MACHINE LEARNING FOR CAUSE-EFFECT PAIRS DETECTION

Mehreen Saeed
CLE Seminar
11 February, 2014.

WHY CAUSALITY….

- Polio drops can cause polio epidemics
  - (The Nation, January 2014)
- A supernova explosion causes a burst of neutrinos
  - (Scientfic American, November 2013)
- Mobile phones can cause brain tumors
  - (The Telegraph, October 2012)
- DDT pesticide my cause Alzhiemer's disease
  - (BBC, January 2014)
- Price of dollar going up causes price of gold to go down
  - (Investopedia.com, March 2011)

OUTLINE

- Causality
- Coefficients for computing causality
  - Independence measures
  - Probabilistic
  - Determining the direction of arrows
- Transfer learning
- Causality challenge
- Conclusions

Observational Vs. Experimental Data

- Observational data is collected by recording values of different characteristics
- Experimental data is collected by changing values of some characteristics of the subject and some values are under the control of an experimenter

Example: Randomly select 100 individuals and collect data on their everyday diet and their health issues

Vs.

Select 100 individuals with diabetes and omit a certain food from their diet and observe the result

Observational Vs. Experimental Data…(contd)

- Observational data: Google receives around 2 million requests/minute, Facebook users post around 680,000 pieces of content/minute, email users send 200,000,000 messages in a minute

VS.

- Experimental data: expensive, maybe unethical, maybe not possible

15 years ago it was thought that inferring causal relationships from observational data is not possible…. Research of machine learning scientists like Judea Pearl has changed this view

REF: http://mashable.com/2012/06/22/data-created-every-minute/

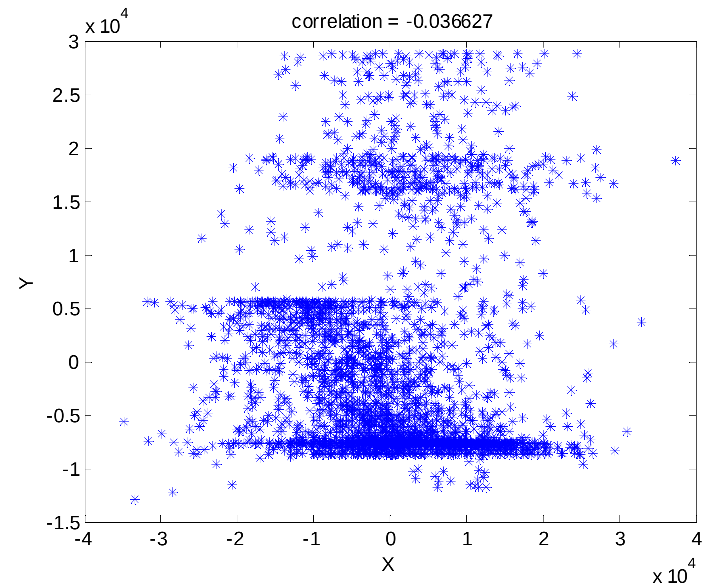
Causality: From Observational Data to Cause Effect Detection

- X->Y *smoking causes lung cancer*
- Y->X *lung cancer causes coughing*
- $X \perp Y$ *winning cricket match and being born in February*
- X->Z->Y $X \perp Y \mid Z$ (Conditional independence)
- X<-Z->Y $X \perp Y \mid Z$ (Conditional independence)

