

Inteligencia Artificial

Técnicas Probabilísticas: Algoritmo de Naive Bayes y Redes de Bayes

Profesor: Andrés Villa

Conceptos Básicos

- Lenguaje de las probabilidades para modelar incertezas
- Probabilidad condicional
- Regla de la cadena
- Teorema de la probabilidad total
- Probabilidad conjunta
- Independencia
- Independencia condicional

Redoble: Sr. Bayes.



Regla de Bayes:

El Sumo de los poderes, alquimia pura

$$P(h/e) == P(e/h)$$

Verdadero o False

Regla de Bayes:

El Sumo de los poderes, alquimia pura

$$P(h/e) == P(e/h)$$

Verdadero o **False**

Regla de Bayes:

El Sumo de los poderes, alquimia pura

$$P(h/e) = \frac{P(e \cap h)}{P(e)}$$

$$P(e \cap h) = P(e/h)P(h)$$

Entonces:

$$P(h/e) = P(e/h) \frac{P(h)}{P(e)}$$

La regla de Bayes hay que interiorizarla. **Es muy importante para lo que viene.**

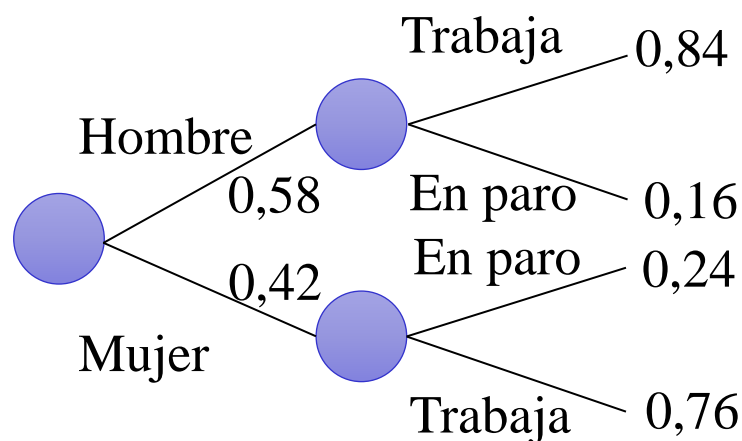
Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370-418



Regla de Bayes:

El Sumo de los poderes, alquimia pura

Supongamos que se elige un adulto al azar para rellenar un formulario y se observa que no tiene trabajo, dado el árbol de probabilidad, **¿Cuál es la probabilidad de que la persona elegida sea mujer?**



Regla de Bayes:

El Sumo de los poderes, alquimia pura

$$P(Mj/Pa) = P(Pa/Mj) \frac{P(Mj)}{P(Pa)}$$

$$P(Pa/Mj) = 0,24$$

$$P(Mj) = 0,42$$

$$P(Pa) = P(Pa/Mj) P(Mj) + P(Pa/Hm) P(Hm)$$

$$P(Pa) = 0,24 * 0,42 + 0,16 * 0,58 = 0,1936$$

Clasificador Inexperto de Bayes (Naive Bayes Classifier)

- Uno de los algoritmos de aprendizaje de máquina más utilizados en la práctica. Diversas aplicaciones exitosas en diagnósticos médicos, clasificación de texto, etc.
- Adecuado para casos en que:
 - Clasificación supervisada
 - Se cuenta con un gran set de entrenamiento

Clasificador Inexperto de Bayes (Naive Bayes Classifier)

Siendo X , es una instancia compuesta por n características distintas:

$$X = (x_1, x_2, x_3, \dots, x_n)$$

¿Cómo podemos calcular la probabilidad que X sea clasificado como y ?

$$P(y/X) = P(y/x_1, x_2, x_3, \dots, x_n)$$

- Supuesto de Naive Bayes :

$$P(a_1 \dots a_n / v_j) = P(a_1 / v_j) \dots P(a_n / v_j)$$

- Entonces el clasificador queda dado por:

$$v_{NB} = \operatorname{argmax}_{v_j \in V} P(v_j) [P(a_1 / v_j) \dots P(a_n / v_j)]$$

¿Algún problema cuando la dimensionalidad de a es grande ?

