



## Redes bayesianas con algoritmos basados en restricciones, scores e híbridos aplicados al problema de clasificación

## Bayesian networks with algorithms based on restrictions, scores and hybrids applied to the classification problem

Carlos López de Castilla Vásquez<sup>1</sup>

<sup>1</sup> Universidad Nacional Agraria La Molina, Lima, Perú. Email: [clopez@lamolina.edu.pe](mailto:clopez@lamolina.edu.pe)

Recepción: 24/04/2018; Aceptación: 05/06/2018

### Resumen

Las redes bayesianas son gráficos acíclicos dirigidos que codifican las relaciones de dependencia e independencia condicional en un conjunto de variables predictoras. En este trabajo de investigación se presentan tres algoritmos que permiten obtener la estructura que define una red bayesiana. Sobre esta estructura se construyeron clasificadores, incluyendo una variable dependiente en el gráfico que tiene las clases o categorías de interés, obteniendo un rendimiento predictivo similar en comparación con los clasificadores por redes bayesianas tradicionales, Naive Bayes y TAN. Se presenta también el algoritmo de selección de variables Statistically Equivalent Signature obteniendo resultados similares a los clasificadores contruidos con todas las variables predictoras.

**Palabras clave:** redes bayesianas; clasificador; Naive Bayes; TAN; independencia condicional; selección de variables.

### Abstract

Bayesian networks are directed acyclic graphs that code the relationships of conditional dependence and independence in a set of predictive variables. In this research work, three algorithms that allow to obtain the structure that defines a bayesian network are presented. Classifiers were built on this structure, including a dependent variable in the graph that has the classes or categories of interest, obtaining a similar predictive performance compared to the classifiers by traditional bayesian networks, Naive Bayes and TAN. The Statistically Equivalent Signature variable selection algorithm is also presented, obtaining similar results to the classifiers constructed with all the predictive variables.

**Keywords:** bayesian networks; classification; Naive Bayes; TAN; conditional independence; variable selection.

**Forma de citar el artículo:** López de Castilla. C. 2019. Redes bayesianas con algoritmos basados en restricciones, scores e híbridos aplicados al problema de clasificación. Anales Científicos 80 (1): 15- 25 (2019).

DOI: <http://dx.doi.org/10.21704/ac.v80i1.1370>

Autor de correspondencia: López de Castilla. C. Email: [clopez@lamolina.edu.pe](mailto:clopez@lamolina.edu.pe)

© Universidad Nacional Agraria La Molina, Lima, Perú.

## 1. Introducción

En muchas áreas de las ciencias, es de interés realizar el proceso de clasificación de un grupo de observaciones, descritas por un conjunto de variables predictoras en diferentes categorías definidas, por una variable respuesta. Uno de los clasificadores bayesianos más populares es Naive Bayes gracias a su simplicidad, eficiencia y bajo error de clasificación (Duda *et al.*, 1973), (Taheri *et al.*, 2014). Otro clasificador bayesiano, basado en el algoritmo de Chow and Liu, es llamado Tree Augmented Naive Bayes, TAN, (Friedman *et al.*, 1997) que considera relaciones de dependencia entre una variable predictora y, a lo sumo, otra. Ambos clasificadores tienen, en términos generales, un buen desempeño predictivo.

Una de las ventajas de usar clasificadores bayesianos es la posibilidad de poder observar de manera sencilla las relaciones de dependencia e independencia condicional que existen entre las variables predictoras (Friedman *et al.*, 1997) a diferencia de aquellos clasificadores, como las redes neuronales, que utilizan algoritmos complejos que son difíciles de entender e interpretar. A pesar de contar en la actualidad con muchos modelos que permiten realizar el proceso de clasificación, estos imponen algunas restricciones sobre el conjunto de variables predictoras, dejando de lado las relaciones naturales que existen entre ellas y que pueden ser aprovechadas para mejorar el comportamiento predictivo del clasificador.

En este trabajo de investigación, se presentan tres algoritmos que permiten obtener la estructura de dependencia e independencia condicional entre las variables predictoras, necesaria para la posterior construcción del clasificador. Además, se presenta el algoritmo Statistically Equivalent Signature, SES, como un método de selección de variables predictoras inspirado en los principios del algoritmo basado en restricciones (Tsamardinos *et al.*, 2013) que permite obtener grupos de variables predictoras con comportamiento predictivo equivalente.

Esta investigación tiene por objetivo evaluar la precisión de los clasificadores por redes bayesianas obtenidos con los algoritmos basados en restricciones, scores e híbridos, en comparación con los

clasificadores tradicionales Naive Bayes y TAN. Para el proceso de selección de variables, se usa el algoritmo Statistically Equivalent Signatures, comparando el poder predictivo de los clasificadores obtenidos antes y después de su aplicación.

