

REDES BAYESIANAS APLICADAS A LA DETECCION DE DIABETES MELLITUS TIPO 2.

Ricardo Cruz Sánchez

21 de Agosto de 2020

Índice

1 Representación

2 Inferencia

3 Aprendizaje

4 Caso de estudio

Definición

Red bayesiana

Una **red bayesiana** (**RB**) es un grafo acíclico dirigido $\mathcal{G} = (V, A)$ que representa de manera compacta a una probabilidad conjunta P sobre una colección de variables $\mathbf{X} = (X_1, \dots, X_n)$ y el conjunto de independencias condicionales existentes entre las variables. Cada uno de los nodos en V está asociado a cada una de las variables aleatorias en \mathbf{X} y las aristas entre dos nodos indican la dependencia directa entre dos variables..

Ejemplo

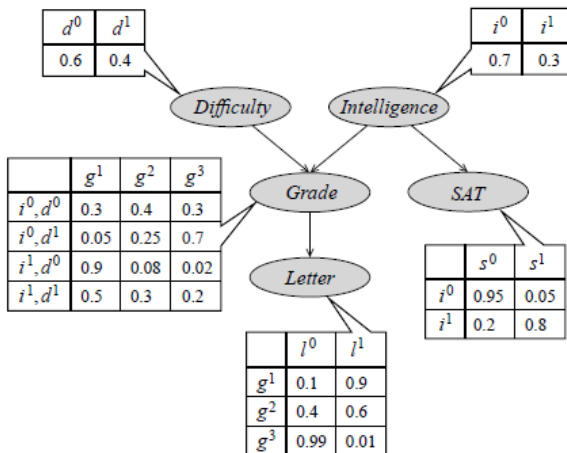


Figure 1: Ejemplo de una RB.

Índice

1 Representación

2 Inferencia

3 Aprendizaje

4 Caso de estudio

- Se basa en especificar la estructura y parámetros del modelo. Se entiende como estructura al DAG y los parámetros son las funciones de probabilidad condicional seleccionadas para cada nodo.
- El principal supuesto que se realiza sobre las redes bayesianas se conoce como *el supuesto de Markov*, el cual establece que sobre un grafo \mathcal{G} , cualquier nodo $X \in V$ es independiente de cualquier nodo que no sea descendiente de X , (conjunto $NoDescendientes_X$) una vez que se conoce el valor de los nodos que son padres de X , es decir:

$$(X \perp NoDescendientes_X \mid Pa_X) \quad \forall X \in V \quad (1)$$

Ejemplo

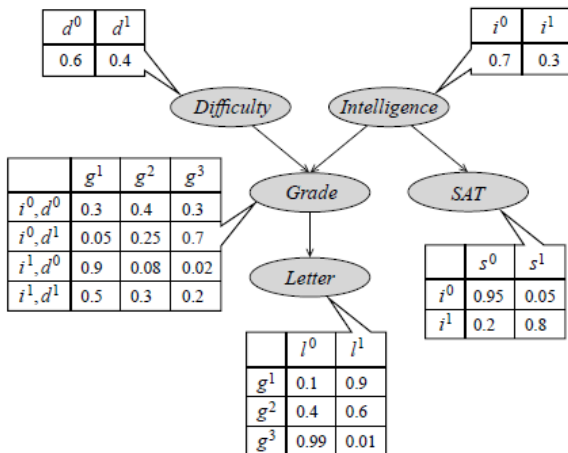


Figure 2: Ejemplo de una RB.

$$P(\mathbf{X}) = P(X_1, \dots, X_n)$$

v.s.

$$P(\mathbf{X}) = P(\mathbf{Y}) = \prod_{i=1}^n P(Y_i \mid Y_1, \dots, Y_{i-1}) = \prod_{i=1}^n P(Y_i \mid Pa_{Y_i})$$

donde Y es una ordenación ancestral.

