Introducción a la Inferencia Causal

Modelos Gráficos Causales

Sebastián Bejos

8 de julio de 2019

Instituto Nacional de Astrofísica Óptica y Electrónica

Manipulación contra condicionamiento

Ejemplo (Compañía aseguradora)

Supongamos una población con las siguientes v.a. en el tiempo t.

rw_t : *número promedio de copas de vino tinto consumidas por día en 5 años antes que t.*

bmi_t : *índice de masa corporal de una persona en el tiempo t.*

sex : *0=mujer 1=hombre.*

ha_t : *si el individuo a tenido un ataque al corazón o no, en los 5 años antes de t.*

Manipulación contra condicionamiento

Ejemplo

1. *Supongamos que una aseguradora desea determinar qué tasa cobrar a un individuo por un seguro médico en el tiempo* t*, con* $rw_t = 1$, $bmi_t = 25$*, y* $sex = 0$*, donde la* **tasa está basada en la probabilidad de tener un ataqué al corazón en los próximo 5 años**.
2. *Asumiendo que la densidad es* **estable a través del tiempo para los próximos 5 años**, *entonces se podría calcular la tasa en base a:*

$$P(ha_t \mid rw_{t-5} = 1, bmi_{t-5} = 25, sex = 0)$$

Manipulación contra condicionamiento

Problema (Modelación predicativa)
Entrada: *Muestras de una densidad* $P(\mathbf{O})$ *(Donde* $\mathbf{O}$ *es un conjunto de variables aleatorias observadas) y dos conjuntos de variables* $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{O}$.
Salida: *Una estimación eficiente y consistente de* $P(\mathbf{Y} \mid \mathbf{X})$.

Manipulación contra condicionamiento

Ejemplo (Epidemiólogo)

- *Supongamos que un epidemiólogo está decidiendo si recomendar o no, incentivos para que los adultos beban un promedio de 1 a 2 copas de vino tinto al día para prevenir ataques al corazón.*
- *La densidad de los ataques cardiacos, condicionalmente observados a beber un promedio de 1 a 2 vasos de vino tinto al día, no es la densidad relevante para responder a esta pregunta.*
- *Supongamos que beber vino tinto no previene los ataques cardíacos, pero la tasa de ataques cardiacos es más baja entre los bebedores moderados de vino tinto, porque algunas variables socio-económicas causan, tanto el consumo moderado de vino tinto, como otras opciones de estilos de vida saludables, que previenen los ataques cardíacos.*

Manipulación contra condicionamiento

Ejemplo (Epidemiólogo)

- *En ese caso, después de que se establezcan los incentivos para beber vino tinto, la densidad del estatus socio-económico entre los bebedores de vino tinto será diferente a la anterior, y la densidad condicional de ataques cardíacos entre los bebedores moderados de vino tinto no será la misma después de que los incentivos fueron adoptados como antes de su adopción.*
- *Por lo tanto, el uso de densidades condicionales de observación para predecir ataques cardíacos después de que se hayan implementado los incentivos conducirá a predicciones incorrectas.*

Manipulación contra condicionamiento

Ejemplo (Epidemiólogo)

- *La densidad que es relevante para determinar si hacer la recomendación o no, no es la densidad de los ataques cardíacos entre las personas que han elegido beber vino tinto, si no la densidad de los ataques cardíacos entre las personas que beberían vino tinto después de los incentivos.*
- *La densidad de los ataques cardíacos entre las personas que beberían vino tinto, después de que se implementan los incentivos, es aproximadamente igual a la densidad de los ataques cardíacos, entre las personas asignadas a beber cantidades moderadas de vino tinto en un estudio experimental.*
- *Esta es una densidad manipulada, que resulta de tomar medidas en una población determinada.*

Manipulación contra condicionamiento

- Las densidades manipuladas son apropiadas cuando se hacen predicciones sobre los efectos de tomar acciones ("manipular.º "hacer") en una población determinada (por ejemplo, asignar el consumo de vino tinto), en lugar de observar ("ver") los valores de las variables dadas.
- En contraste con las probabilidades condicionales, que pueden estimarse a partir de muestras de una población, el estándar para estimar densidades manipuladas es un experimento, a menudo un ensayo aleatorio. Sin embargo, en muchos casos, los experimentos son demasiado caros, demasiado difíciles o no son éticos de realizar.