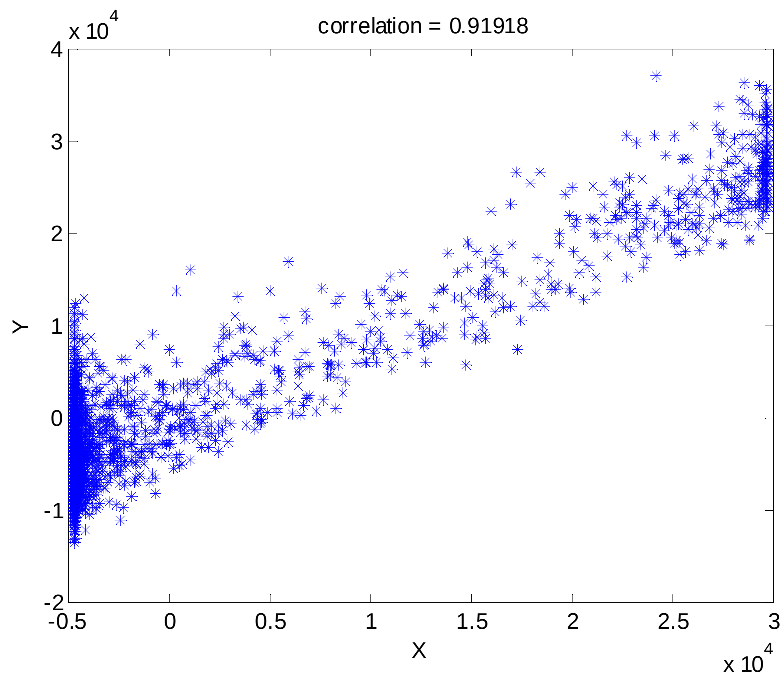
OUTLINE

- Causality
- Coefficients for computing causality
  - Independence measures
  - Probabilistic
  - Determining the direction of arrows
- Transfer learning
- Causality challenge
- Conclusions

Correlation
ρ={E(XY)-E(X)E(Y)}/STD(X)/STD(Y)