## 2. Materiales y métodos

### Tipo de investigación

El trabajo de investigación es de naturaleza cuantitativa, predictiva y correlacional.

### Diseño de la investigación

El diseño es no experimental y de corte transversal, ya que los conjuntos de datos corresponden a un determinado espacio y tiempo.

### Población y muestra

Los conjuntos de datos que se muestran en la [Tabla 1](#) fueron tomados del repositorio UCI Machine Learning (Dheeru y Karra, 2017) y son utilizados frecuentemente para comparar clasificadores por redes bayesianas.

Tabla 1: Conjuntos de datos

N°	Nombre	Variables	Categorías	Tipo
1	Australian	14	2	Cuantitativas y cualitativas
2	Breast	9	2	Cualitativas
3	Corral	6	2	Cualitativas
4	Diabetes	8	2	Cuantitativas
5	German	20	2	Cuantitativas y cualitativas
6	Glass	9	6	Cuantitativas
7	Heart	13	2	Cuantitativas y cualitativas
8	Hepatitis	19	2	Cuantitativas y cualitativas
9	Iris	4	3	Cuantitativas
10	Vehicle	18	4	Cuantitativas
11	Vote	16	2	Cualitativas

### Metodología aplicada

#### Redes bayesianas

Las redes bayesianas son una clase de modelos gráficos que permiten una

representación concisa de la estructura probabilística de datos multivariantes utilizando gráficos. Las redes bayesianas están formadas por lo siguiente:

- Un conjunto de variables aleatorias  $X = \{X_1, X_2, \dots, X_p\}$  que describen las variables de interés. La distribución de probabilidad conjunta de  $X$  se llama la distribución global, mientras que las que se asocian con cada  $X_i \in X$  se llaman distribuciones locales.
- Un gráfico acíclico dirigido, GAD, denotado por  $G=(V,A)$ . Cada nodo  $v \in V$  está asociado con una variable  $X_i$ . Los arcos dirigidos  $a \in A$ , que los conectan representan dependencias probabilísticas directas. Si no existe un arco que conecta dos nodos de las correspondientes variables entonces son independientes o condicionalmente independientes dado un subconjunto de las variables restantes.

Una red bayesiana permite la representación de la distribución de probabilidad conjunta de las variables aleatorias en  $X$  como el producto de las distribuciones locales asociadas con cada variable  $X_i$ . Lo anterior es una aplicación directa de la regla de la cadena:

$$\Pr(X) = \prod_{i=1}^p \Pr(X_i | \Pi(X_i)) \quad (1)$$

donde  $\Pi_{X_i}$  es el conjunto llamado los padres de  $X_i$ .

**Independencia condicional y separación gráfica**

Las relaciones directas e indirectas entre dos variables se pueden obtener desde el GAD comprobando si están conectadas de alguna manera. Si las variables dependen directamente entre sí, habrá un solo arco que conecta sus nodos. Si la dependencia es indirecta, habrá dos o más arcos que pasan a través de los nodos que median la asociación. En general, dos conjuntos de variables  $X$  y  $Y$  son independientes dado un tercer conjunto de variables  $Z$  si no existe un conjunto de arcos que los conecta que no esté bloqueado por las variables condicionantes. En otras palabras, se dice que  $X$  y  $Y$  están  $d$ -separados por  $Z$ .

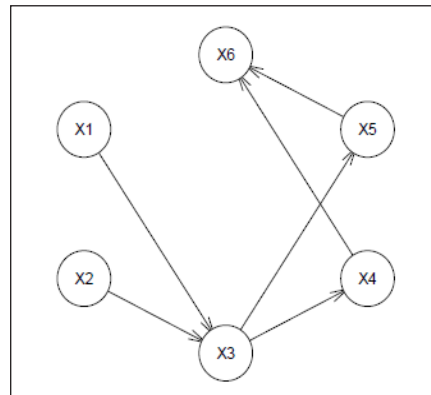


Figura 1: Red bayesiana

**El manto de Markov**

El manto de Markov (Pearl, 1988) representa el conjunto de nodos que  $d$ -separan completamente un nodo específico del resto de nodos en el gráfico. En general, el manto de Markov de un nodo  $A$  es el conjunto formado por los padres de  $A$ , los hijos de  $A$  y todos los demás nodos que compartan un hijo con  $A$ . Si el nodo  $A$  está en el manto de Markov de  $B$ , entonces  $B$  está en el manto de Markov de  $A$ .

