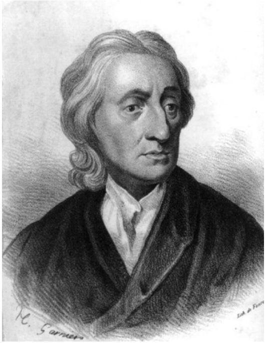
Causality and directed acyclic graphs

Get to know and have some intuition about

- Causality in philosophy (of science)
- Conditional independence relations
- Causal discovery algorithms
- Confounds and back doors

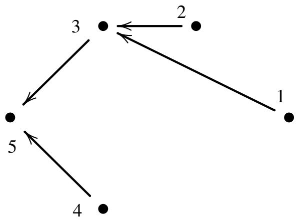


David Hume (1711-1776)

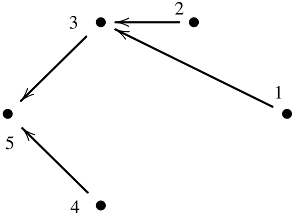
Inductive Causation algorithm

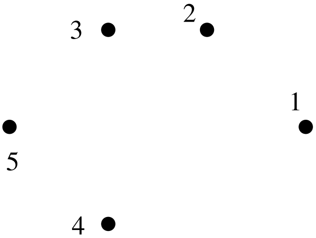
Inductive Causation algorithm

Inductive Causation algorithm

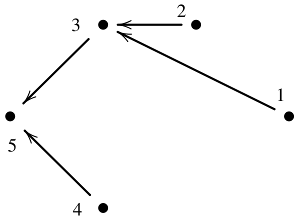


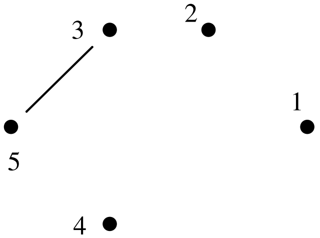
Inductive Causation algorithm



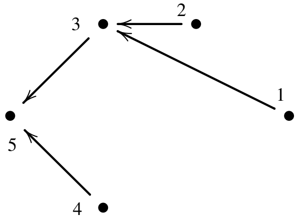


Inductive Causation algorithm

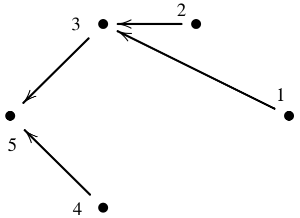




Inductive Causation algorithm

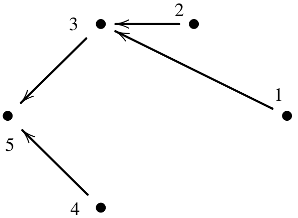


Inductive Causation algorithm



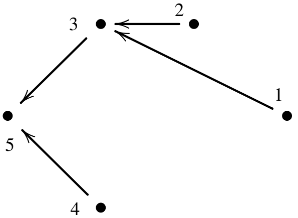
	$r(3,5)$	$r(3,5 \mid 1)$	$r(3,5 \mid 2)$	$r(3,5 \mid 4)$	$r(3,5 \mid 1,2)$	$\cdots$
$3-5$						

Inductive Causation algorithm



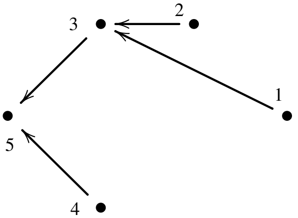
	$r(3,5)$	$r(3,5\|1)$	$r(3,5\|2)$	$r(3,5\|4)$	$r(3,5\|1,2)$	$\cdots$
$3-5$	TRUE					

Inductive Causation algorithm



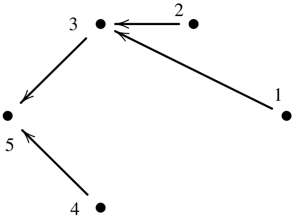
	$r(3,5)$	$r(3,5\|1)$	$r(3,5\|2)$	$r(3,5\|4)$	$r(3,5\|1,2)$	$\cdots$
$3-5$	TRUE	TRUE				

