

Clase 2:

Redes Bayesianas

Agenda

- Explicación: Redes Bayesianas
- Hands-On
- Break
- Cierre

Redes Bayesianas

¿A qué llamamos Red Bayesiana?
¿Cuándo la utilizamos?

Introducción

- Las redes bayesianas modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas. Dado este modelo, se puede hacer inferencia bayesiana; es decir, estimar la probabilidad posterior de las variables no conocidas, en base a las variables conocidas.
- Estos modelos pueden tener diversas aplicaciones, para clasificación, predicción, diagnóstico, etc. Además, pueden dar información interesante en cuanto a cómo se relacionan las variables del dominio, las cuales pueden ser interpretadas en ocasiones como relaciones de causa-efecto.

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística; son utilizadas para proveer:

- una forma compacta de representar el conocimiento y métodos flexibles de razonamiento.

El obtener una red bayesiana a partir de datos es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico.

- Las redes bayesianas o probabilísticas se fundamentan en la teoría de la probabilidad y combinan la potencia del teorema de Bayes con la expresividad semántica de los grafos dirigidos; las mismas permiten representar un modelo causal por medio de una representación gráfica de las independencias / dependencias entre las variables que forman parte del dominio de aplicación.

Se puede interpretar a una red bayesiana de dos formas:

- 1). **Distribución de probabilidad**: Representa la distribución de la probabilidad conjunta de las variables representadas en la red.
- 2). **Base de reglas**: Cada arco representa un conjunto de reglas que asocian a las variables involucradas. Dichas reglas están cuantificadas por las probabilidades respectivas.

A continuación se describirán los fundamentos teóricos de las redes bayesianas y distintos algoritmos de propagación.

Definición formal

Una red bayesiana es un grafo acíclico dirigido en el que los nodos representan variables aleatorias que pueden ser continuas o discretas; en las siguientes definiciones se utilizarán letras mayúsculas para denotar los nodos (X) y las correspondientes letras minúsculas para designar sus posibles estados (i_x).

Los **estados** que puede tener una variable deben cumplir con dos propiedades:

1. Ser **mutuamente excluyentes**, es decir, un nodo sólo puede encontrarse en uno de sus estados en un momento dado.
2. Ser un **conjunto exhaustivo**, es decir, un nodo no puede tener ningún valor fuera de ese conjunto.

Terminología

- *Nodo*

Un nodo X es una variable aleatoria que puede tener varios estados x . La probabilidad de que el nodo X esté en el estado x se denotará como $P(x) = P(X = x)$.

- *Arco*

Es la unión entre dos nodos y representa la dependencia entre dos variables del modelo.

Un arco queda definido por un par ordenado de nodos (X,Y) .

Representación del conocimiento

1. Un conjunto de nodos $\{X_i\}$ que representan cada una de las variables del modelo. Cada una de ellas tiene un conjunto exhaustivo de estados $\{x_i\}$ mutuamente excluyentes.
2. Un conjunto de enlaces o arcos (X_i, X_j) entre aquellos nodos que tienen una relación causal. De esta manera todas las relaciones están explícitamente representadas en el grafo.
3. Una tabla de probabilidad condicional asociada a cada nodo X_i indicando la probabilidad de sus estados para cada combinación de los estados de sus padres. Si un nodo no tiene padres se indican sus probabilidades a priori.

Representación del conocimiento

La estructura de una red bayesiana se puede determinar de la siguiente manera:

1. Se asigna un vértice o nodo a cada variable (X_i) y se indica de qué otros vértices es una causa directa; a ese conjunto de vértices “causa del nodo X_i ” se lo denota como el conjunto X_i^p y se lo llamará “padres de X_i ”.
2. Se une cada padre con sus hijos con flechas que parten de los padres y llegan a los hijos.
3. A cada variable X_i se le asigna una matriz $(P_{ij})_{X_i^p \times p}$ que estima la probabilidad condicional de un evento i $X = x$ dada una combinación de valores de los X_i^p .