Manipulación contra condicionamiento

- Esto plantea el problema de qué se puede determinar acerca de las densidades manipuladas a partir de muestras de una población, posiblemente en combinación con un número limitado de ensayos aleatorios.
- El problema difícil porque la inferencia se hace, a partir de un conjunto de variables aleatorias medidas $\mathbf{O}$, de muestras que pueden no contener variables que sean causa de múltiples variables en $\mathbf{O}$.

Manipulación contra condicionamiento

Problema (Construcción de modelos causales con datos muestrales)
Entrada: *Muestras de una población con densidad* $P(\mathbf{O})$*, un conjunto (posiblemente vacío) de densidades manipuladas* $P(\mathbf{O} \parallel M_1), \ldots P(\mathbf{O} \parallel M_w)$ *y supuestos.*
Salida: *Un conjunto de modelos causales, tan pequeño como sea posible, y que contenga un modelo causal verdadero, el cual contenga al menos las variables en* $\mathbf{O}$*.*

Manipulación contra condicionamiento

Problema (Predicción de los efectos de manipulaciones a partir de modelos causales)
Entrada: *Una densidad no manipulada* $P(\mathbf{O})$, *un conjunto* $\mathbf{C}$ *de modelos causales que contienen al menos las variables de* $\mathbf{O}$, *una manipulación* M, *y los conjuntos* $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{O}$
Salida: *Una función* g *tal que* $g(P(\mathbf{O}), \mathbf{C}, M, \mathbf{X}, \mathbf{Y}) = P(\mathbf{Y} \mid \mathbf{X} \parallel M)$ *si existe alguna, y una salida de no existe la función en otro caso.*

Manipulación contra condicionamiento

Problema (Modelación predictiva contrafactual)

Entrada: *Una densidad no manipulada* $P(\mathbf{O})$*, un conjunto* $\mathbf{C}$ *de modelos causales que contienen al menos las variables de* $\mathbf{O}$*, una manipulación contrafactual* M *y los conjuntos* $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{O}$

Salida: *Una función* g *tal que* $g(P(\mathbf{O}), \mathbf{C}, M, \mathbf{X}, \mathbf{Y}) = P(\mathbf{Y} \mid \mathbf{X} \parallel M)$ *si existe alguna, y una salida de "no existe la función.en otro caso.*

Definición (Condición local de Markov)

Una red bayesiana es un par (G, P), *donde* G *es un grafo acíclico dirigido (DAG), cuyos vértices son variables aleatorias, y* P *es una densidad tal que cada variable* V *en* G, *es condicionalmente independiente de variables que no son descendientes de* V *en* G, *dados los padres de* V *en* G. *En este caso, se dice que* P *satisface la* **condición local de Markov** *para* G.

Definición (Condición global de Markov)

Si $\mathbf{X}$ *es independiente condicional de* $\mathbf{Y}$ *dado* $\mathbf{Z}$ *en* P, *siempre que* $\mathbf{X}$ *es d-separable de* $\mathbf{Y}$ *dado* $\mathbf{Z}$ *en* G, *entonces* P *satisface la* **condición global de Markov** *para* G.