**Estimación de una red bayesiana**

La estimación de una red bayesiana se realiza como un proceso de dos pasos:

1. La estimación de la estructura del GAD.
2. La estimación de los parámetros de las distribuciones locales definidas por el GAD.

Sea  $D$  un conjunto de datos y  $B=(G,\theta)$  una red bayesiana, donde  $\theta$  denota los parámetros de la distribución global de  $X$ . El aprendizaje de la red bayesiana se puede formalizar como:

$$\Pr(B | D) = \Pr(G, \theta | D) = \underbrace{\Pr(G | D)}_{\text{est estructura}} \underbrace{\Pr(\theta | G, D)}_{\text{est param}} \quad (2)$$

La descomposición de  $\Pr(G, \theta | D)$  refleja los dos pasos descritos anteriormente. La estimación de la estructura se puede realizar

en la práctica por la búsqueda de  $G$  que maximiza:

$$\Pr(G|D) \propto \Pr(G) \Pr(D|G) = \Pr(G) \int \Pr(D|G, \theta) \Pr(\theta|G) d\theta \quad (3)$$

La elección más común para  $\Pr(G)$  es una a priori no informativa sobre el espacio de los posibles GAD, asignando la misma probabilidad a cada estructura. El cálculo de  $\Pr(D|G)$  es problemático desde el punto de vista computacional y algebraico. Como resultado, se han desarrollado dos alternativas al uso de  $\Pr(D|G)$  en la estimación de la estructura. La primera de ellas es el uso del criterio de información bayesiano BIC:

$$\text{BIC}(G, D) = \sum_{i=1}^p \left[ \log \Pr(X_i | \Pi_{X_i}) - \frac{|\theta_{X_i}|}{2} \log n \right] \quad (4)$$

La segunda alternativa es utilizar pruebas de independencia condicional de información mutua:

$$\text{MI}(X, Y|Z) = \sum_{i=1}^R \sum_{j=1}^C \sum_{k=1}^L \frac{n_{ijk}}{n} \log \frac{n_{ijk} n_{++k}}{n_{ijk} n_{++k}} \quad (5)$$

## Estimación de la estructura

### Algoritmos basados en restricciones

Los algoritmos basados en restricciones se basan en el trabajo de Pearl, en los mapas y su aplicación a los modelos gráficos causales. Su algoritmo de Causalidad Inductiva proporciona un marco para la estimación de la estructura de un GAD usando pruebas de independencia condicional. El algoritmo Grow-Shrink, (Margaritis, 2003) permite obtener la estructura de la red bayesiana identificando en el primer paso el manto de Markov de cada nodo, usando las pruebas de independencia condicional. Los pasos se presentan a continuación:

---

### Algoritmo: Grow-Shrink

---

1. Para cada nodo  $A$  en  $V$  se obtiene su manto de Markov.
  2. Para cada par de nodos  $A$  y  $B$  en  $V$  se busca el conjunto  $SA \subset BV$  tales que  $A$  y  $B$  son independientes dado  $S_{AB}$  donde  $A, B \in S_{AB}$ . Si no existe tal conjunto, poner un arco no dirigido entre  $A$  y  $B$ .
  3. Para cada par de nodos no adyacentes  $A$  y  $B$  con un vecino común  $C$ , comprobar si  $C \in S_{AB}$ . Si esto no es cierto, tomar la dirección de los arcos  $A-C$  y  $C-B$  como  $A \rightarrow C$  y  $C \leftarrow B$ .
  4. Establecer la dirección de los arcos que todavía no se encuentran dirigidos aplicando de forma recursiva las dos reglas siguientes:
    - a. Si  $A$  es adyacente a  $B$  y hay un camino estrictamente dirigido desde  $A$  hacia  $B$  entonces, establecer la dirección de  $A-B$  como  $A \rightarrow B$ .
    - b. Si  $A$  y  $B$  no son adyacentes pero  $A \rightarrow C$  y  $C \rightarrow B$ , entonces  $C \rightarrow B$ .
  5. Devolver el GAD o GAPD resultante.
- 

### Algoritmos basados en scores

Los algoritmos basados en scores representan la aplicación de técnicas de optimización heurística al problema de estimación de la estructura de la red bayesiana. A cada red candidata se asigna un puntaje de acuerdo a su estructura, el cual refleja su grado de bondad de ajuste. Algunos ejemplos de esta clase de algoritmos son: Hill-Climbing que explora el espacio de búsqueda a partir de la estructura de la red y añade, borra o invierte un arco a la vez hasta que el score ya no se pueda, Genetic algorithms que imitan la evolución natural a través de la selección iterativa de los modelos más aptos y Simulated annealing que realiza una búsqueda local aceptando cambios que incrementan o disminuyen el score de la red con una probabilidad inversamente proporcional a la disminución del score.