Representación del conocimiento

La estructura de una red bayesiana se puede determinar de la siguiente manera:

1. Se asigna un vértice o nodo a cada variable (X_i) y se indica de qué otros vértices es una causa directa; a ese conjunto de vértices “causa del nodo X_i ” se lo denota como el conjunto X_i^p y se lo llamará “padres de X_i ”.
2. Se une cada padre con sus hijos con flechas que parten de los padres y llegan a los hijos.
3. A cada variable X_i se le asigna una matriz $(P_{ij})_{X_i^p \times p}$ que estima la probabilidad condicional de un evento i $X = x$ dada una combinación de valores de los X_i^p .

Representación del conocimiento

Una vez que se ha diseñado la estructura de la red y se han especificado todas las tablas de probabilidad condicional se está en condiciones de conocer la probabilidad de una determinada variable dependiendo del estado de cualquier combinación del resto de variables de la red; para ello se debe calcular la probabilidad a posteriori de cada variable condicionada a la evidencia;

- Estas probabilidades a posteriori se podrán obtener de forma inmediata a partir de la probabilidad conjunta de todas las variables.
- Este proceso se ve simplificado al aplicar la propiedad de independencia condicional que permite obtener la probabilidad conjunta a partir de las probabilidades condicionales de cada nodo en función de sus padres.

Independencia condicional

La topología o estructura de una red bayesiana no sólo representa explícitamente dependencias probabilísticas entre variables, sino que también describe implícitamente las independencias condicionales existentes entre ellas.

La siguiente definición muestra las condiciones que deben darse para que dos variables sean condicionalmente independiente de otra Y dada una tercer variable Z , si el conocer Z hace que X e Y sean independientes. Es decir, si conozco Z , Y no tiene influencia en X . Esto es: $P(X | Y, Z) = P(X | Z)$.

Esta definición se traduce a que cada variable es independiente de todos aquellos nodos que no son sus “descendientes” una vez que se conocen sus propios nodos padres.

Algoritmos de propagación

Los algoritmos de propagación se dividen inicialmente en “exactos” o “aproximados” según cómo calculen los valores de las probabilidades.

- Los *métodos exactos* calculan los valores por medio del teorema de Bayes mientras que los *métodos aproximados* utilizan técnicas iterativas de muestreo en las que los valores se aproximarán más o menos a los exactos dependiendo del punto en que se detenga el proceso.
- Los algoritmos de propagación dependen del tipo de estructura de la red bayesiana, existiendo las siguientes dos topologías de red:

Algoritmos de propagación

- Los algoritmos de propagación dependen del tipo de estructura de la red bayesiana, existiendo las siguientes dos topologías de red:
 - Árboles
 - Poliárboles

El aprendizaje en las redes bayesianas

- El aprendizaje es una de las características que definen a los sistemas basados en inteligencia artificial ; es difícil definir el término “aprendizaje”, pero la mayoría de las autoridades en el campo coinciden en que es una de las características de los sistemas adaptativos que son capaces de mejorar su comportamiento en función de su experiencia pasada, por ejemplo al resolver problemas similares.
- Los sistemas de aprendizaje son capaces de generar nuevo conocimiento y de ajustar el conocimiento existente. El aprendizaje en la redes bayesianas consiste en definir la red probabilística a partir de datos almacenados en bases de datos en lugar de obtener el conocimiento del experto.

El aprendizaje en las redes bayesianas

- A continuación se resume cada una de estas dos fases:

Aprendizaje estructural: obtiene la estructura de la red bayesiana a partir de bases de datos, es decir, involucradas las relaciones de dependencia e independencia entre las variables

- Las técnicas de aprendizaje estructural dependen del tipo de estructura o topología de la red (árboles, poliárboles o redes multiconectadas). Otra alternativa es combinar conocimiento subjetivo del experto con aprendizaje, para lo cual se parte de la estructura dada por el experto y se la valida y mejora utilizando datos estadísticos.