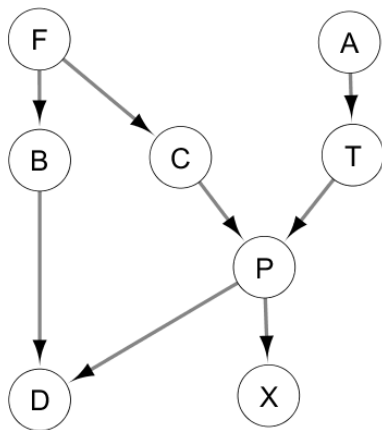


Figure 3: Ejemplo de una RB y el impacto de la regla de la cadena en RB.

Estructura

Indica las relaciones que se satisfacen entre las variables

D-separación

En un grafo dirigido $\mathcal{G} = (V, A)$, se dice que dos conjuntos disjuntos $X, Y \subset V$ están *separados dirigidamente* o *d-separados* por un conjunto disjunto Z si y solo si en todo camino no dirigido entre un nodo de X a un nodo de Y existe al menos un nodo B tal que:

- 1 Si B es el hijo en una v-estructura en el grafo \mathcal{G} ($A \rightarrow B \leftarrow C$), entonces B ni ninguno de sus descendientes ($Descendientes_B$) existe en el conjunto separador Z .
- 2 Si B no es el hijo en una v-estructura en el grafo \mathcal{G} , entonces B existe en el conjunto separador Z .

Parámetros

determinan de manera numérica la intensidad de cada una de las relaciones descritas por la estructura

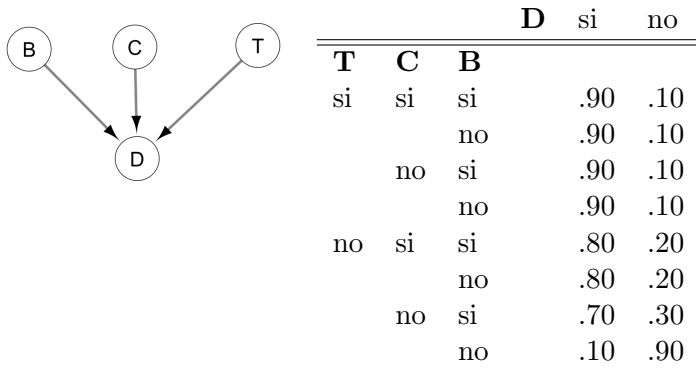


Figure 4: Disnea como efecto de 3 enfermedades y su tabla de probabilidad.

Índice

1 Representación

2 Inferencia

3 Aprendizaje

4 Caso de estudio

Una vez definida la RB sobre un espacio \mathbf{X} , se espera poder utilizar el conocimiento incluido en la red para obtener la probabilidad de una asignación \mathbf{x} . Sin embargo, en muchas ocasiones, no se cuenta con información de todas las variables y es justo para lo que los modelos están diseñados. Con base en un conjunto de datos, el modelo es capaz de inferir el comportamiento de otro conjunto de variables no observadas hasta ese momento.

Todos los algoritmos buscan encontrar $P(\mathbf{X}|\mathbf{E})$, donde \mathbf{X} y \mathbf{E} son conjuntos disjuntos. \mathbf{E} representará el conjunto de variables observadas (evidencia) y \mathbf{X} las variables de interés.

Algoritmos exactos

- Suma-producto
- Pearl o propagación de probabilidad
- Condicionamiento
- Árbol unión

Algoritmos aproximados

- Aceptación-rechazo
- Muestreo uniforme
- Verosimilitud pesante
- MCMC

Índice

1 Representación

2 Inferencia

3 Aprendizaje

4 Caso de estudio

En ocasiones la recopilación del conocimiento de un experto puede ser una tarea imposible debido a los recursos con los que se cuenta, además de que puede ser una tarea subjetiva en la cual no es posible medir la calidad del conocimiento transferido.

Se puede recurrir a estimar o aprender la representación de una RB a través de datos ya observados de las variables de interés. Es decir, las mismas observaciones muestran cual es el grafo y funciones de distribución condicionales adecuadas para las variables X_1, \dots, X_n consideradas en el problema.

Paramétrico

Se asumen una estructura ya proporcionada y con base en los datos y el grafo se infieren las tablas de probabilidad.