Los pasos del algoritmo Hill-Climbing se presentan a continuación.

---

 Algoritmo: Hill-Climbing
 

---

1. Elija una estructura de red  $G$  sobre  $V$ , no necesariamente vacía.
  2. Calcular el score de  $G$  denotado por  $\text{Score}_G = \text{Score}(G)$ .
  3. Sea  $\text{maxScore} = \text{Score}_G$ .
  4. Repita los siguientes pasos conforme  $\text{maxScore}$  aumenta:
    - a. Para cada posible arco agregado, eliminado o invertido que no resulte en una red cíclica:
      1. Calcular el score de la red modificada  $G^* \text{Score}_{G^*} = \text{Score}(G^*)$ .
      2. Si  $\text{Score}_{G^*} > \text{Score}_G$ , tomar  $G = G^*$  y  $\text{Score}_G = \text{Score}_{G^*}$ .
    - b. Actualizar  $\text{maxScore}$  con el nuevo valor de  $\text{Score}_G$ .
  5. Devolver el GAD  $G$ .
- 

### Algoritmos híbridos

Los algoritmos híbridos combinan algoritmos basados en restricciones y scores para compensar las respectivas debilidades de cada método y así producir estructuras más confiables en una amplia variedad de situaciones. Los dos miembros más conocidos de esta familia son el algoritmo Sparse Candidate, SC, (Friedman *et al.*, 1999) y el algoritmo Max-Min Hill-Climbing, MMHC, (Tsamardinos *et al.*, 2006). El algoritmo MMHC se ilustra a continuación.

---

 Algoritmo: Max-Min Hill-Climbing
 

---

1. Elija una estructura de red  $G$  sobre  $V$ , no necesariamente vacía.
  2. Realice los siguientes pasos:
    - a. **Fase de restricción:** seleccionar un conjunto  $C_i \subset V$  de los padres e hijos candidatos para cada nodo  $X_i \in V$ .
    - b. **Fase de maximización:** usar el algoritmo Hill-Climbing para encontrar la estructura de la red  $G^*$  que maximiza  $\text{Score}(G^*)$  entre las redes en las que los padre o hijos de cada nodo  $X_i$  se encuentren incluidos en el correspondiente conjunto  $C_i$ .
  3. Devolver el GAD  $G$ .
- 

### Estimación de parámetros

Una vez que la estructura de la red bayesiana se ha obtenido a partir de la data, la tarea de estimación de los parámetros de la distribución global se simplifica en gran medida por la descomposición en las distribuciones locales. Dos enfoques son comunes en la literatura: la estimación de máxima verosimilitud y la estimación bayesiana. Para las redes bayesianas discretas los parámetros se estiman usando las probabilidades condicionales en las distribuciones locales. Estos parámetros pueden ser estimados con las correspondientes frecuencias empíricas en el conjunto de datos. A pesar de que este enfoque parece ser una buena aproximación, hay que tomar en cuenta que muchas veces el conjunto de datos no abarca todas las posibles combinaciones de los valores de las variables predictoras, por lo que este tipo de estimación puede producir probabilidades iguales a cero. Como alternativa, se pueden estimar las probabilidades condicionales en un escenario bayesiano. En el caso de las redes bayesianas discretas se asume una muestra multinomial,  $X_i | \Pi_{X_i} \sim M(\theta_{X_i} | \Pi_{X_i})$ , donde  $\theta_{X_i} | \Pi_{X_i}$  representa las probabilidades condicionales  $\pi_{ijk} = \Pr(X_i = k | \Pi_{X_i} = j)$ . Si se considera una distribución a priori conjugada Dirichlet,  $\theta_{X_i} | \Pi_{X_i} \sim D(\alpha_{ijk})$ , se obtiene una distribución posterior  $D(\alpha_{ijk} + n_{ijk})$ .

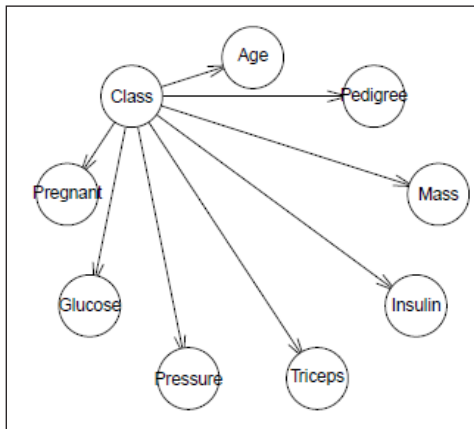
### Clasificadores por redes bayesianas

Las redes bayesianas son particularmente útiles en el proceso de clasificación supervisada ya que permiten simplificar la distribución conjunta de las variables predictoras usando distribuciones locales con una estructura más sencilla, a partir de las relaciones de dependencia e independencia condicional existentes entre las variables. Las redes bayesianas permiten representar la estructura probabilística de las variables predictoras que se puede utilizar para predecir la clase de pertenencia de cada observación usando el teorema de Bayes (Nagarajan *et al.*, 2013).

