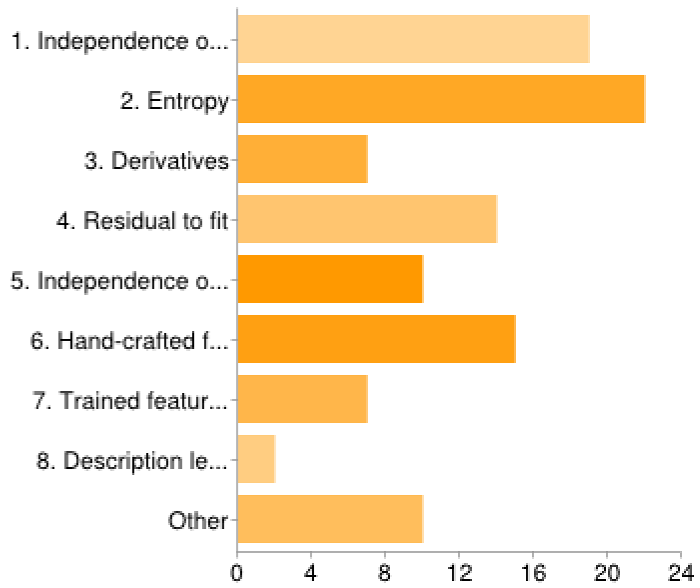


1. Variable normalization	**13**	48%
2. Smoothing	**7**	26%
3. Binning or discretization	**12**	44%
4. Functional transform (e.g. log)	**9**	33%
5. Spectral transform	**0**	0%
6. Outlier removal	**4**	15%
7. Grouping of categorical variables	**9**	33%
Other	**6**	22%

People may select more than one checkbox, so percentages may add up to more than 100%.

Feature extraction

Feature extraction

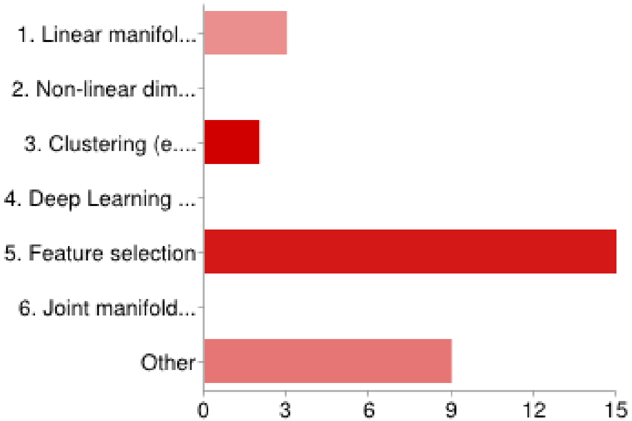


1. Independence of variables	**19**	70%
2. Entropy	**22**	81%
3. Derivatives	**7**	26%
4. Residual to fit	**14**	52%
5. Independence of input and residual	**10**	37%
6. Hand-crafted features	**15**	56%
7. Trained feature extractors	**7**	26%
8. Description length or complexity of model	**2**	7%
Other	**10**	37%

People may select more than one checkbox, so percentages may add up to more than 100%.