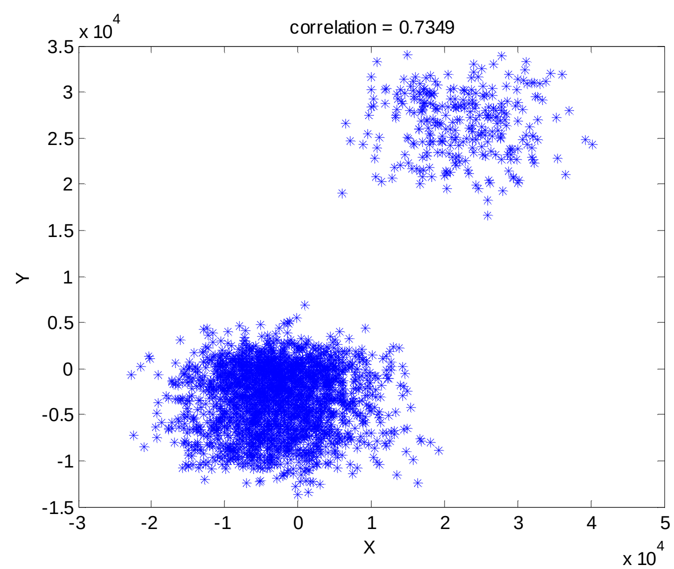


X->Y correlation = -0.04



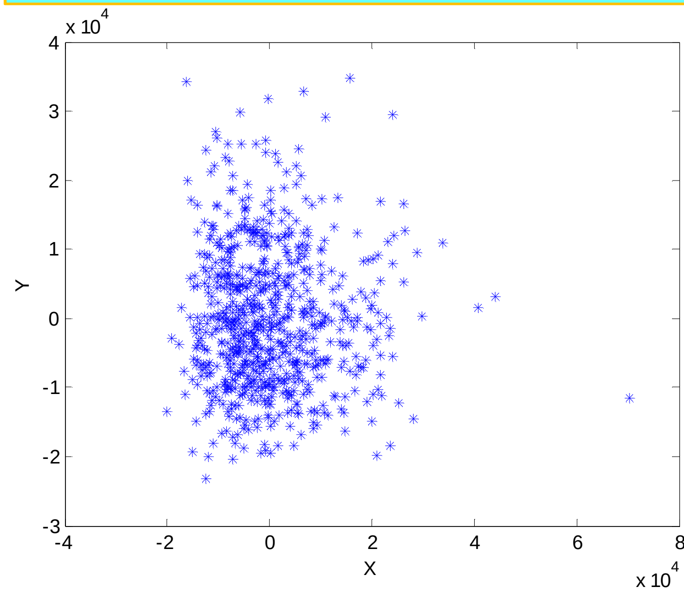
X->Y correlation = 0.9

Correlation does not necessarily imply causality

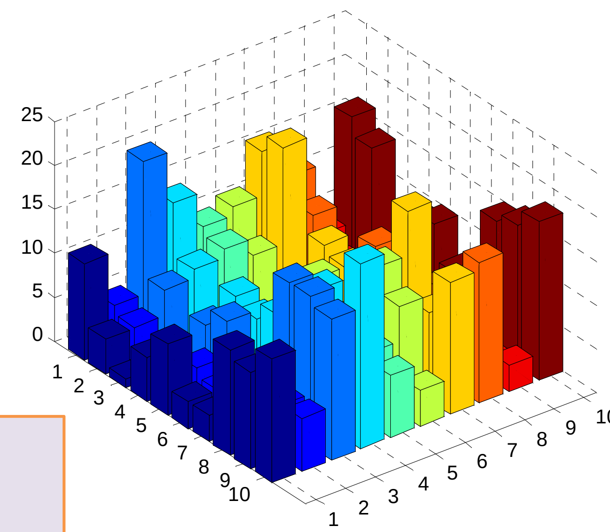


X⊥Y correlation = 0.73

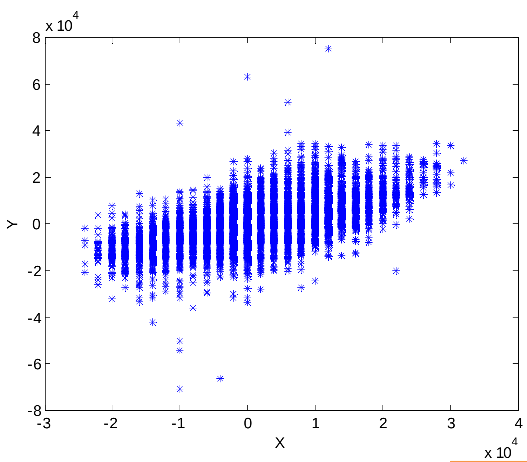
χ^2 Test For Independence



truth: X ⊥ Y



p-value = 0.99
dof = 81
chi2value = 52.6



truth: X ⊥ Y

p-value = 0
dof = 63
chi2value = 3255
corr = 0.5948

Again this test does not tell us anything about causal inference

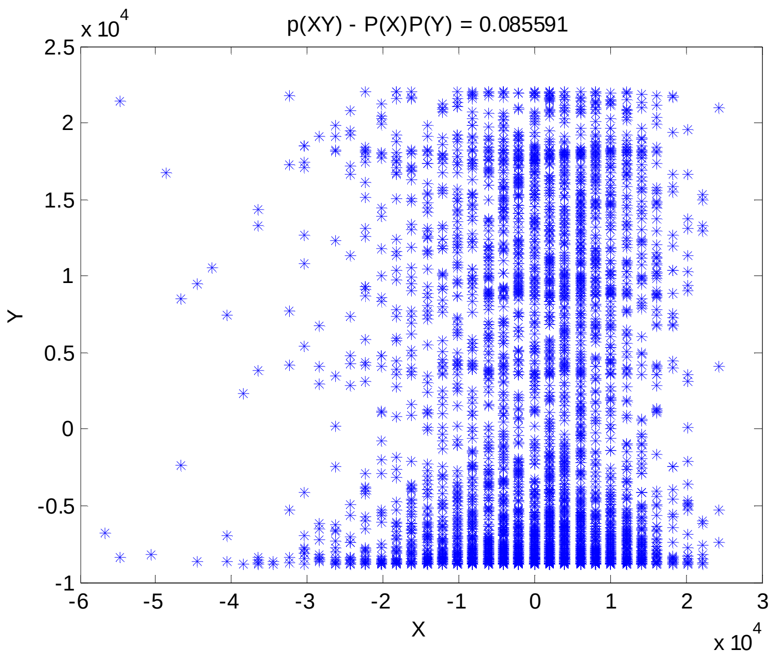
STATISTICAL INDEPENDENCE

FOR TWO INDEPENDENT EVENTS:

$$P(XY)=P(X)P(Y)$$

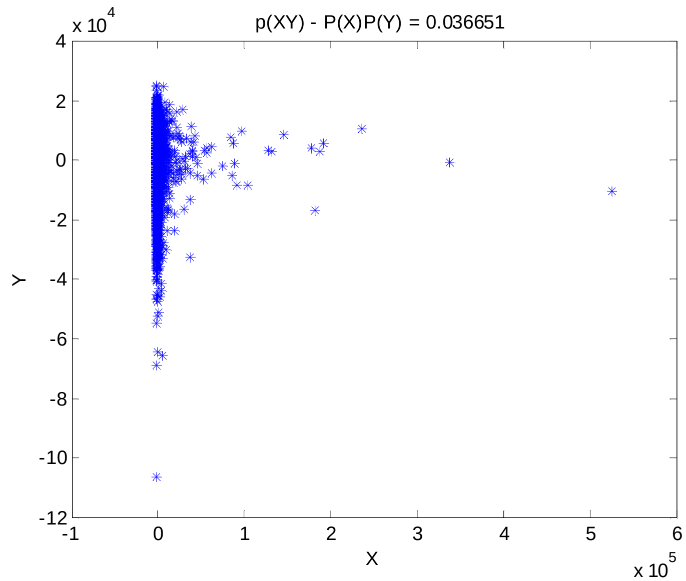
STATISTICAL INDEPENDENCE…CONTD…

Measuring P(XY)-P(X)P(Y)