### Naive Bayes

Naive Bayes y sus variantes (Martínez *et al.*, 2016), se encuentran entre los algoritmos más conocidos para construir clasificadores

de documentos de texto, clasificación de galaxias y reconocimiento de emociones. A pesar de su simplicidad, es comparable con clasificadores sofisticados como las redes neuronales y los árboles de decisión, ya que posee alta precisión y velocidad cuando es aplicada a conjuntos grandes de datos. El clasificador Naive Bayes fue popularizado por [Duda \*et al.\*, 1973](#) gracias a su simplicidad, eficiencia y bajo error de clasificación. Este clasificador supone que todas las variables son condicionalmente independientes dado el valor de la variable de clase. La estructura de este clasificador puede representarse usando una red bayesiana, tal como se muestra en la [Figura 2](#).

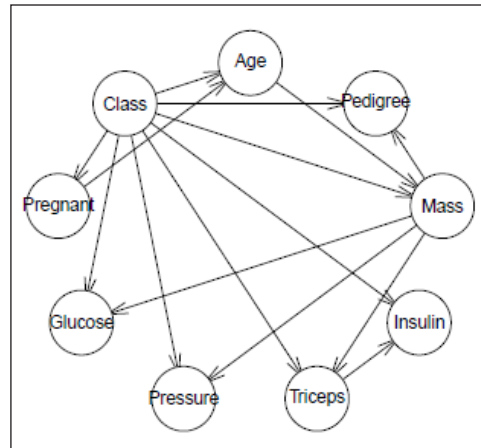


[Figura 2](#): Clasificador Naive Bayes para la data diabetes

### Tree Augmented Network

El clasificador Tree Augmented Network, TAN, fue propuesto por [Friedman \*et al.\* \(1997\)](#) como una extensión de Naive Bayes que mantiene su estructura básica. Para mejorar el comportamiento del clasificador, propuso aumentar la estructura de Naive Bayes con arcos entre las variables predictoras, cuando estos fueran necesarios, desechando así el supuesto de independencia. Estas estructuras son llamadas redes aumentadas Naive Bayes y estos arcos, arcos aumentados. En una estructura aumentada, un arco desde  $X_j$  hacia  $X_i$  implica que la influencia de  $X_j$  en la asignación de la variable de clase  $Y$  también depende del valor de  $X_i$ . El

propósito está en encontrar una red Naive Bayes de árbol aumentado, TAN, en la que la variable de clase no tenga padres y las variables predictoras tengan como padres a la variable de clase y, a lo más, alguna otra variable, tal como se muestra en la [Figura 3](#). El procedimiento para obtener estos arcos está basado en el algoritmo de [Chow y Liu \(1968\)](#) para estimar las relaciones de dependencia entre un conjunto de variables usando una estructura de árbol.



[Figura 3](#): Clasificador TAN para la data diabetes

### Proceso de discretización Chi-Merge

El algoritmo de discretización ChiMerge ([Kerber, 1992](#)) realiza un proceso de fusión de abajo hacia adelante, donde los intervalos adyacentes se juntan continuamente hasta que se cumple cierta condición de parada. Se ordenan los datos a discretizar y se considera que cada uno de ellos conforma un intervalo diferente. El proceso de fusión contiene dos pasos: (1) calcular el valor del estadístico de prueba  $\chi^2$  de independencia para cada par de intervalos adyacentes y (2) fusionar el par de intervalos adyacentes con el valor  $\chi^2$  más bajo. El proceso continúa hasta que todos los pares de intervalos tienen valores  $\chi^2$  que exceden cierto umbral predefinido; es decir, hasta que todos los intervalos adyacentes se consideren significativamente diferentes por la prueba de independencia. El valor del umbral se determina seleccionando un nivel de significación y calculando luego el percentil correspondiente a la distribución

chi-cuadrado con  $k - 1$  grados de libertad. La fórmula para calcular el estadístico de prueba  $\chi^2$  es:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (6)$$

### Algoritmo SES para selección de variables predictoras

El proceso de selección de variables es una de las tareas fundamentales en el área del *machine learning*. En general, se busca identificar un subconjunto de variables predictoras que son relevantes con respecto a una tarea específica; por ejemplo, en regresión y clasificación se busca seleccionar y retener el subconjunto de variables predictoras con el más alto poder predictivo.