Clasificador Inexperto de Bayes (Naive Bayes Classifier)

Siendo X , es una instancia compuesta por n características distintas:

$$X = (x_1, x_2, x_3, \dots, x_n)$$

¿Cómo podemos calcular la probabilidad que X sea clasificado como y ?

$$P(y/X) = P(y/x_1, x_2, x_3, \dots, x_n)$$

¿Qué pasa cuando tenemos un numero de características muy alta?

$$P(y/X) = \frac{P(X/y)P(y)}{P(X)}$$

$$P(X/y) = P(x_1, x_2, x_3, \dots, x_n/y)$$

$$\begin{aligned} P(X/y) &= P(x_1/y) P(x_2, x_3, \dots, x_n/y, x_1) \\ &= P(x_1/y) P(x_2 /y, x_1) P(x_3, \dots, x_n/y, x_1, x_2) \end{aligned}$$

Clasificador Inexperto de Bayes (Naive Bayes Classifier)

¿Qué se puede hacer para simplificar lo anterior?

El modelo de Naive Bayes, asume que la probabilidad en la que ocurre una característica es independiente de las demás. Lo cual es un supuesto ingenuo o Naive.

Por tanto...

$$\begin{aligned} P(X/y) &= P(x_1, x_2, x_3, \dots, x_n/y) \\ &= P(x_1/y)P(x_2/y) \dots P(x_n/y) \\ P(X/y) &= \prod_{i=1}^n P(x_i/y) \end{aligned}$$

Clasificador Inexperto de Bayes (Naive Bayes Classifier)

Entonces la ecuación final del modelo es ...

$$P(y/X) = \frac{P(y)}{P(X)} \prod_{i=1}^n P(x_i/y)$$
$$P(X) = P(x_1)P(x_2) \dots P(x_n)$$

Como permanece constante para una entrada dada, podemos dejarlo de lado, entonces ...

$$P(y/X) \propto P(y) \prod_{i=1}^n P(x_i/y)$$

Clasificador Inexperto de Bayes (Naive Bayes Classifier)

Por tanto, el objetivo del clasificador es ...

$$C = \underset{y_j \in Y}{\operatorname{argmax}} P(y_j) \prod_{i=1}^n P(x_i/y_j)$$

Naive Bayes Classifier

- Ejemplo: aprender a clasificar textos (clasificar sitios web por tipo, clasificar noticias de acuerdo a relevancia nacional, etc...)
- Naive Bayes escala muy bien a grandes set de datos
 - 10.000 atributos, papita!
 - Millones de registros, pan comido!

Ejemplo: Clasificación de textos en newsgroups

- **Datos**

Archivos de textos provenientes de 20 grupos de noticias (deportes, política, etc.)

- **Objetivo**

Usar set de entrenamiento para construir un clasificador de textos de noticias basado en Naive Bayes

- **¿Cómo?**

¿Alguna idea?

- **Resultados**

Exactitud en la clasificación sobre 90%

- Código disponible en:

<http://www.cs.cmu.edu/~mccallum/bow/rainbow/>

Naive Bayes Classifier

- Problemas:

Supuesto de independencia condicional comúnmente no se cumple. Sin embargo en la práctica el clasificador entrega buenos resultados. En realidad no se necesita estimar la función a posterior sino sólo la clase que la maximiza

En la práctica hay problemas cuando una de las clases nunca contiene un cierto atributo:

Solución: $\hat{P}(a_i / v_j) = 0 \Rightarrow \hat{P}(v_j) \prod_i \hat{P}(a_i / v_j) = 0$

$$\hat{P}(a_i / v_j) \approx \frac{n_c + mp}{n + m} \quad (\text{m-estimate})$$

tamaño de muestra
equivalente

Probabilidad a priori para
las clases, usualmente uniforme

Volvamos al caso de la regla de Bayes

- B=la regla de Bayes es super
- N=hoy está nublado
- $P(B,N)$ contiene 4 posibles entradas, ¿Cuáles?
- Gracias a la independencia sólo necesitamos especificar $P(B)$ y $P(N)$

Ejemplo $P(B)=0.9$ $P(N)=0.3$



B	N	$P(B,N)$
T	T	
T	F	
F	T	
F	F	

¿Qué pasa si tenemos más variables?