X->Y

P(XY)-P(X)P(Y) = 0.09



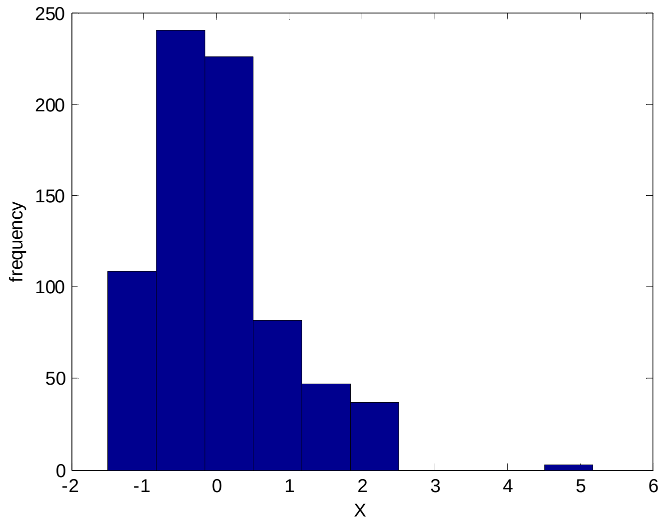
X ⊥ Y

P(XY)-P(X)P(Y) = 0.04

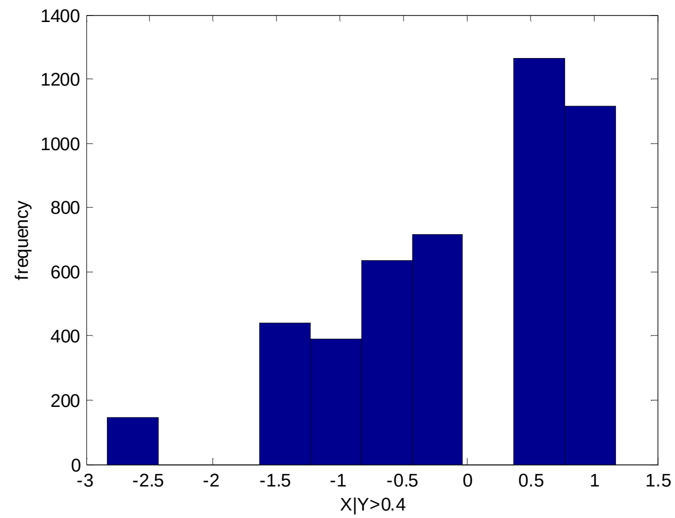
X->Y VS. Y->X

CAUSALITY & DIRECTION OF ARROWS

CONDITIONAL PROBABILITY



P(X)



P(X|Y)

Does the presence of another variable alter the distribution of X?

- **P(cause and effect) more likely explained by P(cause)P(effect|cause) as compared to P(effect)P(cause|effect)**
- ALSO
- if (PX)=P(X|Y) it may indicates that X is independent of Y

Determining The Direction Of Arrows

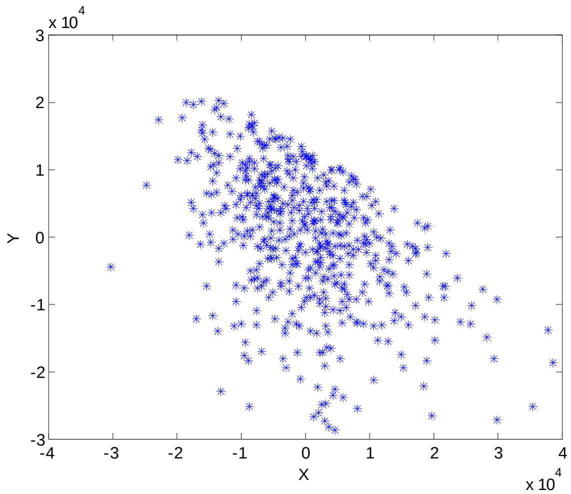
ANM	Fit $Y=f(X)+e_x$ check independence of X and e_x to determine strength of X->Y
PNL	Fit $Y=g(f(X)+e_x)$ and check independence of X and e_x
IGCI	If X->Y then KL-divergence between P(Y) and a reference distribution is greater than KL-divergence between P(X) and a reference distribution
GPI-MML ANM-MML ANM-GAUSS	Likelihood of observed data given X->Y is inversely related to the complexity of P(X) and P(Y|X)
LINGAM	Fit $Y=aX+e_x$ and $X=bY+e_Y$ X->Y if a>b