Inductive Causation algorithm



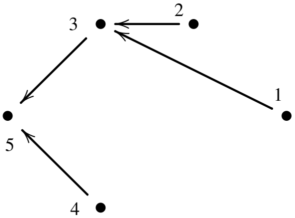
	$r(3,5)$	$r(3,5\|1)$	$r(3,5\|2)$	$r(3,5\|4)$	$r(3,5\|1,2)$	$\cdots$
$3-5$	TRUE	TRUE	TRUE			

Inductive Causation algorithm



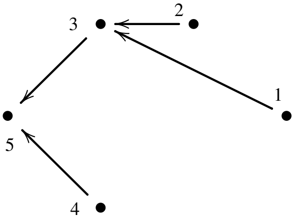
	$r(3,5)$	$r(3,5\|1)$	$r(3,5\|2)$	$r(3,5\|4)$	$r(3,5\|1,2)$	$\cdots$
$3-5$	TRUE	TRUE	TRUE	TRUE		

Inductive Causation algorithm



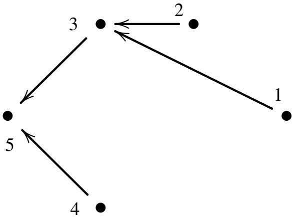
	$r(3,5)$	$r(3,5 \mid 1)$	$r(3,5 \mid 2)$	$r(3,5 \mid 4)$	$r(3,5 \mid 1,2)$	$\cdots$
$3-5$	TRUE	TRUE	TRUE	TRUE	TRUE	

Inductive Causation algorithm



	$r(3,5)$	$r(3,5\|1)$	$r(3,5\|2)$	$r(3,5\|4)$	$r(3,5\|1,2)$	$\cdots$
$3-5$	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

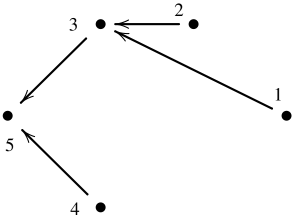
Inductive Causation algorithm



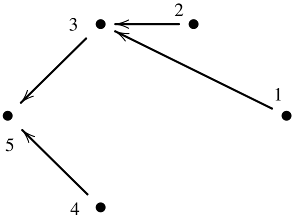
	$r(3,5)$	$r(3,5\|1)$	$r(3,5\|2)$	$r(3,5\|4)$	$r(3,5\|1,2)$	$\cdots$
$3 - 5$	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

There is no (set of) node(s) x such that $(3 \perp\!\!\!\perp 5 \mid x)$ holds; and so connection $3 - 5$ is TRUE

Inductive Causation algorithm

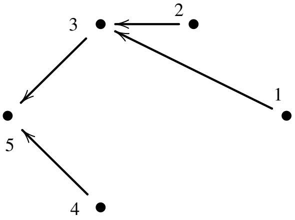


Inductive Causation algorithm



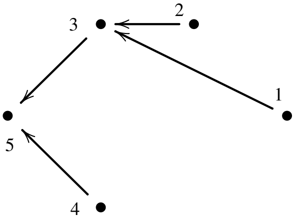
	$r(3,4)$	$r(3,4\|1)$	$r(3,4\|2)$	$r(3,4\|5)$	$r(3,4\|1,2)$	$\cdots$
$3-4$						

Inductive Causation algorithm



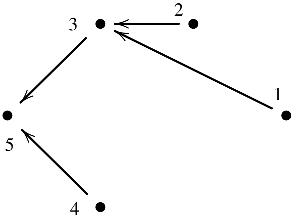
	$r(3,4)$	$r(3,4\|1)$	$r(3,4\|2)$	$r(3,4\|5)$	$r(3,4\|1,2)$	$\cdots$
$3 - 4$	FALSE					

Inductive Causation algorithm



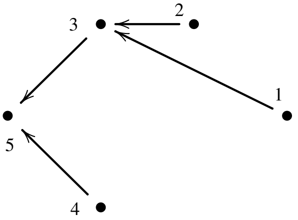
	$r(3,4)$	$r(3,4 \mid 1)$	$r(3,4 \mid 2)$	$r(3,4 \mid 5)$	$r(3,4 \mid 1,2)$	$\cdots$
$3-4$	FALSE	FALSE				

