Distinguishing Cause and Effect:

Inferring causal direction in two variables

Lohit Jain (11390) lohit@cse.iitk.ac.in

Balram Meena (11189) balram@cse.iitk.ac.in

Guide: Prof. Amitabha Mukerjee amit@cse.iitk.ac.in

Dept of Computer Science and Engineering, IIT Kanpur

Introduction

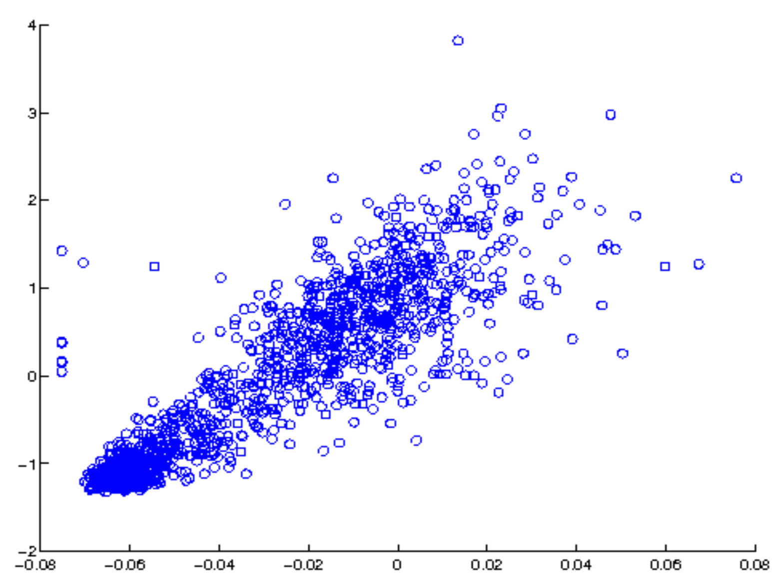
The problem of attributing causes to effects is common in almost every aspects of our everyday life involving human reasoning and decision making.

Detecting causal directions using such data can possibly be a better solution in terms of time and effort.
Consider for instance a target variable B, like occurrence of "lung cancer" in patients. The goal would be to find whether a factor A, like "smoking", might cause B. This information can be further used to detect the best way forward in such social problems.

Problem

The problem is limited to pairs of variables deprived of their context, rendering constraint-based methods (relying on conditional independence tests and/or graphical models) non applicable.

Are you skeptic? Try guessing for this plot:



X Axis - Bytes sent at minute t.
Y Axis - Open http connections at the moment t.

Given distribution of two variables, say X and Y, determine the causal relation between them. The relation can be X = > Y, Y => X, X | Y and X-Y.

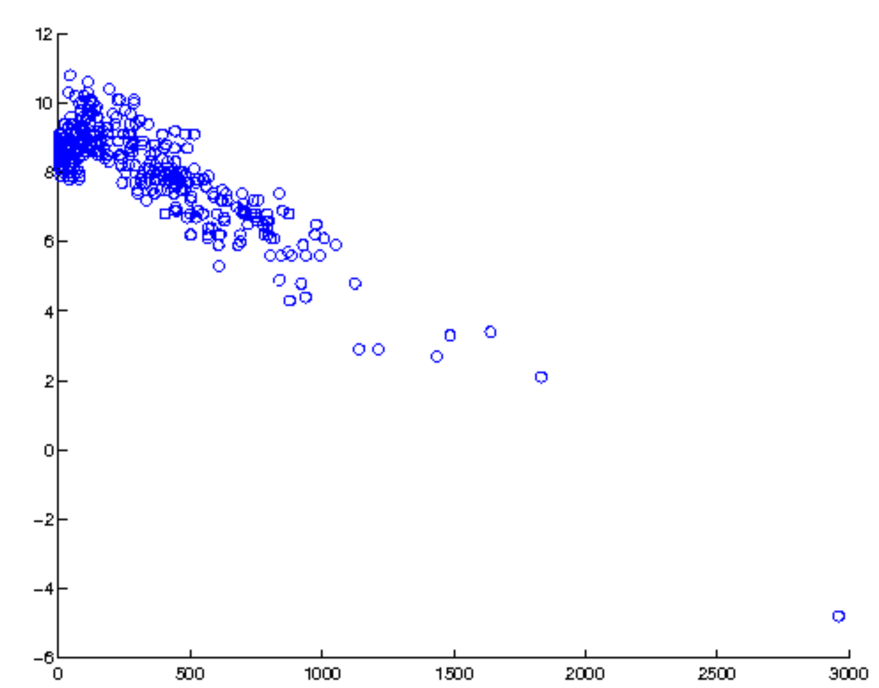
Additionally, there is no feedback looping, explicit time information and the pairs are independent of each other.

There are 4000 sets of such pairs.

Output is confidence score between +inf and -inf for a pair, where +ve indicates X => Y, -ve Y => X and 0 others. The task is to maximize the AUC of ROC for the dataset.