Note: There are assumptions associated with each method, not stated here

REF: Statnikov *et al.*, new methods for separating causes from effects in genomics data, BMC Genomics, 2012

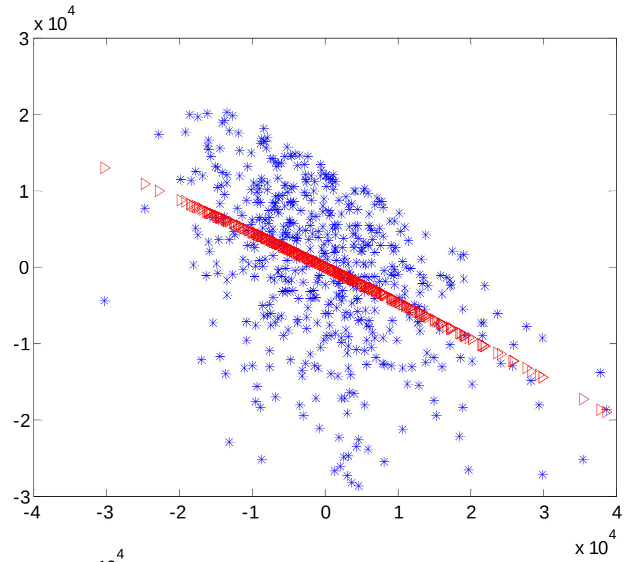
USING REGRESSION

Determine the direction of causality idea behind ANM …

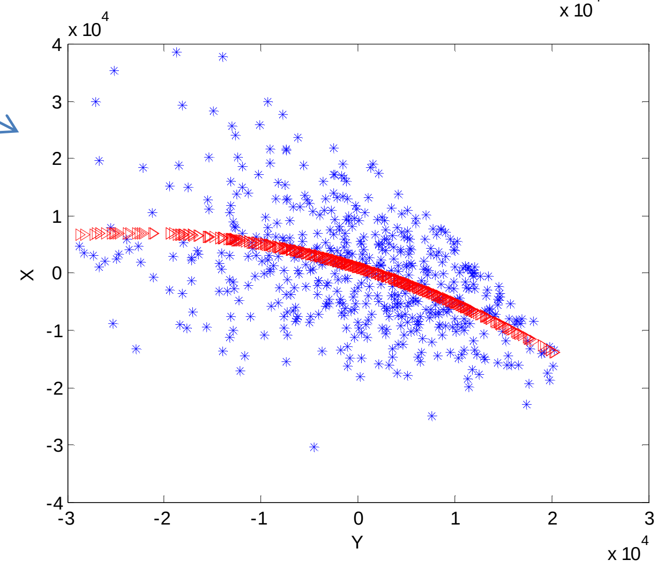


Truth: X->Y

Fit $Y=f(X)+e_x$

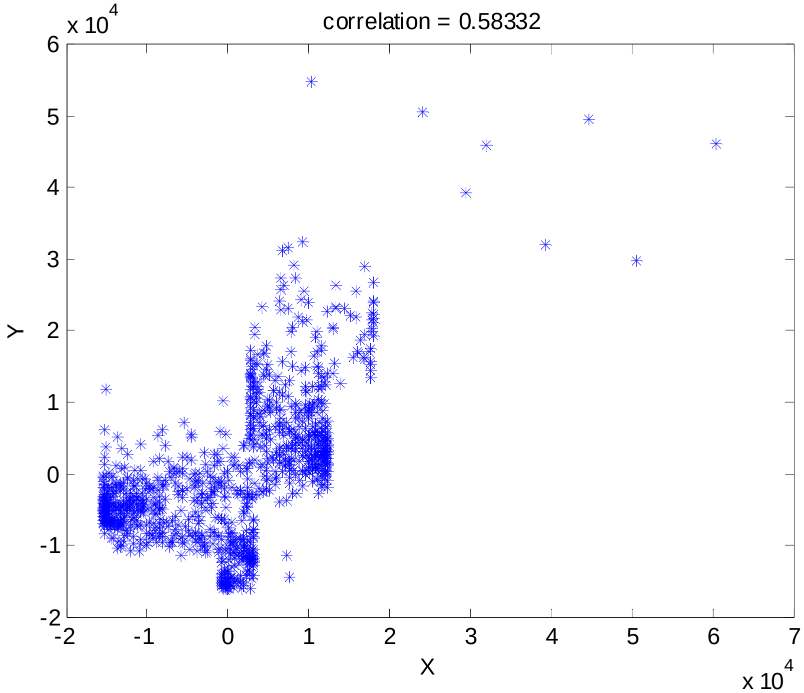


Fit $X=f(Y)+ e_y$



Check the independence of X and e_x and Y and e_y

IDEA BEHIND LINGAM…



y=0.58x-0.02

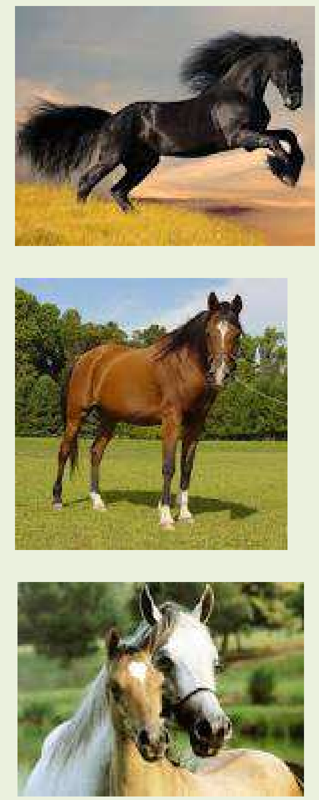
x=.6y+0.01

truth: Y->X

OUTLINE

- Causality
- Coefficients for computing causality
  - Independence measures
  - Probabilistic
  - Determining the direction of arrows
- Transfer learning
- Causality challenge