Se han desarrollado algoritmos que generan múltiples conjuntos de variables equivalentes (Huang *et al.*, 2014). El algoritmo Statistically Equivalent Signature, SES, (Tsamardinos *et al.*, 2013) permite identificar múltiples subconjuntos de variables con rendimientos estadísticamente equivalentes. De esta forma, SES se presenta como una extensión de los métodos tradicionales de selección de variables predictoras.

Suponga que  $X$  es una variable predictora y  $W$  un subconjunto de variables predictoras. Se desea probar si  $Y$  es independiente de  $X$  dado  $W$  comparando el modelo de regresión logística multinomial que incluye como predictores a  $X$  y  $W$ , con el modelo de regresión logística multinomial alternativo que solo incluye a  $W$ . Cuando el modelo completo muestra una mejora significativa en comparación con el modelo alternativo, entonces se puede concluir que las variables  $X$  y  $Y$  se encuentran relacionadas dado .

La prueba de independencia condicional basada en la regresión logística multinomial se plantea como una prueba de comparación entre el logaritmo de la función de verosimilitud del modelo completo, MC, y el logaritmo de la función de verosimilitud del modelo alternativo, MA. El estadístico de prueba:

$$D=2(\log L_{MC}-L_{MA})\sim\chi_1^2 \quad (7)$$

## 3. Resultados y discusión

### Etapa de preprocesamiento

A cada conjunto de datos se le aplicó una etapa inicial de preprocesamiento que consta de dos procesos. El primero, elimina los valores perdidos antes de aplicar los algoritmos propuestos. El segundo, consiste en discretizar los valores correspondientes a las variables predictoras cuantitativas usando el algoritmo Chi-Merge, disponible en la librería *dprep*, dentro del software estadístico R.

### Etapa de estimación de la estructura de red

En la segunda etapa se realizó el proceso de estimación de la estructura presente en las variables predictoras, usando los algoritmos propuestos. El algoritmo basado en restricciones utiliza la prueba de independencia condicional de información mutua, mientras que el algoritmo basado en scores e híbrido usa el criterio de información bayesiano BIC. En esta etapa se aplicó también el proceso de selección de variables SES para construir la estructura, usando una menor cantidad de variables predictoras. De esta forma, es posible comparar el rendimiento predictivo de los clasificadores construidos con todas las variables predictoras y los clasificadores construidos con las variables predictoras seleccionadas por el algoritmo SES.

### Etapa de estimación del error de clasificación

En la tercera etapa se realizó el proceso de validación cruzada 10 para calcular la precisión del clasificador usando la tasa de elementos correctamente clasificados. Los pasos a seguir son los siguientes:

1. Se divide aleatoriamente el conjunto de datos en 10 partes.
2. Se realiza el proceso de estimación de parámetros usando 9 de estas partes como muestra de entrenamiento y evaluando el clasificador con la parte restante, llamada muestra de prueba, calculando la proporción de elementos correctamente clasificados.
3. Se repite el procedimiento anterior hasta que cada una de las 10 partes haya sido

utilizada como muestra de prueba.

Para efectos de evaluar la precisión del estimador, se repite el proceso de validación cruzada 20 veces. Luego, se calcula un promedio de las proporciones de observaciones correctamente clasificadas obtenidas en cada repetición. Se usó la librería `bnlearn` (Denis y Scutari, 2014) dentro del software estadístico R que ofrece una amplia variedad de algoritmos para la estimación de la estructura, estimación de parámetros y técnicas de inferencia en redes bayesianas.

### Clasificadores con todas las variables predictoras

La [Tabla 2](#) muestra la tasa de elementos correctamente clasificados usando los clasificadores bayesianos Naive Bayes, TAN y los obtenidos con los algoritmos propuestos, usando todas las variables predictoras de cada conjunto de datos.

**Tabla 3:** Variables seleccionadas con el algoritmo SES

N	Nombre	Variables	Algoritmo SES
1	Australian	14	8
2	Breast	9	4
3	Corral	6	6
4	Diabetes	8	5
5	German	20	4
6	Glass	9	3
7	Heart	13	6
8	Hepatitis	19	3
9	Iris	4	2
10	Vehicle	18	13
11	Vote	16	3

La [Tabla 4](#) muestra la tasa de elementos correctamente clasificados para Naive Bayes, TAN y los clasificadores obtenidos con los algoritmos propuestos, previo proceso de selección de variables.

