Naïve Bayes

Introducción

Es un método que no sólo ofrece un análisis cualitativo de las atributos y valores que pueden intervenir en el problema, sino da cuenta también de la importancia cuantitativa de esos atributos.

En el aspecto **cualitativo** se puedes representar cómo se relacionan esos atributos ya sea en una forma causal, o señalando simplemente de la correlación que existe entre esas variables (o atributos).

Introducción

Cuantitativamente (y el gran aporte de los métodos bayesianos), da una medida probabilística de la importancia de esas variables en el problema (y por lo tanto una probabilidad explícita de las hipótesis que se formulan).

Características

Cada ejemplo observado va a modificar la probabilidad de que la hipótesis formulada sea correcta (aumentándola o disminuyéndola).

Es decir, una hipótesis que no concuerda con un conjunto de ejemplos más o menos grande no es desechada por completo sino que lo que harán será disminuir esa probabilidad estimada para la hipótesis.

Características

Estos métodos son robustos al posible ruido presentes en los ejemplos de entrenamiento y a la posibilidad de tener entre esos ejemplos de entrenamiento datos incompletos o posiblemente erróneos.

Los métodos bayesianos permiten tener en cuenta en la predicción de la hipótesis el conocimiento a priori o conocimiento del dominio en forma de probabilidades. El problema puede surgir al tener que estimar ese conocimiento estadístico sin disponer de datos suficientes.

Clasificación de patrones

Dado un conjunto de datos (divididos en dos conjuntos de entrenamiento y de test) representados por pares <*atributo*, *valor*>, el problema consiste en encontrar una función *f(x)* (llamada hipótesis) que clasifique dichos ejemplos.

La idea de usar el teorema de Bayes en cualquier problema de aprendizaje automático, es que se puede estimar las probabilidades a posteriori de cualquier hipótesis consistente con el conjunto de datos de entrenamiento para así escoger la hipótesis más probable.

Clasificador basado en Naïve Bayes

Dado un ejemplo *x* representado por *k* valores el clasificador Naïve Bayes se basa en encontrar la hipótesis más probable que describa a ese ejemplo.

Si la descripción de ese ejemplo viene dada por los valores $<a_1, a_2, ..., a_n>$ (atributos), la hipótesis más probable será aquella que cumpla:

$$v_{MAP} = argmax_{v_j \in V} P(v_j | a_1, \dots, a_n)$$

donde v_j son las posibles clases o valores de *f(x).*

MAP - "Maximum a Posteriori"

Clasificador basado en Naïve Bayes

Es decir, la probabilidad de que conocidos los valores que describen a ese ejemplo, éste pertenezcan a la clase v_j (donde v_j es el valor de la función de clasificación *f(x)* en el conjunto finito *V*).

Por el teorema de Bayes:

$$v_{MAP} = argmax_{x_j \in V} \frac{P(a_1, \ldots, a_n | v_j) p(v_j)}{P(a_1, \ldots, a_n)}$$

$$= argmax_{x_j \in V} \mathrm{P}(a_1, \ldots, a_n | v_j) p(v_j)$$

Clasificador basado en Naïve Bayes

Se puede estimar $P(v_j)$ contando las veces que aparece el ejemplo v_j en el conjunto de entrenamiento y dividiéndolo por el número total de ejemplos que forman este conjunto.

Para estimar el término $P(a_1, \ldots, a_n|v_j)$, es decir, las veces en que para cada categoría aparecen los valores del ejemplo *x*, debo recorrer todo el conjunto de entrenamiento. Siendo esto **imposible**.

Clasificador basado en Naïve Bayes

Por lo que se recurre a la hipótesis de independencia condicional con el objeto de poder factorizar la probabilidad. Esta hipótesis dice lo siguiente:

Los valores a_j que describen un atributo de un ejemplo cualquiera x son independientes entre sí conocido el valor de la categoría a la que pertenecen.

Así la probabilidad de observar la conjunción de atributos a_j dada una categoría a la que pertenecen es el producto de las probabilidades de cada valor por separado: $P(a_1, \ldots, a_n | v_i) = \prod_i P(a_i | v_j)$,

Teorema de Bayes

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Donde:

- *P(h)* es la probabilidad a priori de la hipótesis h.
- *P(D)* es la probabilidad de observar el conjunto de entrenamiento D.
- *P(D|h)* es la probabilidad de observar el conjunto de entrenamiento D en un universo donde se verifica la hipótesis h.
- *P(h|D)* es la probabilidad a posteriori de h, cuando se ha observado el conjunto de entrenamiento D.

Teorema de Bayes

Máximo a posteriori (MAP): Se denomina así a la hipótesis más probable aplicando el teorema de Bayes.

$$h_{MAP} = \underset{h \in H}{argmax} P(h|D)$$

$$= \underset{h \in H}{argmax} \frac{P(D|h)P(h)}{P(D)}$$

$$= \underset{h \in H}{argmax} P(D|h)P(h)$$

Teorema de Bayes

En algunos casos la multiplicación por *P(h)* no tiene sentido ya que las hipótesis son equiprobables:

$$h_{ML} \equiv \underset{h \in H}{argmax}\, P(D|h)$$

A este resultado se le denomina máxima verosimilitud (**maximum likelihood**).

Ejemplo

X1	X2	X3	Y
0	0	1	0
0	1	0	0
1	1	0	0
0	0	1	1
1	1	1	1
0	0	1	1
1	1	0	1

P(Y= 0) = 3/7 P(Y = 1) = 4/7

A que clase será asignado el registro (X_1=0,X_2=0,X_3=1)?

P(X_1=0,X_2=0,X_3=1|Y=0) = P(x_1=0|Y=0)P(X_2=0|Y=0)P(x_3=1|Y=0) = (2/3)(1/3)(1/3) = 2/27

P(X_1=0,X_2=0,X_3=1|Y=1) = P(x_1=0|Y=1)P(X_2=0|Y=1)P(x_3=1|Y=1) = (2/4)(2/4)(3/4) = 12/64 = 6/32 = 3/16

Ejemplo

Como

$$(3/7)(2/27) < (4/7)(3/16)$$

$$0,032 = 0,429*0,074 < 0,571*0,188 = 0,107$$

entonces (X_1=0, X_2=0, x_3=1) será asignado a la clase 1.

Si el objeto esta asignado a la clase 0 entonces NB comete un error.

Atributos

Atributos discretos: la probabilidad condicional es la frecuencia de aparición en los datos de entrenamiento, es decir el número de casos en los que los atributos toma un determinado valor para la una clase, dividido el número total de casos. (Hay otros estimadores, por ej. Laplace.)

Atributos continuos: suponer que el atributo sigue una distribución normal, calcular media y desviación estándar.

Características

Desventajas:

- Sobreajuste con pocos valores
- Interpretabilidad del modelo
- Supone que las variables son independientes