- Conclusions

TRANSFER LEARNING



Can we use our knowledge from one problem and transfer it to another???

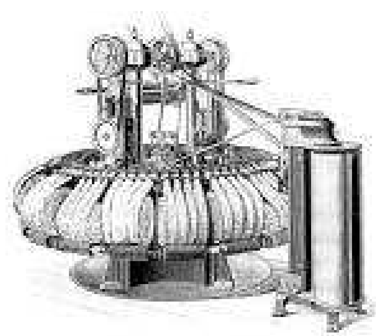
REF: Pan and Yang, A survey on transfer learning, IEEE TKDE, 22(10), 2010.

TRANSFER LEARNING…ONE POSSIBLE VIEW

SOURCE DOMAIN

Lots of labeled data
Truth values are known

feature construction



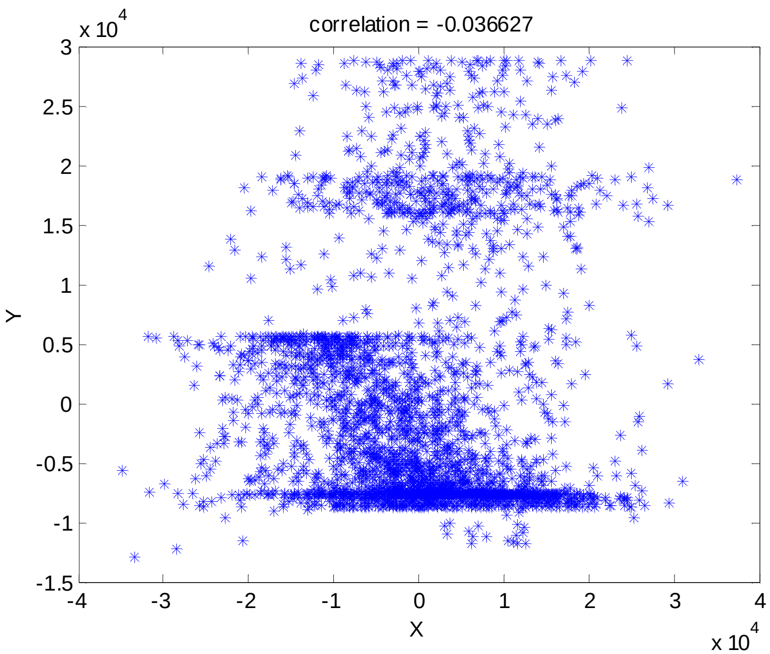
Classification machine

Output labels

TARGET DOMAIN

same features

CAUSALITY & FEATURE CONSTRUCTION FOR TRANSFER LEARNING



If we know the truth values for X and Y relationship then construct features such as:

independence based:
correlation
chi square and so on
causality based
IGCI
ANM
PNM and so on
statistical
percentiles
medians and so on
machine learning
errors of prediction and so on