Dimensionality reduction

Dimensionality reduction

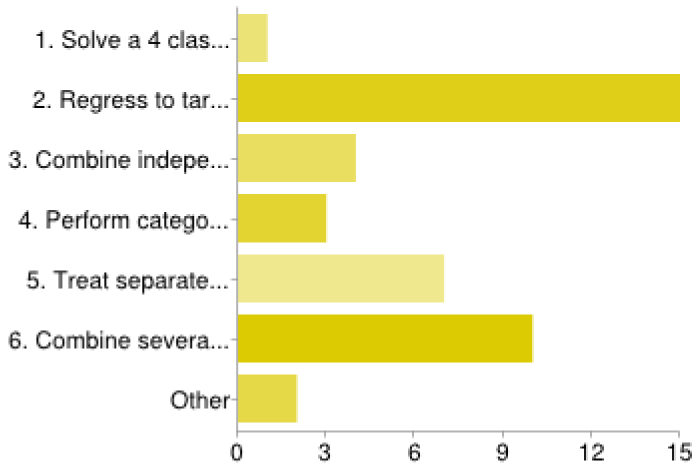


1. Linear manifold transformations (e.g. factor analysis, PCA, ICA)	**3**	11%
2. Non-linear dimensionality reduction (e.g. KPCA, MDS, LLE, Laplacian Eigenmaps, Kohonen maps)	**0**	0%
3. Clustering (e.g. K-means, hierarchical clustering)	**2**	7%
4. Deep Learning (e.g. stacks of auto-encoders, stacks of RBMs)	**0**	0%
5. Feature selection	**15**	56%
6. Joint manifold data fusion	**0**	0%
Other	**9**	33%

People may select more than one checkbox, so percentages may add up to more than 100%.

Recognition

Architecture

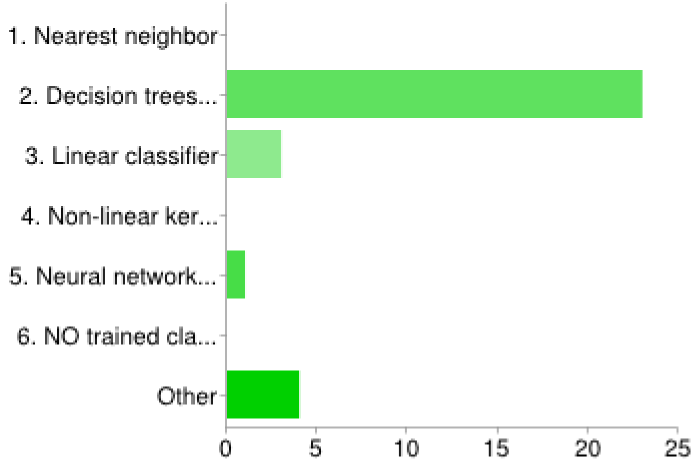


1. Solve a 4 class problem then combine	**1**	4%
2. Regress to targets -1/0/1	**15**	56%
3. Combine independence and causal direction scores	**4**	15%
4. Perform categorical regression	**3**	11%
5. Treat separately categorical variables	**7**	26%
6. Combine several strategies	**10**	37%
Other	**2**	7%

People may select more than one checkbox, so percentages may add up to more than 100%.