Inductive Causation algorithm



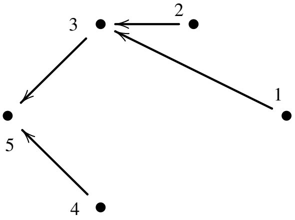
	$r(3,4)$	$r(3,4\|1)$	$r(3,4\|2)$	$r(3,4\|5)$	$r(3,4\|1,2)$	$\cdots$
$3-4$	FALSE	FALSE	FALSE			

Inductive Causation algorithm



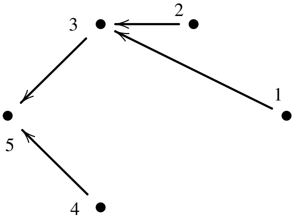
	$r(3,4)$	$r(3,4\|1)$	$r(3,4\|2)$	$r(3,4\|5)$	$r(3,4\|1,2)$	$\cdots$
$3-4$	FALSE	FALSE	FALSE	TRUE		

Inductive Causation algorithm



	$r(3,4)$	$r(3,4\|1)$	$r(3,4\|2)$	$r(3,4\|5)$	$r(3,4\|1,2)$	$\cdots$
$3-4$	FALSE	FALSE	FALSE	TRUE	FALSE	

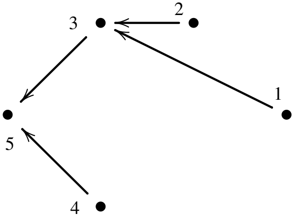
Inductive Causation algorithm



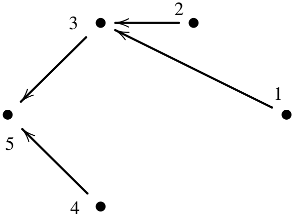
	$r(3,4)$	$r(3,4\|1)$	$r(3,4\|2)$	$r(3,4\|5)$	$r(3,4\|1,2)$	$\cdots$
$3 - 4$	FALSE	FALSE	FALSE	TRUE	FALSE	

No correlation between 3 and 4 and so no (direct) connection $3 - 4$, but conditioning on 5 gives correlation $r(3,4)$ and so a collider $3 \rightarrow 5 \leftarrow 4$.

Inductive Causation algorithm

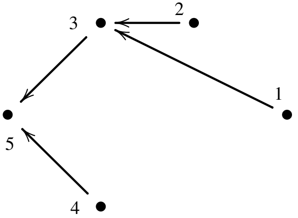


Inductive Causation algorithm



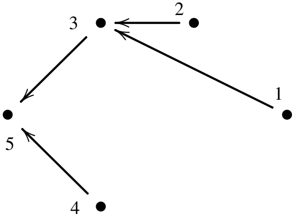
	$r(1,4)$	$r(1,4\|2)$	$r(1,4\|3)$	$r(1,4\|5)$	$r(1,4\|3,5)$
$1-4$					

Inductive Causation algorithm



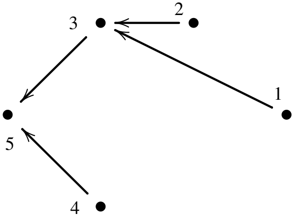
	$r(1,4)$	$r(1,4 \mid 2)$	$r(1,4 \mid 3)$	$r(1,4 \mid 5)$	$r(1,4 \mid 3,5)$
$1 - 4$	FALSE				

Inductive Causation algorithm



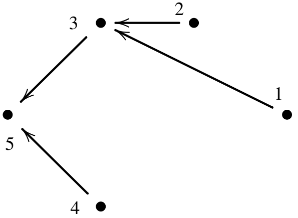
	$r(1,4)$	$r(1,4 \mid 2)$	$r(1,4 \mid 3)$	$r(1,4 \mid 5)$	$r(1,4 \mid 3,5)$
$1-4$	FALSE	FALSE			