- B=la regla de Bayes es super
- N=hoy está nublado
- C=los alumnos de IA están contentos
- ¿Cuántas entradas tiene $P(B,N,C)$
(nuevamente consideramos variables binarias) ?
- ¿Cuántas si tenemos 10 variables ?

Dos Extremos

- **Mr. Hard:** te dije que esto sería así, lo sabia, ahora tenemos que modelar toda la probabilidad conjunta...
- **Mr. Naive Bayes:** relajate, vive la vida, asumamos que todo es independiente dado la clase...

Mr. Bayes Net al Rescate

- Que tal si usamos la super arma: causalidad (i.e, independencia, conocimiento, creencia, belief, tarot, karma....)

Mr. Bayes Net al rescate

- B=la regla de Bayes es super
- N=hoy está nublado
- C=los alumnos de Miguel están contentos

CAUSALIDAD

$$P(B/N)=P(B)$$

$$P(B)=$$

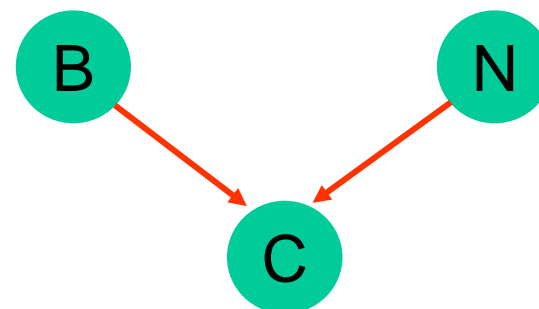
$$P(N)=$$

$$P(C/N,B)=$$

$$P(C/N,\neg B)=$$

$$P(C/\neg N,B)=$$

$$P(C/\neg N,\neg B)=$$



Platón: Sócrates esto es fantástico...

Sócrates: realmente lo es, pero aún quedan algunas malas noticias.

Platón: ¿malas noticias maestro ?

Sócrates: si, pero por ahora vivamos nuestro momento de gloria.

Momento de Gloria

- Probabilidades son excelentes para razonar bajo incerteza.
- Probabilidad conjunta el sueño dorado que se hace realidad gracias a Mr. Bayes.
- Ahora podemos razonar probabilísticamente sin miedos
 - $P(\text{problema/síntomas})$
 - ¿Cuán anómalo es cierto caso ?
 - ¿Cuál es la probabilidad de estar interesado en cierto producto?
 - ...

Importante: una red de Bayes representa una factorización de la probabilidad conjunta, por tanto son equivalentes y permiten inferir cualquier relación entre las variables.

Redes de Bayes de terno y corbata, o traje 2 piezas según corresponda

- Una red de Bayes (Bayes net or Bayesian belief network): es una representación gráfica acíclica que permite expresar distribuciones de probabilidad conjunta y que esta determinada por un set (V, E) :
 - V : es el set de vértices
 - E : es el set de uniones dirigidas que no pueden formar ciclos cerrados (loops)
- Cada vértice V_i en V contiene información de una variable aleatoria y de las probabilidades condicionales de esta variable dado sus nodos padres.