Classifier

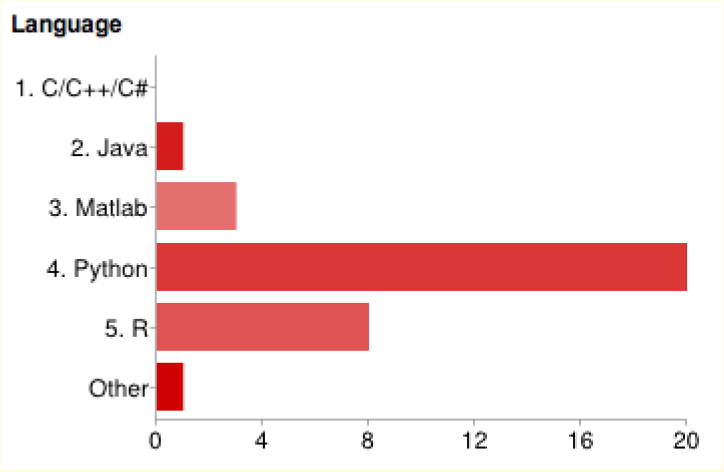
Classifier

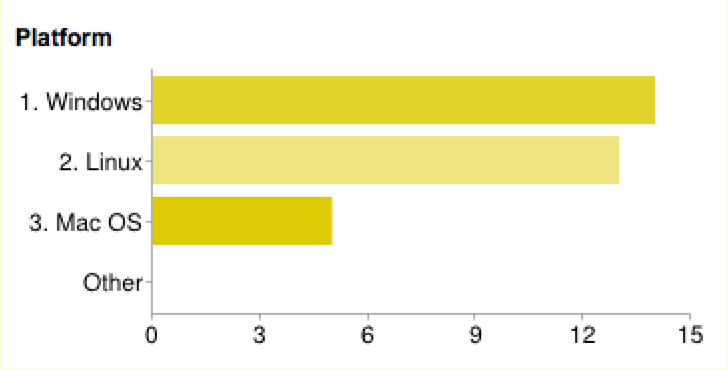


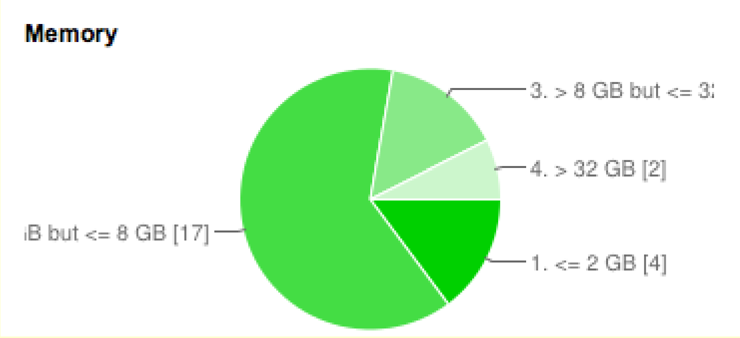
1. Nearest neighbor	**0**	0%
2. Decision trees or random forests	**23**	85%
3. Linear classifier	**3**	11%
4. Non-linear kernel method	**0**	0%
5. Neural networks or deep learning	**1**	4%
6. NO trained classifier	**0**	0%
Other	**4**	15%

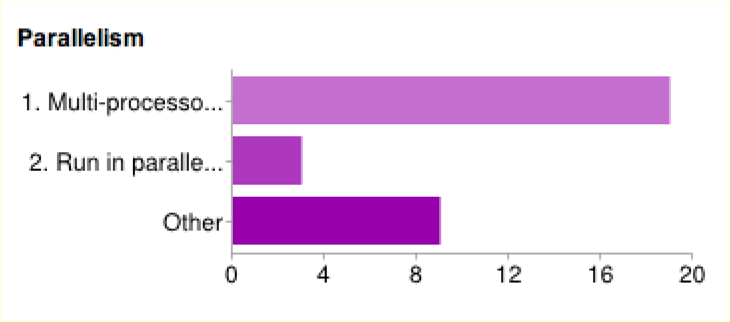
People may select more than one checkbox, so percentages may add up to more than 100%.

