



Naive Bayes



Probabilidad condicional.

La **probabilidad condicionada** de un suceso A a otro suceso B, es la probabilidad de que ocurra A sabiendo que B ya ha ocurrido. Esta probabilidad se denota como $P(A|B)$ y se calcula de la siguiente manera:

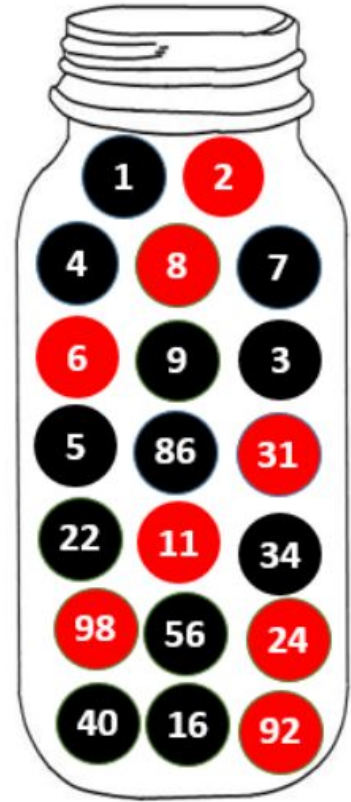
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Ejemplo:

Supongamos que tenemos un frasco con 20 bolas, 8 de ellas son rojas y 12 son negras. Todas ellas están numeradas.

Consideremos A como el evento de sacar una bola par y B como sacar una bola negra. Si suponemos que B ya ha ocurrido, entonces la probabilidad condicionada de A a B es

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{7}{20}}{\frac{12}{20}} = \frac{7}{12}$$





Teorema de Bayes.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{\text{Anterior} \times \text{Probabilidad evidencia}}{\text{evidencia}}$$



Naive bayes.

- Es una **familia de modelos** de clasificación
- Se basa en elegir la categoría con **más probabilidad** de acuerdo a las características dadas

Formulación:

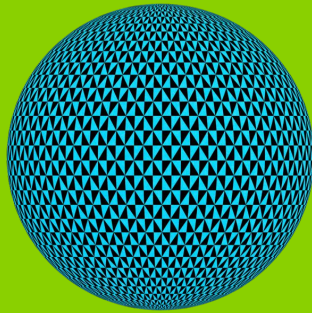
Consideramos una clase **C** que **desconocemos**, esta depende de ciertas características F_1, F_2, \dots, F_n que son **independientes** entre sí. Por ejemplo: C puede ser cualquiera de estas frutas.



F_1, F_2, \dots, F_n son las características de las frutas, el color, peso, forma, longitud, grosor, volumen, etc ...

Lo que queremos encontrar es la probabilidad de que C sea una manzana, una pera, una sandía o alguna otra fruta, de acuerdo a las características de la fruta que nosotros queremos clasificar.

180 gramos



Si sabemos que la fruta es verde, pesa 180 gramos y es esférica, Cuál es la probabilidad de que sea una sandía?





En términos matemáticos:

$$P(C|F_1, F_2, \dots, F_n) = \frac{P(C)P(F_1, F_2, \dots, F_n|C)}{P(F_1, F_2, \dots, F_n)}$$



Como suponemos F_1, F_2, \dots, F_n variables independientes, tenemos que:

$$P(C|F_1, F_2, \dots, F_n) = \frac{P(C)P(F_1, F_2, F_3, \dots, F_n|C)}{P(F_1, F_2, \dots, F_n)} = \frac{P(C)P(F_1|C)P(F_2|C), \dots, P(F_n|C)}{P(F_1, F_2, \dots, F_n)}$$

En la práctica se desprecia el denominador, pues este es constante y el numerador es variable.



Por lo anterior, tenemos que

$$P(C|F_1, \dots, F_n) \propto P(C) \prod_{i=1}^n P(F_i|C)$$



Para clasificar elegimos la categoría que maximice:

$$\hat{C} = \arg_C \max P(C) \prod_{i=1}^n P(F_i|C)$$

- $P(C)$ se calcula como la **frecuencia** de la categoría C en los datos de entrenamiento.
- $P(F_i|C)$ es la **probabilidad dada por una función de distribución** supuesta por el usuario y con parámetros determinados por los datos de entrenamiento.



Tipos de clásicos de Naive bayes:

- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Bernoulli Naive Bayes

Todos ellos calculan sus parámetros por medio de máxima verosimilitud con ayuda de los datos de entrenamiento.



Ejemplo:

Supongamos la siguiente tabla de alturas y pesos.

Construiremos un clasificador de Naive bayes.

	Altura(pies)	Peso(lbs)
Mujer	5.75	150
Mujer	5.42	130
Mujer	5.5	150
Mujer	5	100
Hombre	5.92	165
Hombre	5.58	170
Hombre	5.92	190
Hombre	6	180



Supondremos una distribución normal y sus parámetros son:

	Media Altura	Varianza Altura	Media Peso	Varianza Altura
Mujer	5.41	0.09	132.5	558.33
Hombre	5.85	0.03	176.25	122.92

Como de ambas categorías hay el mismo número, entonces las frecuencias de ambas son $P(\text{Mujer})=P(\text{Hombre})=0.5$.



Después para cualquier entrada a predecir con dato (altura,peso) se calcula:

Mujer:

$$0.5 \times \frac{1}{\sqrt{2\pi \times 0.09^2}} \exp \left(\frac{-(altura-5.41)^2}{2 \times 0.09^2} \right) \times \frac{1}{\sqrt{2\pi \times 558.33^2}} \exp \left(\frac{-(peso-132.5)^2}{2 \times 558.33^2} \right)$$

Hombre:

$$0.5 \times \frac{1}{\sqrt{2\pi \times 0.03^2}} \exp \left(\frac{-(altura-5.85)^2}{2 \times 0.03^2} \right) \times \frac{1}{\sqrt{2\pi \times 122.92^2}} \exp \left(\frac{-(peso-176.25)^2}{2 \times 122.92^2} \right)$$

Este dato se clasifica, según cuál de las dos numeros es mayor.

Evaluación del modelo.

Para problemas de clasificación se puede hacer uso de matrices de confusión. Para casos binarios la matriz tiene la siguiente forma y métricas:

Confusion matrix

Actual classes	Negative 0	True Negatives (TN)	False Positive (FP)
	Positive 1	False Negative (FN)	True Positive (TP)
		Negative 0	Positive 1
		Predicted classes	

$$F_1\text{score} = \frac{2TP}{2TP+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Accuracy} = \frac{TP+TN}{P+N}$$



Algunas propiedades de las métricas:

- **Precision y Recall** son muy parecidas, una da más importancia a los verdaderos positivos y otra a falsos positivos. Ambas son útiles en datos desbalanceados, pero siempre es conveniente elegir una.
- **F1 score** es una media armónica entre precision y recall. Entre más cercano 1 es mejor el modelo y entre más cerca a 0 es peor.



Ventajas:

- Es **simple** y **básico**.
- Es **rápido** y **eficiente** computacionalmente.
- Puede manejar datos en **alta dimensión**.
- Necesita **pocos datos** para determinar los parámetros para clasificar.



Desventajas:

- Está sujeto a que las **variables de entrada independientes** , lo cual se ve raramente en aplicaciones a la vida real.
- Se debe de saber elegir la **distribución** a usar.
- Es superado por otros modelos más recientes como boosted trees o random forest.



Aplicaciones:

- Filtrado de spam.
- Clasificación de documentos.
- Análisis de sentimientos.
- Predicciones del estado mental.