Redes de Bayes

- Necesitamos
 1. Estructura
 2. Probabilidades condicionales $P(\text{nodo}/\text{padres})$
 3. Algoritmos de Inferencia

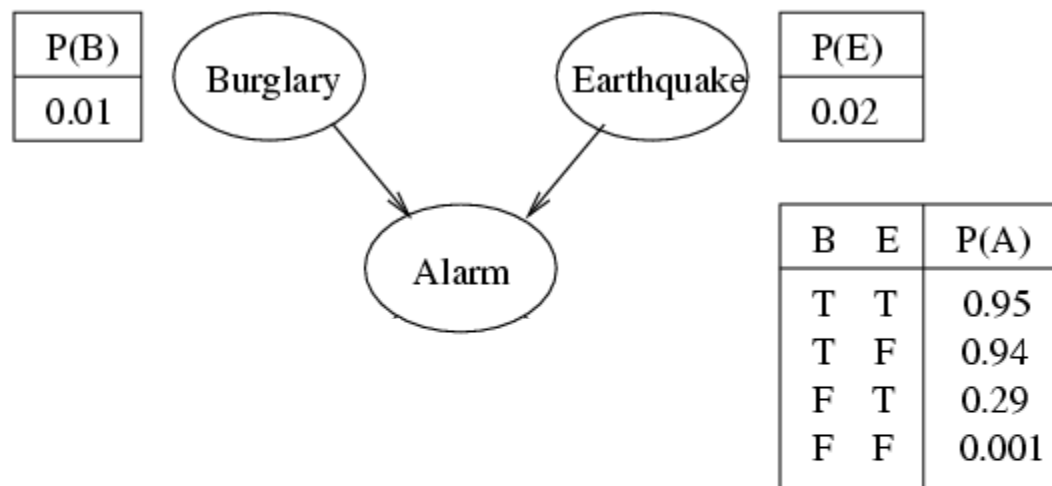
Redes de Bayes: Estructura y Probabilidades

- Algunas veces es posible obtener la estructura de la red y las probabilidades condicionales manualmente usando conocimiento de un experto.
- Algunas **más** veces es posible obtener la estructura usando conocimiento de un experto y datos para inferir las probabilidades condicionales.
- Algunas **más más** veces la única posibilidad es obtener tanto la estructura como las probabilidades condicionales usando datos.

Ejemplo 1: Estructura Manual, Caso Temblor

- Nuevo sistema de alarma en la casa
- Funciona para detectar a los malos pero también se activa con temblores
- ¿Cuáles son las variables relevantes para hacer inferencia sobre la alarma?
- ¿Algún comentario sobre independencia?
- ¿Algún comentario sobre la estructura causal de la red?
- ¿Algún comentario sobre las pdf's?
- ¿Probabilidad alarma? ¿Alarma dado robo?