**Tabla 2.** Tasa de elementos correctamente clasificados antes de aplicar el algoritmo SES

Nº	Nombre	Naive Bayes	TAN	Restricciones	Scores	Híbridos
1	Australian	0,894	0,816	0,895	0,896	0,897
2	Breast	0,976	0,962	0,973	0,966	0,973
3	Corral	0,865	0,995	0,829	0,865	0,865
4	Diabetes	0,843	0,749	0,836	0,833	0,841
5	German	0,924	0,674	0,919	0,908	0,908
6	Glass	0,783	0,730	0,772	0,760	0,772
7	Heart	0,869	0,674	0,860	0,838	0,863
8	Hepatitis	0,943	0,957	0,947	0,947	0,948
9	Iris	0,960	0,957	0,939	0,952	0,961
10	Vehicle	0,694	0,760	0,723	0,785	0,723
11	Vote	0,915	0,957	0,941	0,950	0,922

### Clasificadores con las variables predictoras seleccionadas por SES

En esta sección se construyen los clasificadores propuestos usando las variables predictoras seleccionadas por el algoritmo Statistically Equivalent Signature. En la [Tabla 3](#) se muestra cada conjunto de datos, el número total de variables predictoras y las que fueron seleccionadas luego de la aplicación del algoritmo SES.

### Comparación entre los clasificadores antes y después de aplicar el algoritmo SES

En esta sección, se comparan los clasificadores construidos con todas las variables predictoras y los clasificadores obtenidos luego de aplicar el algoritmo de selección de variables SES. Se consideran los tres algoritmos propuestos para la construcción de la estructura presente entre las variables predictoras. Los resultados obtenidos en las secciones anteriores se resumen en la [Tabla 5](#).



**Tabla 4:** Tasa de elementos correctamente clasificados luego de aplicar el algoritmo SES

Nº	Nombre	Naive Bayes	TAN	Restricciones	Scores	Híbridos
1	Australian	0,898	0,802	0,860	0,896	0,896
2	Breast	0,976	0,948	0,971	0,964	0,976
3	Corral	0,865	0,995	0,829	0,865	0,865
4	Diabetes	0,848	0,751	0,832	0,847	0,847
5	German	0,925	0,833	0,899	0,923	0,923
6	Glass	0,721	0,685	0,677	0,716	0,716
7	Heart	0,866	0,774	0,846	0,853	0,868
8	Hepatitis	0,916	0,938	0,921	0,916	0,916
9	Iris	0,972	0,971	0,971	0,971	0,971
10	Vehicle	0,715	0,749	0,743	0,748	0,734
11	Vote	0,962	0,961	0,960	0,960	0,960

**Tabla 5:** Tasa de elementos correctamente clasificados antes y después de aplicar el algoritmo SES

Nº	Nombre	Restricciones		Scores		Híbridos	
		Sin SES	Con SES	Sin SES	Con SES	Sin SES	Con SES
1	Australian	0,895	0,860	0,896	0,896	0,897	0,896
2	Breast	0,973	0,971	0,966	0,964	0,973	0,976
3	Corral	0,829	0,829	0,865	0,865	0,865	0,865
4	Diabetes	0,836	0,832	0,833	0,847	0,841	0,847
5	German	0,919	0,899	0,908	0,923	0,908	0,923
6	Glass	0,772	0,677	0,760	0,716	0,772	0,716
7	Heart	0,860	0,846	0,838	0,853	0,863	0,868
8	Hepatitis	0,947	0,921	0,947	0,916	0,948	0,916
9	Iris	0,939	0,971	0,952	0,971	0,961	0,971
10	Vehicle	0,723	0,743	0,785	0,748	0,723	0,734
11	Vote	0,941	0,960	0,950	0,960	0,922	0,960

#### 4. Conclusiones

Los clasificadores por redes bayesianas construidos con los algoritmos basados en restricciones, scores e híbridos presentan un buen comportamiento predictivo, según los valores de la [Tabla 3](#). Al ser comparados con los clasificadores tradicionales, se obtuvieron tasas de elementos correctamente clasificados muy cercanos a los obtenidos con Naive Bayes y ligeramente superiores en el caso de TAN. Sin embargo, los clasificadores obtenidos tienen la ventaja de estar construidos sobre estructuras que representan las relaciones existentes entre las variables predictoras y en este caso los

clasificadores propuestos tienen, en términos generales, un mejor comportamiento predictivo que TAN. El algoritmo basado en restricciones tiene la desventaja de obtener estructuras donde algunos de los arcos podrían no estar dirigidos. Lo anterior supone la tarea previa de decidir la dirección de los arcos antes de construir los clasificadores. Sin embargo, la elección es arbitraria y no determina cambios importantes ni en las medidas de score calculadas sobre la estructura obtenida, ni sobre la tasa de elementos correctamente clasificados. El esfuerzo computacional requerido por los algoritmos propuestos para obtener la estructura entre variables predictoras, y

posterior construcción del clasificador, es ligeramente mayor al que se requiere con los clasificadores tradicionales Naive Bayes y TAN. Las librerías `bnlearn`, `dprep` y `MXM` no presentaron inconveniente alguno al trabajar con los conjuntos de datos usados en el presente trabajo de investigación.