- máxima verosimilitud de los datos
- enfoque bayesiano

Paramétrico

Cuando el conjunto de datos D es incompleto se puede emplear el algoritmo *expectation-maximization*, EM.

Consiste en simular los parámetros en el paso *expectation* y evaluar su verosimilitud en el paso *maximization*.

Este algoritmo es iterativo y depende del punto de inicio, por lo que se recomienda ejecutarlo en varias ocasiones con diferentes puntos iniciales.

Estructural

El aprendizaje de la estructura de una RB solo depende de un conjunto de datos. Similar al proceso de aprendizaje paramétrico se buscará una estructura que satisfaga algún criterio con el cual determinar cuál es la mejor selección de la estructura de acuerdo a las observaciones otorgadas.

- *Score*
- *BScore*

Hill climbing

El algoritmo considera un grafo inicial y pasa a un siguiente grafo a través de los *scores* y 3 operaciones en grafos.

Se pueden agregar restricciones a este algoritmo para considerar arcos obligatorios o prohibidos.

Existen medidas para determinar lo significativas que pueden ser algunas características de la red, por ejemplo, la fuerza del arco, la cual realiza remuestreo Bootstrap y evalúa que tan frecuente es la aparición de los arcos

Índice

1 Representación

2 Inferencia

3 Aprendizaje

4 Caso de estudio

Datos

La base de datos seleccionada, proviene del Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales, en Estados Unidos. Cada registro contiene las características relevantes de un paciente para poder determinar la presencia o ausencia de diabetes tipo 2.

Los datos fueron recolectados de la población femenina mayor de 21 años de un grupo tribu de nativos norteamericanos conocidos como Pima, con el objetivo de determinar las causas del alto índice de prevalencia de este padecimiento en este grupo.

Datos

La base de datos contiene 768 registros en los que se detallan 9 variables. La variable Diabetes se considera como la variable objetivo y las 8 restantes son las características particulares de cada paciente.

- **Embarazo**
- **Glucosa** (mg/dL)
- **Presión** ($mmHg$)
- **Tricep** (mm)
- **Insulina** (muU/mL)
- **IMC**
- **Pedigree**
- **Edad**
- **Diabetes**

Datos

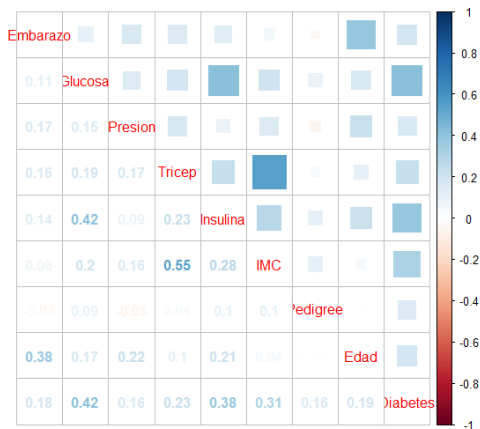


Figure 5: Matriz de correlación

Representación y aprendizaje

La representación será modelada con técnicas de aprendizaje, es decir, los datos mostrarán la estructura y parámetros con los cuales se pueda modelar la red bayesiana. Comenzando con un modelo sin ninguna restricción y a partir de este se agregan las opiniones de expertos acerca de las relaciones que guardan las variables.

La proporción de los conjuntos de entrenamiento y prueba son 80% (614) y 20% (154).

Se opta por utilizar el algoritmo E-M para el aprendizaje estructural con 100 inicios aleatorios y el algoritmo Hill-Climbing con el enfoque de verosimilitud y el BIC como *score*.

Representación y aprendizaje

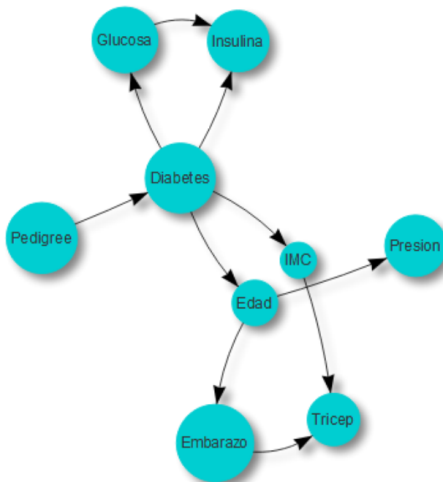


Figure 6: Grafo con algoritmo E-M sin restricciones.