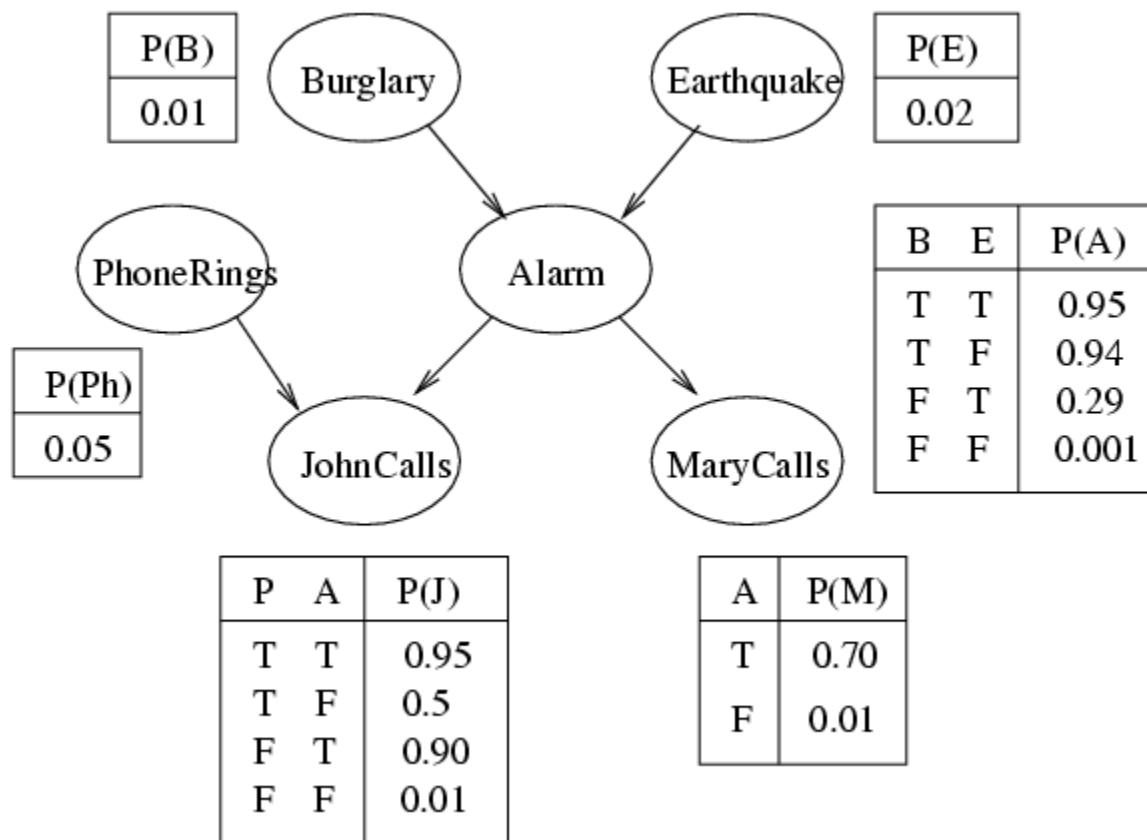
Ejemplo 1: Estructura Manual, Caso Temblor



Ejemplo 1 Extendido

- Nos vamos de viaje y nuestros vecinos John y Mary quedan encargados de llamar si pasa algo en nuestra casa
- John siempre llama cuando suena la alarma pero a veces se confunde entre la alarma y el sonido del teléfono
- Mary es un poco sorda así que no es tan confiable
- Dada evidencia de un llamado y su autor, ¿Qué podríamos inferir acerca de un posible robo?

Ejemplo 1 Extendido



Ejemplo 2: Caso Cáncer

- Un estudio ha revelado que dos importantes causas en el cáncer al pulmón son la contaminación y el fumar
- Dos formas de detectar este tipo de cáncer son rayos-X y observar problemas respiratorios en el paciente
- ¿Cuál sería la red?

Ejemplo 2: Caso Cáncer Revisado

- ¿Qué pasa si ahora agregamos bronquitis y tuberculosis a nuestro problema ?
- Tuberculosis también produce problemas respiratorios y rayos-X detectan esta enfermedad
- Bronquitis produce problemas respiratorios pero no produce rayos-X positivos
- Suponga también que es posible contar con info acerca de si paciente estuvo en Asia, donde la tuberculosis es un gran problema
- ¿Cuál sería la red?

Redes de Bayes: Algoritmos de Inferencia

- **Mr. AguaFiesta:** alguien me dijo por ahí que inferencia en una BN era un problema de complejidad NP.
- **Mr. Positivo:** algo hay de eso pero fíjate que existen casos en que la complejidad es lineal, por ejemplo cuando la estructura de la red corresponde a un árbol.

En este curso no profundizaremos en algoritmos de inferencia para redes de Bayes.

Más información sobre esto en: IIC3635 Métodos Estadísticos para el Aprendizaje de Máquina.

La Hora del recuento

- Probabilidades permiten modelar incerteza.
- Probabilidad condicional permite modelar problemas del tipo $P(\text{hipótesis/observación})$.
- Gran problema es encontrar funciones de densidad de probabilidad.
- Bayes permite unir observaciones y conocimiento previo.
- Naive Bayes permite escalar a problemas multidimensionales usando independencia condicional.

La Hora del recuento

- Software con aplicaciones de estas tecnologías:

www.hugin.com, www.norsys.com.