El algoritmo de selección de variables *Statistically Equivalent Signatures*, utilizado antes de la aplicación de los algoritmos de estimación de la estructura, permitió obtener clasificadores sobre una menor cantidad de variables predictoras. Sin embargo, no se obtiene una diferencia importante en la tasa de elementos correctamente clasificados en comparación con los clasificadores construidos con todas las variables predictoras, según los valores de la [Tabla 5](#). Por otro lado, elegir las variables predictoras a utilizar en una etapa inicial permite un menor gasto computacional al momento de aplicar los algoritmos propuestos.

## 5. Recomendaciones

Los clasificadores usados en el presente trabajo de investigación requieren una etapa inicial de discretización de las variables cuantitativas continuas para llevar a cabo el proceso de estimación de la estructura, a partir de los algoritmos propuestos. Sin embargo, esto puede llevarnos a una inevitable pérdida de información por lo que es importante considerar metodologías que permitan trabajar con ambos tipos de variables.

Es posible usar las redes bayesianas gaussianas para los conjuntos de datos que tienen solamente variables cuantitativas continuas. En este tipo de redes, se asume que la distribución global es normal multivariada y que cada distribución local se puede expresar como un modelo lineal gaussiano clásico de regresión donde el nodo es la variable respuesta y sus padres son las variables explicativas.

Las redes bayesianas mixtas permiten trabajar con variables discretas y continuas, pudiendo utilizar casi cualquier modelo de probabilidad para las distribuciones locales dentro de lo razonable. Desafortunadamente, esta mayor flexibilidad hace que la red bayesiana sea más compleja. No existe

en R alguna librería que permita manejar este tipo de redes bayesianas por lo que el proceso de estimación de la estructura y de los parámetros requieren de un esfuerzo de programación por parte del usuario.

## 6. Literatura citada

- Chow, C.; Liu, C. 1968. Aproximación de distribuciones de probabilidad discretas con árboles de dependencia. *Transacciones IEEE sobre teoría de la información*. 14 (3): 462-467.
- Denis, J.; Scutari, M. 2014. Réseaux bayésiens avec r: Élaboration, manipulation et utilisation en modélisation appliquée.
- Duda, R.O.; Hart, P.E.; Stork, D.G. *et al.* 1973. *Pattern classification*, volume 2. Wiley New York.
- Dheeru, D.; y Karra, E. 2017. UCI machine learning repository.
- Friedman, N.; Geiger, D.; Goldszmidt, M. 1997. Bayesian network classifiers. *Machine learning*, 29 (2-3): 131-163.
- Friedman, N.; Nachman, I.; Peér, D. 1999. Learning bayesian network structure from massive datasets: the «sparse candidate» algorithm. In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 206-215 p.
- Huang, G.T.; Tsamardinos, I.; Raghu, V.; Kaminski, N.; Benos, P.V. 2014. T-recs: stable selection of dynamically formed groups of features with application to prediction of clinical outcomes. In: *Pacific Symposium on Biocomputing Co-Chairs*, World Scientific. 431-442 p.
- Kerber, R. 1992. Chimerge: Discretization of numeric attributes. In *Proceedings of the tenth national conference on Artificial intelligence*, Aaai Press. 123-128 p.
- Margaritis, D. 2003. Learning bayesian network model structure from data. Technical report, Carnegie-Mellon Univ. Pittsburgh Pa School of Computer Science.
- Martinez, A.M.; Webb, G.I.; Chen, S.; Zaidi, N.A. 2016. Scalable learning of bayesian network classifiers. *Journal of Machine Learning Research*, 17 (44): 1-35.
- Nagarajan, R.; Scutari, M.; Lèbre, S. 2013. Bayesian networks in r. Springer, 122:

125–127.

- Pearl, J. 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kauffman Pub.
- Taheri, S.; Yearwood, J.; Mammadov, M.; Seifollahi, S. 2014. Attribute weighted naive bayes classifier using a local optimization. *Neural Computing and Applications* 24 (5): 995–1002.
- Tsamardinos, I.; Brown, L.E.; Aliferis, C F. 2006. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65 (1): 31–78.
- Tsamardinos, I.; Lagani, V.; Pappas, D. 2013. Discovering multiple, equivalent biomarker signatures. In: 7th Conference of the Hellenic Society for Computational Biology and Bioinformatics (HSCBB12).