Representación y aprendizaje

Desde	Hacia	arco	dirección
Diabetes	Embarazo	0.275	0.5727273
Diabetes	Glucosa	0.965	0.4818653
Diabetes	Presión	0.220	0.6590909
Diabetes	Tricep	0.235	0.4893617
Diabetes	Insulina	0.485	0.5792079
Diabetes	IMC	0.900	0.4361111
Diabetes	Pedigree	0.715	0.6290323
Diabetes	Edad	0.265	0.5094340

Table 1: Fuerza del arco y dirección, m=200 sin restricciones

Representación y aprendizaje

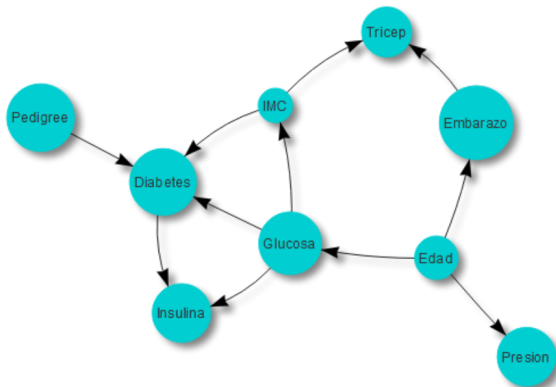


Figure 7: Grafo con algoritmo E-M y restricciones.

Representación y aprendizaje

Desde	Hacia	arco	dirección
Embarazo	Tricep	1.000	1.0000
Glucosa	Insulina	1.000	0.9725
Glucosa	Diabetes	1.000	1.0000
Insulina	IMC	0.875	0.9800
IMC	Tricep	1.000	1.0000
IMC	Diabetes	1.000	1.0000
Pedigree	Diabetes	1.000	1.0000
Edad	Embarazo	1.000	1.0000
Edad	Glucosa	0.815	1.0000
Edad	Presion	0.600	1.0000

Table 2: Fuerza del arco y dirección, $m = 200$ con restricciones

Inferencia

Se realiza inferencia de manera supervisada sobre la variable Diabetes.

Paciente	$P(D = no E = e)$	$P(D = si E = e)$
1	0.3030303	0.69696970
2	0.6885246	0.31147541
3	0.5036869	0.49631310
4	0.8846154	0.11538462
5	0.8613281	0.13867187
\vdots	\vdots	\vdots
610	0.6472196	0.35278042
611	1.0000000	0.00000000
612	0.9049786	0.09502141
613	0.8613281	0.13867187
614	0.9310345	0.06896552