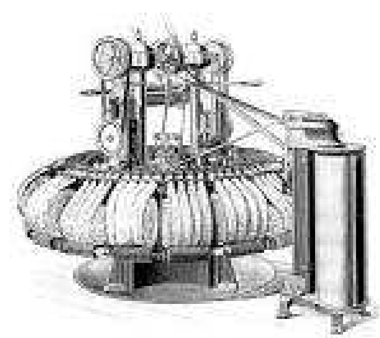
CAUSALITY AND TRANSFER LEARNING…
THE WHOLE PICTURE

PAIR 1		PAIR 2		PAIR 3	
X->Y		Y->X		X⊥Y	
0.1215	0.1855	0.307	-0.064	0.0225	0.6551
0.1557	0.3448	0.5005	-0.1891	0.0537	0.4515
0.1692	0.2291	0.3983	-0.06	0.0388	0.7383
0.1114	0.3994	0.5108	-0.288	0.0445	0.2788
0.1947	0.3059	0.5006	-0.1113	0.0596	0.6363
0.3416	0.2861	0.6278	0.0555	0.0978	1.1939
0.2519	0.4929	0.7449	-0.241	0.1242	0.5111
0.1769	0.1232	0.3002	0.0537	0.0218	1.4356

PAIR 1	LABEL	CORR	IG	CHI-SQ	ANM	…
PAIR 2	LABEL	CORR	IG	CHI-SQ	ANM	…
PAIR 3	LABEL	CORR	IG	CHI-SQ	ANM	…

PAIR i		PAIR j		PAIR k	
unknown		unknown		unknown	
0.0783	0.5261	0.6045	-0.4478	0.0412	0.1488
0.0902	0.2827	0.3728	-0.1925	0.0255	0.319
0.125	0.5065	0.6314	-0.3815	0.0633	0.2468
0.1408	0.3727	0.5135	-0.232	0.0525	0.3777
0.4615	0.4928	0.9543	-0.0314	0.2274	0.9364

features



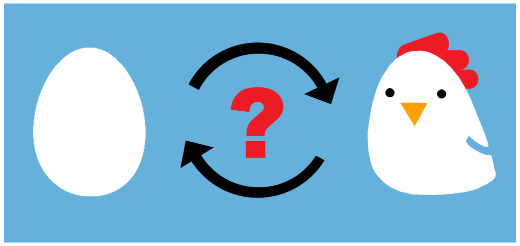
Classification machine

Output

OUTLINE

- Causality
- Coefficients for computing causality
  - Independence measures
  - Probabilistic
  - Determining the direction of arrows
- Transfer learning
- Causality challenge
- Conclusions