Related Works

- Constraint-based approaches (Pearl, 2000; Spirtes et al., 1993)
- Hoyer et al. [2009] introduce additive noise models, modeling the causal direction X => Y as Y=f(X)+E.
- Generalization of above model is the post-nonlinear model of Zhang and Hyv arinen [2009] where an additional non-linear transformation on noise and effect is allowed, Y=h(f(X)+E).
- Latent variable models (LINGAM): Model as linear combination of cause and noise, Y=pX+qE.
- Complexity-based models: GPI. Applies general non-parametric priors, assuming Y=f(X,E). The causal direction can then be inferred by using standard Bayesian model selection.
  IGCI : function and the probability density of the cause are chosen independently, then the distribution of the effect will, in a certain sense, depend on the function. This is used to generate non symmetry to detect causal direction.



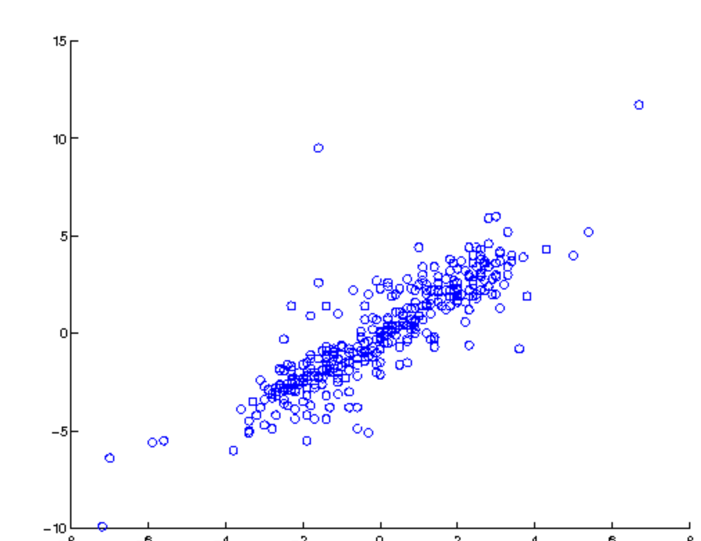
The above plot is between
x axis: altitude
y axis: temperature (average over 1961-1990)
for German Cities

Our Result is 3.1763, ie. A=> B

Methodology

The three major steps involved:

- Evaluate GPI and IGCI models:
  - Compute cost involved in applying gpi and igci models to the each pair of data using unsupervised learning.
    - *IGCI (Information Geometry Causality Inference)*: Involves differences in logarithms of step distributions of the values among the two variables.
    - *GPI (Gaussian Process Inference)*: Gaussian processes (GPs) are used for Bayesian non-parametric estimation of latent functions.
      - The gp function for Bayesian inference and prediction with Gaussian processes for scalar targets.
      - The negative log marginal likelihood and its partial derivatives wrt hyperparameters is computed. The residual hyperparameters are used as inital noise.
  - Store cost of applying in both directions (X=>Y and Y=>X) along with the minimum description length.
- Extract features and train classifier:
  - Extract features.
    Gini-coefficient, normalized hsic, normalized entropy, kurtosis, skewness, divergence, count and unique counts, fit and fit error, joint entropy, discrete mutual information and adjusted mutual information.
  - Use these features and train binary *SVM classifier* (linear kernel) for detecting whether dependency exists.
- Combine results:
  - 0, If the classifier detects pair are not causally related.
  - Combine the cost and other feature values to output confidence. Multiply the cost of gpi and result of igci. Combine with fit and normalized to get the output.



The above plot is between
X Axis : Average annual rate of change of population.
Y Axis : Average annual rate of change of total dietary consumption for total population (kcal/day).
Our Result is -0.241, ie. B=>A. Wrong!

Results

- The AUC of ROC on data was found to be 0.658.
- The accuracy of independence classifier was around 73%.
- The python implementation baseline was 0.570, whereas the top 3 were 0.819, 0.810 and 0.807.

Conclusions

- GPI and IGCI have application in causal direction inference in two numeric variables.
- Ensemble of GPI and IGCI models work better on real life data due to the noise models involved in the two.
- Application of SVM classifier to detect the absence of causal direction works reasonably well.
- IGCI works slightly better than GPI on noise free simulated data.

References

- Mooij, Joris M., et al. "Probabilistic latent variable models for distinguishing between cause and effect." NIPS. 2010.
- Daniusis, Povilas, et al. "Inferring deterministic causal relations." arXiv preprint arXiv:1203.3475 (2012).
- Hoyer, Patrik O., et al. "Nonlinear causal discovery with additive noise models." NIPS. Vol. 21. 2008.
- Peters, Jonas, Dominik Janzing, and Bernhard Scholkopf. "Causal inference on discrete data using additive noise models." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.12 (2011): 2436-2450.
- Janzing, Dominik, et al. "Information-geometric approach to inferring causal directions." Articial Intelligence 182 (2012): 1-31.
- CauseEffect: Challenge - Causality Workbench. 2014. [ONLINE] Available at: http://www.causality.inf.ethz.ch/cause-effect.php. [Accessed 14 April 2014].
- Causal inference at the Max Planck Institute for Intelligent Systems Tübingen. 2014. Causal inference at the Max Planck Institute for Intelligent Systems Tübingen. [ONLINE] Available at: http://webdav.tuebingen.mpg.de/causality/. [Accessed 14 April 2014].