Aprendizaje paramétrico: dada una estructura y las bases de datos, obtiene las probabilidades a priori y condicionales requeridas.

Clasificador Bayes ingenuo

- Uno de los clasificadores más utilizados en Machine Learning por su simplicidad y rapidez, es el Clasificador Bayes ingenuo. El cual es una técnica de clasificación supervisada basada en el teorema de Bayes que asume que existe una independencia entre los atributos.
- En términos simples, un Clasificador Bayes ingenuo asume que la presencia de una característica particular en una clase no está relacionada con la presencia de cualquier otra característica.

Clasificador Bayes ingenuo

- Por ejemplo, una fruta puede considerarse como una manzana si es roja, redonda y de aproximadamente 9 cm de diámetro. Incluso si estas características dependen unas de otras o de la existencia de otras características, todas estas propiedades contribuyen independientemente a la probabilidad de que esta fruta sea una manzana.
- Se lo llama ingenuo ya que asumir independencia absoluta entre todos los atributos, no es algo que se suele dar en la realidad. El modelo Bayes ingenuo es fácil de construir y particularmente útil para conjuntos de datos muy grandes.
- A pesar de su simplicidad y de su irrealista postulado de independencia, este clasificador se ha mostrado muy efectivo y se suele utilizar como el estándar para evaluar el rendimiento de otros modelos de Machine Learning.

Clasificador Bayes ingenuo

El Clasificador Bayes ingenuo se utiliza en múltiples escenarios de la vida real, tales como:

- Clasificación de texto: Es uno de los algoritmos conocidos más exitosos cuando se trata de la clasificación de documentos de texto, es decir, si un documento de texto pertenece a una o más categorías (clases).
- Detección de spam: Es un ejemplo de clasificación de texto. Se ha convertido en un mecanismo popular para distinguir el correo electrónico spam del correo electrónico legítimo.

Clasificador Bayes ingenuo

- Análisis de sentimientos: Puede ser utilizado para analizar el tono de tweets, comentarios y revisiones, ya sean negativos, positivos o neutrales.
- Sistema de Recomendaciones: El algoritmo Bayes ingenuo en combinación con el filtrado colaborativo se utiliza para construir sistemas de recomendación híbridos que ayudan a predecir si un usuario desea un recurso determinado o no.

Clasificador Bayes ingenuo

- Ejemplo:

```
# Ejemplo Naive Bayes usando iris dataset
iris = datasets.load_iris()
X = iris.data
y = iris.target

# Dividir los datos en entrenamiento y evaluación
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.7,
                                                    random_state=0)

# inicializar el clasificador Naive Bayes
bayes_ingenuo = GaussianNB()

# predicción
y_pred = bayes_ingenuo.fit(X_train, y_train).predict(X_test)

# Matriz de confusión
cnf_matrix = confusion_matrix(y_test, y_pred)

print("Cantidad de errores de clasificación sobre un total de {0} casos: {1}"
      .format(y_test.shape[0], (y_test != y_pred).sum()))
print("Efectividad del algoritmo: {0: .2f}"
      .format(1 - (y_test != y_pred).sum()/y_test.shape[0]))

# Graficando la matriz de confusión
sns.heatmap(cnf_matrix.T, square=True, annot=True, fmt='d', cbar=False)
plt.xlabel('Clase verdadera')
plt.ylabel('Clase predecida')
plt.title('Matriz de Confusión')
plt.show()
```

Cantidad de errores de clasificación sobre un total de 105 casos: 7
Efectividad del algoritmo: 0.93

Redes Bayesianas

- Redes Bayesianas (Sucar)
<http://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>
- Microsoft Bayesian Networks - Basics of Knowledge Engineering (Locke)
<http://www.freelock.com/sites/default/files/KE.pdf>