Inductive Causation algorithm



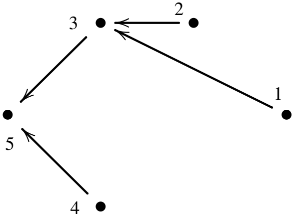
	$r(1,4)$	$r(1,4\|2)$	$r(1,4\|3)$	$r(1,4\|5)$	$r(1,4\|3,5)$
$1 - 4$	FALSE	FALSE	FALSE		

Inductive Causation algorithm



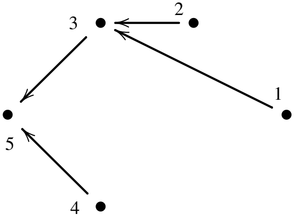
	$r(1,4)$	$r(1,4\|2)$	$r(1,4\|3)$	$r(1,4\|5)$	$r(1,4\|3,5)$
$1-4$	FALSE	FALSE	FALSE	TRUE	

Inductive Causation algorithm



	$r(1,4)$	$r(1,4 \mid 2)$	$r(1,4 \mid 3)$	$r(1,4 \mid 5)$	$r(1,4 \mid 3,5)$
$1-4$	FALSE	FALSE	FALSE	TRUE	

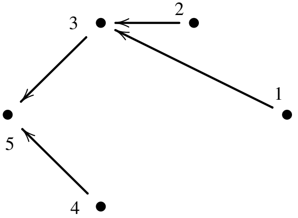
Inductive Causation algorithm



	$r(1,4)$	$r(1,4\|2)$	$r(1,4\|3)$	$r(1,4\|5)$	$r(1,4\|3,5)$
$1 - 4$	FALSE	FALSE	FALSE	TRUE	

There is no correlation between 1 and 4, and so no (direct) connection $1 - 4$.

Inductive Causation algorithm



	$r(1,4)$	$r(1,4 \mid 2)$	$r(1,4 \mid 3)$	$r(1,4 \mid 5)$	$r(1,4 \mid 3,5)$
$1-4$	FALSE	FALSE	FALSE	TRUE	FALSE

There is no correlation between 1 and 4, and so no (direct) connection $1-4$.
No collider since conditioning on 3 and 5 removes the correlation again.

Inductive Causation algorithm

IC-Algorithm Pearl (1988)

Inductive Causation algorithm

IC-Algorithm Pearl (1988)

Input $\hat{P}$ a sampled distribution

Inductive Causation algorithm

IC-Algorithm Pearl (1988)

Input $\hat{P}$ a sampled distribution

Output some acyclic graph for $\hat{P}$

Inductive Causation algorithm

IC-Algorithm Pearl (1988)

Input $\hat{P}$ a sampled distribution

Output some acyclic graph for $\hat{P}$

1. For each pair a and b, look for $(a \perp\!\!\!\perp b \mid S_{ab})$. If no such S_{ab} exists, then a and b are dependent.

Inductive Causation algorithm

IC-Algorithm Pearl (1988)

Input $\hat{P}$ a sampled distribution

Output some acyclic graph for $\hat{P}$

1. For each pair a and b, look for $(a \perp\!\!\!\perp b \mid S_{ab})$. If no such S_{ab} exists, then a and b are dependent.
2. For each trio (a, b, c) such that $a - c - b$ check if c belongs to S_{ab}. If so, then nothing. If c is not in S_{ab} then make a collider at c, i.e. $a \rightarrow c \leftarrow b$.

Inductive Causation algorithm

IC-Algorithm Pearl (1988)

Input $\hat{P}$ a sampled distribution

Output some acyclic graph for $\hat{P}$

1. For each pair a and b, look for $(a \perp\!\!\!\perp b \mid S_{ab})$. If no such S_{ab} exists, then a and b are dependent.
2. For each trio (a, b, c) such that $a - c - b$ check if c belongs to S_{ab}. If so, then nothing. If c is not in S_{ab} then make a collider at c, i.e. $a \rightarrow c \leftarrow b$.
3. Orient as many of the undirected edges as possible, subject to: (i) no new v-structures and (ii) no cycles.