Table 3: Probabilidades para diabetes en el conjunto de

Entrenamiento

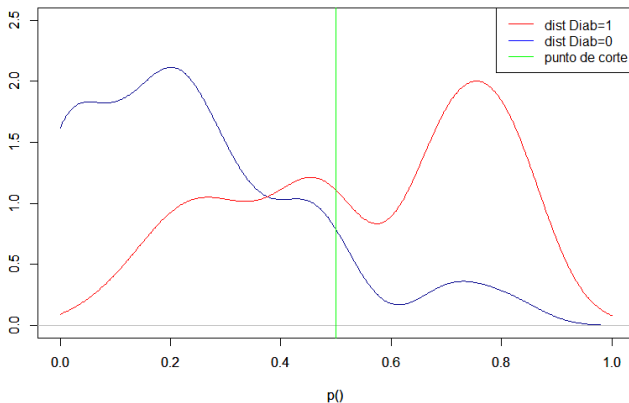


Figure 8: Densidades de los valores originales. $Acc = 77.7\%$
 $Sen = 55.9\%$

Entrenamiento

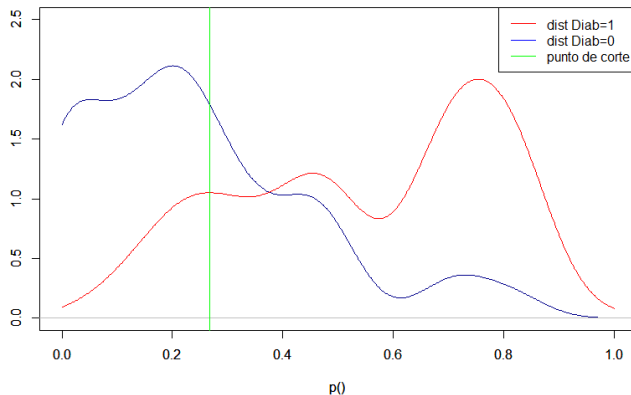


Figure 9: Modificación del punto de corte. $Acc = 73.6\%$ $Sen = 75.1\%$

Prueba

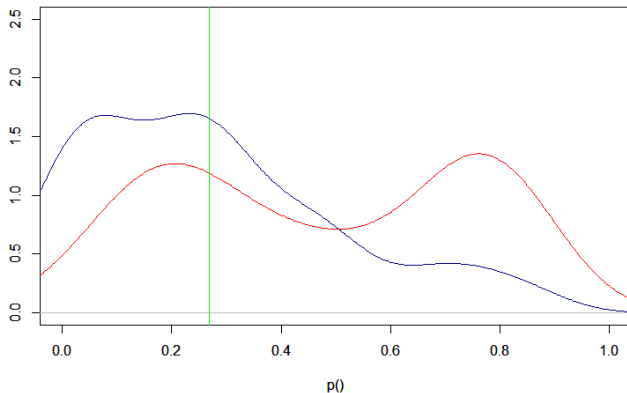


Figure 10: Densidades para el conjunto de prueba. $Acc = 62.3\%$
 $Sen = 64.4\%$

Cross-validation

Fold	Presición entr	Sensib entr	Presición prue	Sesns prue
1	0.7100977	0.8130841	0.6558442	0.6296296
2	0.6894309	0.7892157	0.6993464	0.8593750
3	0.7084691	0.7440758	0.7012987	0.6842105
4	0.6959350	0.7304348	0.7320261	0.8684211
5	0.6758958	0.8403756	0.6428571	0.8363636
Prom	0.6959657	0.7834372	0.6862745	0.7756000

Table 4: Resultados de la validación cruzada con 5 *folds*.

Conclusiones

Las redes bayesianas son una herramienta útil cuando se debe especificar la distribución conjunta de una colección de variables aleatorias, ya que permite incorporar el conocimiento del tema a tratar y visualmente proporciona las relaciones de causalidad existentes.

Son de gran ayuda cuando con frecuencia se realizan consultas de estado de alguna o algunas variables, dado que se conoce el valor de otras. Esto gracias al proceso de inferencia que permite considerar a cualquier subconjunto de variables como variables respuesta, a diferencia de otros modelos en los que la respuesta es fija y cambiarla implicaría entrenar un nuevo modelo.

Conclusiones

En general, las ventajas de optar por una red bayesiana son:

- La incorporación de la naturaleza del fenómeno a través de las independencias.
- La reducción en la especificación de parámetros.
- La obtención de manera eficaz y sencilla de probabilidades condicionales.
- Se puede realizar inferencia sobre cualquier subconjunto de variables.
- Posee un manejo adecuado de los valores faltantes.
- Si se requiere agregar alguna variable al modelo, no necesariamente se tiene que realizar todo el proceso para la construcción de la red de nuevo.
- Visualmente se puede comprender con facilidad.

Conclusiones

Algunas de las desventajas de este modelo son.

- Encontrar a algún experto que pueda transferir el conocimiento del tema.
- Costo computacional alto a medida que el número de variables crece.
- Al usarlo como clasificador puede no ser tan bueno como algoritmos más sofisticados.
- Dependen de la calidad de los datos, pues se construyen a partir de estos.