Implementation

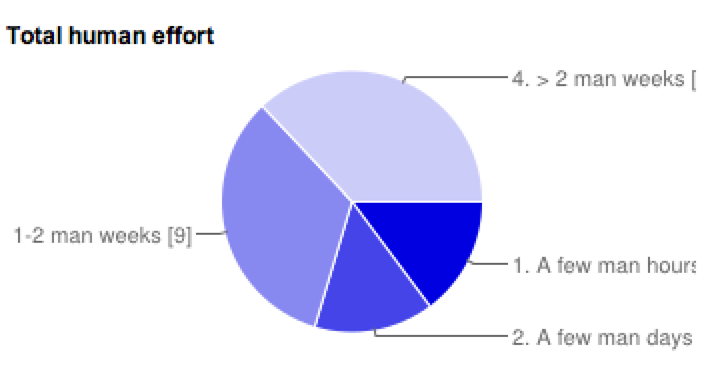




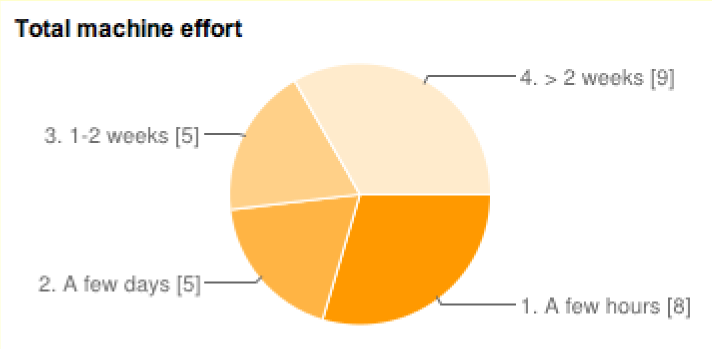




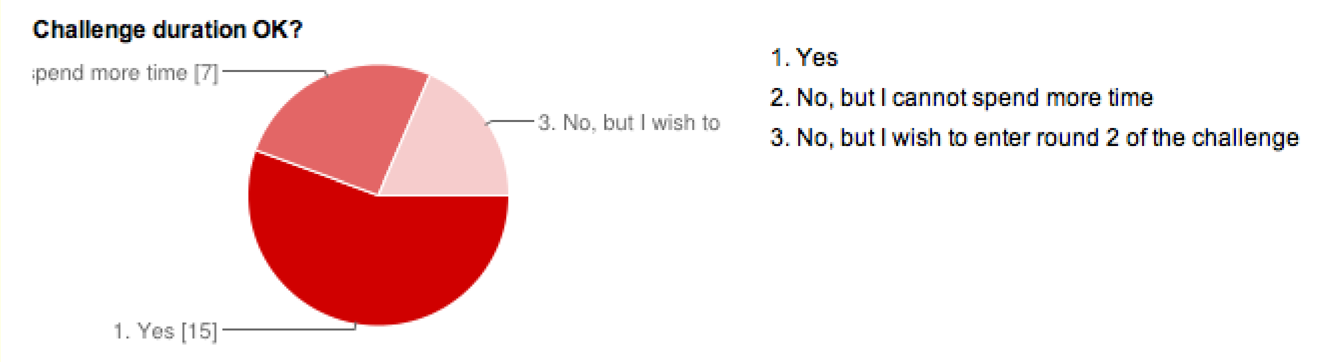
Time spent



Total human effort
4. > 2 man weeks [
1-2 man weeks [9]
1. A few man hours
2. A few man days

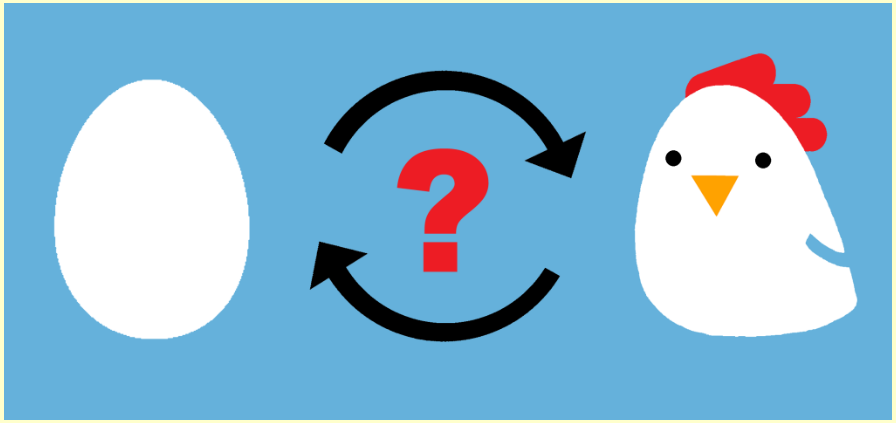


Total machine effort
4. > 2 weeks [9]
3. 1-2 weeks [5]
2. A few days [5]
1. A few hours [8]



Challenge duration OK?
pend more time [7]
3. No, but I wish to
1. Yes [15]
1. Yes
2. No, but I cannot spend more time
3. No, but I wish to enter round 2 of the challenge

Cause-Effect Pairs Challenge



http://clopinet.com/causality