Inductive Causation algorithm

IC-algorithm with multivariate normal (Gaussian) data

$$X \longrightarrow Z \longrightarrow Y \qquad\qquad \Leftrightarrow \rho(X, Y \mid Z), \ldots$$

Inductive Causation algorithm

IC-algorithm with multivariate normal (Gaussian) data

For any X, Y, and Z that have a multivariate normal distribution, we have

$$X \rightarrow Z \rightarrow Y \qquad \Leftrightarrow \rho(X, Y \mid Z), \ldots$$

Inductive Causation algorithm

IC-algorithm with multivariate normal (Gaussian) data

For any X, Y, and Z that have a multivariate normal distribution, we have

1. If X and Y are independent conditional on Z then they are conditionally uncorrelated, i.e. the partial correlation is 0.

$$X \longrightarrow Z \longrightarrow Y \qquad \Leftrightarrow \rho(X, Y \mid Z), \ldots$$

Inductive Causation algorithm

IC-algorithm with multivariate normal (Gaussian) data

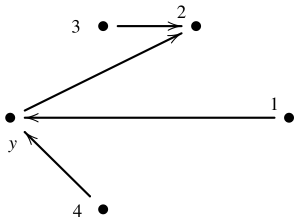
For any X, Y, and Z that have a multivariate normal distribution, we have

1. If X and Y are independent conditional on Z then they are conditionally uncorrelated, i.e. the partial correlation is 0.
2. If X and Y given Z have a partial correlation of 0, then they are conditionally independent.

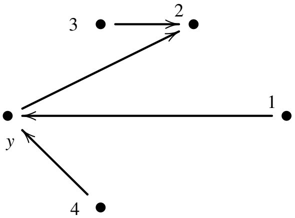
$$X \longrightarrow Z \longrightarrow Y \qquad\qquad \Leftrightarrow \rho(X, Y \mid Z), \ldots$$

What's wrong with regression?

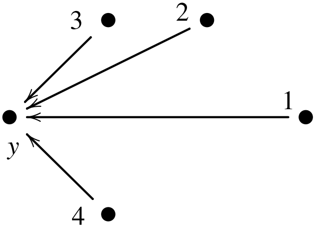
What's wrong with regression?



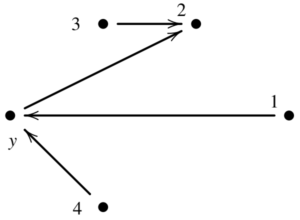
What's wrong with regression?



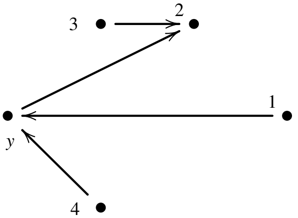
regression framework



What’s wrong with regression?



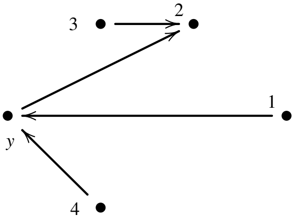
What's wrong with regression?



regression framework

$r(y, 3 \mid 1, 2, 4)$
$y - 3$

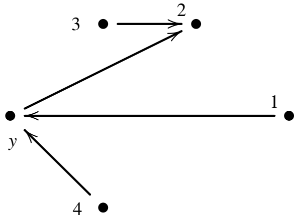
What's wrong with regression?



regression framework

	$r(y,3\|1,2,4)$
$y-3$	TRUE

What's wrong with regression?

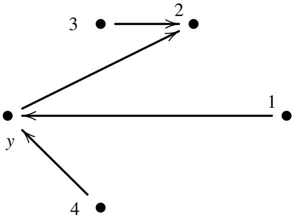


regression framework

	$r(y, 3 \mid 1, 2, 4)$
$y - 3$	TRUE

False inference that 3 has influence on y

What's wrong with regression?



regression framework

	$r(y, 3 \mid 1, 2, 4)$
$y - 3$	TRUE

False inference that 3 has influence on y
Solution: Verify all combinations of partial correlations!