Estudio detallado del algoritmo de Naive – Bayes:

Naïve Bayes es uno de los clasificadores más utilizados por su simplicidad y rapidez.

Se trata de una técnica de clasificación y predicción supervisada que construye modelos que predicen la probabilidad de posibles resultados. Constituye una técnica supervisada porque necesita tener ejemplos clasificados para que funcione, como ya veremos.

ESTÁ BASADA EN EL TEOREMA DE BAYES...

Está basada en el Teorema de Bayes, tambiéń conocido como teorema de la probalibidad condicionada. Existe una gran variedad de fuentes para el que se quiera sumergir más en el mundo de la probablidad condicional, por ahora analizemos un pequeño ejemplo para entenderlo mejor:

"Supongamos que un ingeniero esta buscando agua en un terreno. A priori, se sabe que la probabilidad de que haya agua en dicha finca es del 60%. No obstante, el ingeniero quiere asegurarse mejor y decide realizar una prueba que permite detectar la presencia o no de agua. Dicha prueba tiene una fiabilidad del 90%, es decir, habiendo agua, la detecta en el 90% de los casos. También, cuando realmente no hay agua, la prueba predice que no hay agua en el 90% de los casos.
Por tanto, pudiendo hacer uso de dicha prueba ¿qué es más probable, que haya agua o que no?"

Como vemos, en el enunciado tenemos una propabilidad a priori (60% de que haya agua) que se ve afectada por otra propabilidad (el 90% de acierto de la prueba). Para resolver esta problemática, Thomas Bayes nos dió la solución. La probabilidad de que se dé un suceso (que haya agua) habiendo sucedido otro que influye en el anterior (que la prueba diga que sí hay agua), se define con la siguiente fórmula:

$$P(A|B)=\frac{P(B|A)\cdot P(A)}{P(B)}$$

¿Qué pasaría si existieran más pruebas para detectar si hay agua? Supongamos que se aplican nuevas pruebas, identificadas como las pruebas 1, 2, 3 y 4. En este caso tendriamos que calcular la probabilidad de que haya agua sabiendo que todas las pruebas han dicho que hay agua. Es decir: P(Agua|P1+,P2+,P3+,P4+). ¿Pero cómo se calcula esto?

Tras aplicar el Teorema de Bayes y realizar diversos cálculos, se llega a la conclusión de que:

$$P(A|b_1,b_2,b_3,b_4)=P(A)\cdot(P(b_1|A)\cdot P(b_2|A)\cdot P(b_3|A)\cdot P(b_4|A))$$

Si generalizamos, obtenemos la siguiente expresión:

$$P(A|b_1,b_2,\ldots,b_n)=P(A)\cdot P(b_1,b_2,\ldots,b_n|A)=P(A)\cdot\prod_{i=1}^{n}P(a_i|A)$$

Si quisiéramos resolver nuevamente el enunciado anterior, deberíamos de volver a calcular la probabilidad de que haya agua y de que no haya agua, y quedarnos con la mayor. Como siempre nos quedamos con la mayor, podemos añadir un matiz a la fórmula anterior:

$$Solucion = arg \max_{i=1}^{n} P(c_i) \cdot \prod_{j=1}^{m} P(a_j|c_i)$$

Fórmula de Naïve Bayes

Si volvemos sobre nuestros pasos, podemos apreciar cada hecho (que haya agua y que no haya) como una clase. También nos damos cuenta, que al final siempre nos quedamos con una única clase. Incluyo podemos hacer una correspondencia entre las pruebas de agua y posibles atributos o características de un problema en cuestión. Por tanto, se puede modelar este problema como un problema de clasificación supervisada: dado un nuevo ejemplo, calculamos la probabilidad de que pertenezca a cada clase y nos quedamos con el resultado que emita mayor probabilidad.

Un nuevo ejemplo nos aclarará las ideas. Supongamos un típico problema de clasificación en Minería de Datos, donde tenemos unos cuantos ejemplos que queremos clasificar. Cada ejemplo (también llamados instancias), se caracteriza por una serie de atributos. Además como estamos en clasificación supervisada, dichos ejemplos deben estar etiquetados, es decir, se debe definir de qué clase son. Por tanto, tenemos un conjunto de 4 ejemplos con 3 atributos cada uno y 2 posibles clases.

Ejemplos	Atr. 1	Atr. 2	Atr. 3	Clase
x1	1	2	1	positiva
x2	2	2	2	positiva
x3	1	1	2	negativa
x4	2	1	2	negativa

Caben destacar dos partes en el algoritmo. La primera es la construcción del modelo, y la segunda clasificar un nuevo ejemplo con dicho modelo:

PARTE 1: Crear el modelo.
Para ello se necesitan **cuatro pasos**:

1.Calcular las probabilidades a priori de cada clase.

2.Para cada clase, realizar un recuento de los valores de atributos que toma cada ejemplo. Se debe distribuir cada clase por separado para mayor comodidad y eficiencia del algoritmo.

3.Aplicar la Corrección de Laplace, para que los valores "cero" no den problemas.

4.Normalizar para obtener un rango de valores [0,1].

A continuación se detalla cada paso:

Paso 1: Hay varias formas de establecer la probabilidad a priori a cada clase. La más intuitiva es que todas ellas tengan la misma probabilidad, es decir, 1 divido entre el número de clases. Si contamos con la opinión de un experto, puede que éste nos proporciones dichas probabilidades. En este caso, vamos a considerar cada clase equiprobable. Por tanto:

P(positivo) = 1/2 = 0,5
P(negativo) = 1/2 = 0,5

Paso 2: Se crea una tabla (un otro contenedor) para cada clase y se hace un recuento de valores:

positivo	valor 1	valor 2
atr.1	1	1
atr.2	0	2
atr.3	1	1

negativo	valor 1	valor 2
atr.1	1	1
atr.2	2	0
atr.3	0	2

Paso 3: Se suma una unidad a cada valor de la tabla anterior.

positivo	valor 1	valor 2
atr.1	2	2
atr.2	1	3
atr.3	2	2

negativo	valor 1	valor 2
atr.1	2	2
atr.2	3	1
atr.3	1	3

Paso 4: Se normalizan todos los valores de la tabla del siguiente modo. Cada celda se divide por la suma de los valores de la fila. Por ejemplo, el valor: "valor 1" del atributo: "atr. 2" se actualiza con: 1 / (1+3) = **0,25.**

positivo	valor 1	valor 2
atr.1	0,5	0,5
atr.2	**0,25**	0,75
atr.3	0,5	0,5

negativo	valor 1	valor 2
atr.1	0,5	0,5
atr.2	0,75	0,25
atr.3	0,25	0,75

Paso 2: Se aplica la fórmula. Para cada clase se calcula su probabilidad. Recordemos la fórmula:

$$Solucion = arg \max_{i=1}^{n} P(c_i) \cdot \prod_{j=1}^{m} P(a_j | c_i)$$

Fórmula de Naïve Bayes

Por tanto:

P(positivo | x5) = 0,5 · (0,5 · 0,25 · 0,5) = 0,03125
P(negativo | x5) = 0,5 · (0,5 · 0,75 · 0,25) = ***0,046875***

Como la clase más probable es "negativa", clasificamos el nuevo ejemplo (x5) como **negativo**.

En lo que se refiere al pre-procesamiento de los datos, estos se deberán limpiar pues existen varios registros que están vacios, en cuantoa la integracion de los datos no será necesario pues los datos provienen de una misma fuente, la transformación tampoco será necesaria pues no se trabaja con muchos datos numéricos, lo mismo para la discretización y reducción.