Definición (Factorización sobre un DAG)

Para un conjunto de variables aleatorias $\mathbf{V}$ *en* G*, una densidad* $P(\mathbf{V})$ *se* **factoriza de acuerdo a la DAG** G *si y solo si,*

$$P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P(V \mid \mathbf{Parent}(V, G)),$$

donde $\mathbf{Parent}(V, G)$ *es el conjunto de padres de* V *en* G*.*

Modelos Causales

Definición (Causa directta)

A es **causa directa** *de B, relativa a un conjunto de variables V sobre una población, cuando existen dos manipulaciones de $V \setminus \{B\}$ (Esto es, todas la variables en V, excepto B, son manipuladas con valores específicos) que difieren solamente en los valores de asignados a A y en que producen diferentes densidades para B.*

Definición (DAG Causal)

Una **DAG Causal** *G para una población (conjunto de datos) contiene aristas $A \to B$ si y solo si, A es* **causa directa** *de B, en la población específica.*

Modelos Causales

Definición (Suficiencia causal)
Un conjunto de variables V tiene **suficiencia causal** *si y solo si, no existe una variable C que no está en V tal que sea causa directa de más de una variable en en V.*

Definición (Suposición Markoviana Causal)
Para un conjunto de variables con **suficiencia causal** *V, en una población N con densidad $P(V)$. La densidad $P(V)$ satisface la* **condición local de Markov** *para la DAG causal de N.*

Definición (Regla de manipulación)

Bajo la suposición Markoviana causal, en una **Red Bayesiana Causal**, *una* **manipulación** *de X a $P'(X \mid \mathbf{Y})$ (donde se asume que $\mathbf{Y}$ contiene únicamente no desendientes de X en G), simplemente remplaza el término $P(X \mid$ **Parent(X,G)**$)$ en la factorización de la densidad conjunta por la densidad manipulada $P'(X \mid \mathbf{Y})$, esto es:*

$$P(\mathbf{V} \parallel P'(X \mid \mathbf{Y})) = P'(X \mid \mathbf{Y}) \prod_{V \in \mathbf{V} \setminus \{X\}} P(V \mid \mathbf{Parents}(V, G))$$

Búsqueda de Modelos Causales

La búsqueda de modelos causales se divide en dos partes:

1. Búsqueda de la DAG Causal G.
2. Estimación de los parámetros a partir de datos muéstrales y la gráfica causal G.

Definición (Supuesto de fidelidad causal)
Para un conjunto de variables V con suficiencia causal en una población N, toda relación de independencia condicional verdadera en la densidad sobre V está implicada en la condición local de Markov sobre el DAG causal de N.

Definición (Markov equivalentes)

Dos DAGs diferentes G y G' que tienen el mismo conjunto de relaciones de independencia condicional bajo d-separación se dice que son **Markov equivalentes***.*

Definición (Equivalencia Estadística)

Para cada DAG G, existe un conjunto de densidades $\mathbf{P}(G)$ que satisfacen la condición local de Markov para G. Sea $\mathbf{F} \subset \mathbf{P}$ una sub-familia paramétrica de $\mathbf{P}$ (e. g. el subconjunto Gausiano de $\mathbf{P}$). Dos DAGs G y G' son estadíticamente equivalentes con respecto a $\mathbf{F}$, si y solo si,

$$\mathbf{P}(G) \cap \mathbf{F} = \mathbf{P}(G') \cap \mathbf{F}$$

Definición (Equivalencia causal)

Dos DAGs causales son **equivalentes bajo causalidad** *(con respecto a una familia de densidades* **F***) si y solo si, estos represental al mismo conjunto de densidades (dentro de la familia* **F***) para toda manipulación (incluyendo la manipulación nula)*

Ejemplo

Los DAGs: $A \to B \leftarrow C \leftarrow D$ *y* $A \to B \leftarrow C \to D$ *son Markov equivalentes, pero no son equivalentes bajo causalidad.*

Búsqueda de Modelos Causales

En ausencia de información tal como:

1. Muestras de densidades manipuladas o,
2. Antecedentes de conocimiento del dominio.

Los método estandar de puntuación: Bayesian Information Criterion, Minimum Description Length, etc.

1. Asignan puntuaciones iguales para los DAG en una clase de equivalencia estadística para todos los conjuntos de datos.
2. Más fácil buscar la clase de equivalencia de Markov de los DAG, incluso si se sabe que la clase de equivalencia estadística es un subconjunto propio de la clase de equivalencia de Markov.