CAUSE EFFECT PAIRS CHALLENGE



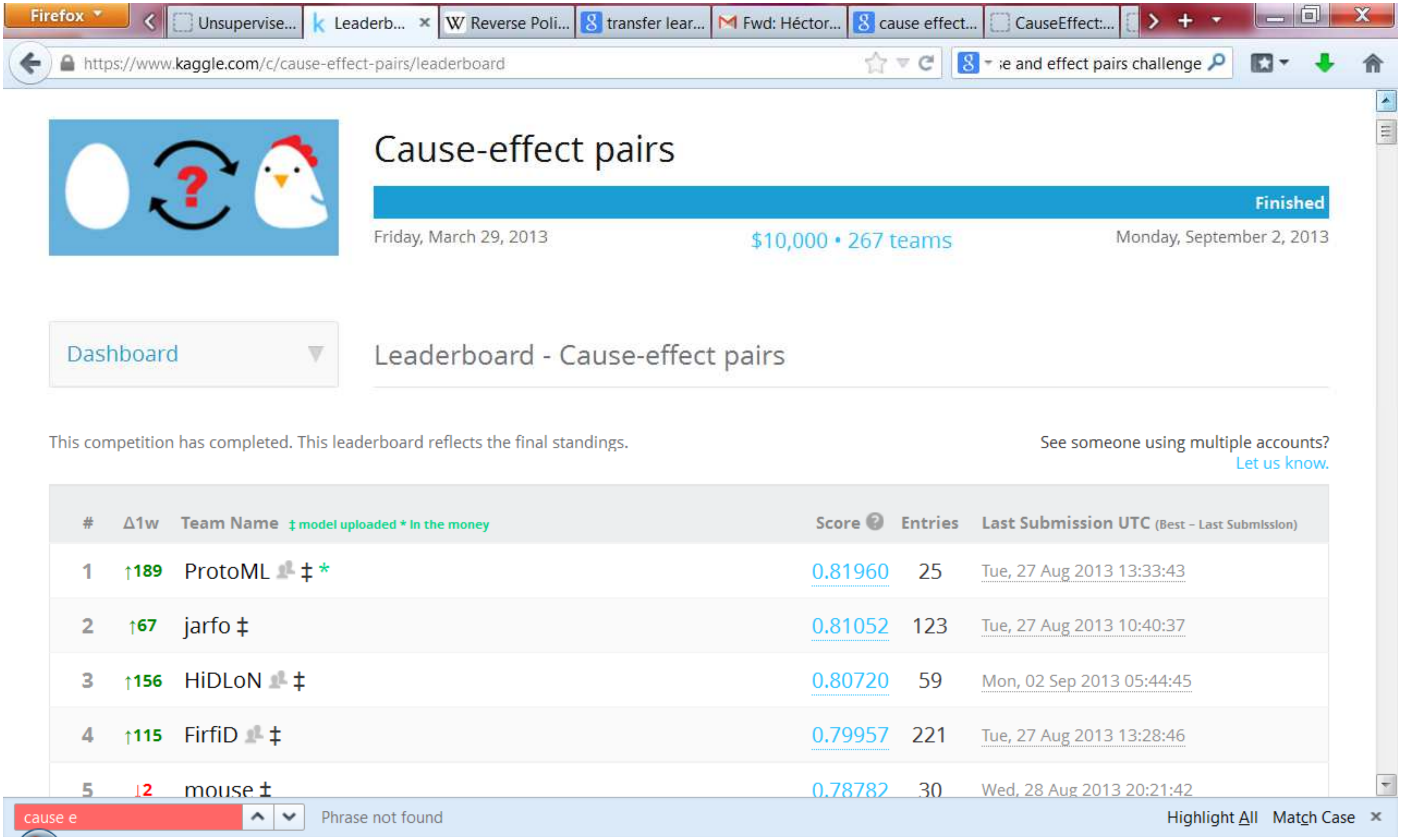
1	SampleID	A	B
2	train1	2092 1143 390 1424 1277 1833 905 980 1488 1451	5651 4449 4012 6124 7310 7608 6201 4618
3	train2	3158 3158 3684 3684 6315 3158 3158 7368 8420 7	2 4 2 1 2 2 2 2 2 2 2 2 2 4 2 4 4 4 2 2 2 2 2 4
4	train3	1699 1808 707 1498 1585 725 1200 1262 1645 855	0 1 0 1 0 1 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 1

Generated from artificial and real data (geography, demographics, chemistry, biology, etc.:
Training Data: 4050 pairs (truth values : known)
Validation Data: 4050 pairs (truth values : unknown)
Test Data: 4050 pairs (truth values : unknown)
Can be categorical, numerical or binary

Identity of variables in all cases: unknown

REF: Guyon, Results and analysis of the 2013 ChaLearn cause-effect pair challenge, NIPS 2013.
REF: http://www.causality.inf.ethz.ch/cause-effect.php

CAUSE EFFECT PAIRS CHALLENGE



https://www.kaggle.com/c/cause-effect-pairs

WHAT WERE THE BEST METHODS

Pre-processing: Smoothing, binning, transforms, noise removal etc.

Feature extraction: Independence, entropy, residuals, statistical features etc.

Dimensionality reduction: Feature selection, PCA, ICA, clustering

Classifier : Random forests, decision trees, neural networks etc.

REF: Guyon, Results and analysis of the 2013 ChaLearn cause-effect pair challenge, NIPS 2013.

INTERESTING RESULTS…
TRANSFER LEARNING

	NO RETRAINING	RETRAINING
Jarfo	0.87	0.997
FirfiD	0.60	0.984
ProtoML	0.81	0.990

3648 gene network cause effect pairs from Ecoli regulatory network

REF: Guyon, Results and analysis of the 2013 ChaLearn cause-effect pair challenge, NIPS 2013.

REF: http://gnw.sourceforge.net/dreamchallenge.html

CONCLUSIONS

- In many cases just one causal coefficient is not enough and so you may have to train a classifier with multiple causal features
- Research on causal inference from the past decade has shown that it is possible to isolate cause and effect pairs from observational data, to a great extent

THANK YOU

REFERENCES

1. Statnikov *et al.*, new methods for separating causes from effects in genomics data, BMC Genomics, 2012.
2. NIPS 2013 Workshop on Causality
   http://clopinet.com/isabelle/Projects/NIPS2013/
3. Pan and Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering, 22(10), 2010
4. Kaggle website on machine learning challenges and cause effect pairs challenge, www.kaggle.com
5. All datasets are taken from the causality challenge:
   https://www.kaggle.com/c/cause-effect-pairs