# Clasificadores y Fundamentos Matemáticos

En este tema se vieron algunos elementos que nos permiten clasificar información habiendo realizado una clasificación previa, para ello también se tuvo que ver algunos temas sobre fundamentos matemáticos para conseguir las operaciones necesarias.

**Clasificador por Centroides**

Este es uno de los clasificadores más sencillos, por lo mismo tiene bastante más margen de errores. El algoritmo consiste en determinar un punto en el plano de dos dimensiones, donde podría decirse que "la clase se concentra", el cual se calcula con el promedio de los valores en cada dimensión.

Ya que se cuenta con los centroides de cada clase, pasamos ahora a la clasificar valores según lo "aprendido". Ello consiste en ubicar un nuevo valor en el plano y calcular la distancia de este elemento a cada centroide de clase, el centroide más cercano será el mejor candidato.

**Clasificador Gaussiano**

Otro clasificador visto es el basado en la campana de gauss, consiste en tomar todos los valores de una clase, y de ellos sacar la media y varianza de los datos.

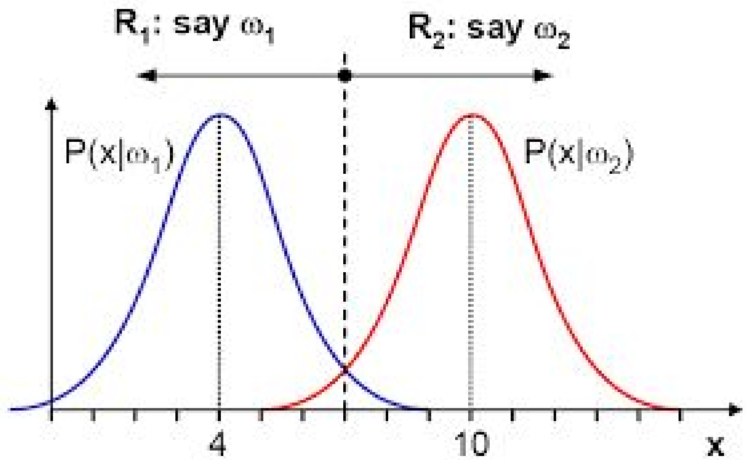
$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

*Media*

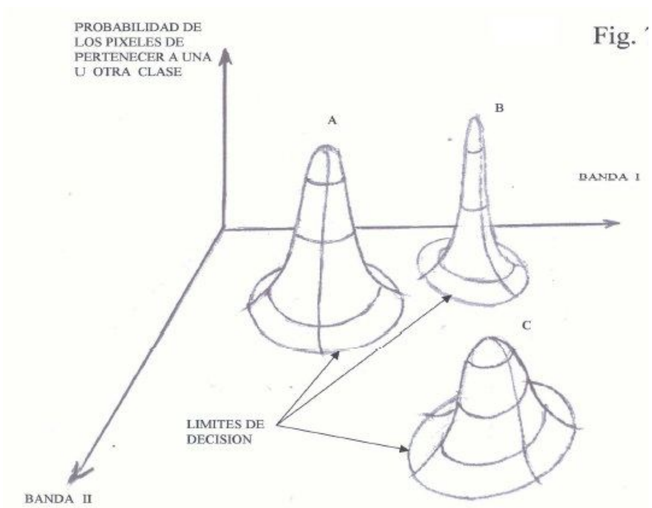
$$s_x^2 = \frac{\sum_{i-1}^{n}(x_i - \overline{x})^2}{n-1}$$

*Varianza (tomando el valor de S)*

Ya que poseemos estos valores, podríamos graficar la campana de gauss, donde el área bajo la curva es igual a 1, y aprovechando esta propiedad podemos calcular valor puntual en la curva de la campana para cada dimensión del parámetro que estamos por clasificar, entonces en la clase donde se adquiera el valor gaussiano más alto, hay una fuerte probabilidad de que el elemento pertenezca a esta clase.




Cuando realizamos el cálculo para dos dimensiones conseguimos como resultado algo como esto, donde los valores de Banda I y Banda II son los valores X y Y del elemento que estamos por clasificar.




Del mismo modo se toma el valor de gauss más grande como el mejor candidato de clase para nuestro elemento.

**Probabilidad a Priori y a Posteriori**

La probabilidad a Priori es aquella probabilidad donde no se tiene base en ningún hecho previo, solo se conocer el porcentaje de posibilidad de un evento el cual se considera independiente. En cambio, la probabilidad a Posteriori es aquella que depende de ciertas condiciones, tiene que haber un evento verdadero previo a la probabilidad que queremos calcular para poder conocerlo; esta probabilidad depende de otros hechos.

La probabilidad conjunta es aquella donde se tienen dos eventos independientes donde se quiere verificar la posibilidad de dos o más hechos. Esta probabilidad es calculada mediante el producto de las dos probabilidades.

La Regla de la Cadena nos dice que la probabilidad conjunta de:

$$P(A,B,C,D...) = P(A) \cdot P(B \mid A) \cdot P(C \mid A,B) \cdot P(D \mid A,B,C)...$$

Y eso nos lleva al teorema de bayes:

$$P(M \mid R) = P(M \mid R) \cdot P(R) = P(R \mid M) \cdot P(M)$$

*Teorema de Bayes*

$$P(M \mid R) = \frac{P(M \mid R) \cdot P(M)}{P(R)}$$

**Clasificador Bayesiano**

El clasificador Bayesiano es ampliamente utilizado como clasificador puesto que permite conseguir un alto nivel efectividad de un modo muy rápido. Algunos de sus usos son detección de spam, diagnósticos de enfermedades, detección de idiomas, entre otros. Para poder conseguir un alto nivel de precisión, se debe haber entrenado con un conjunto extenso ya clasificado.

*Clasificador de Naive Bayes*

$$NB(W) = arg_c\, max\, P(W|C) \cdot P(C)$$
$$C = \{ B, \neg B \}$$

Entonces, al momento de usar este clasificador, podemos decir que la probabilidad de pertenecer a esta clase es el producto de las probabilidades de cada propiedad

$$P(A) = \prod_{i=1}^{N} P(F_i \mid A)$$

Sin embargo, como sabemos, el producto de dos número entre 0 y 1, es un número decimal aún más pequeño, entonces si tuviéramos un producto de varios números ínfimos, no podríamos representarlos con los bits de la computadora, para ello recurrimos a una propiedad de logaritmo:

$$log(A \cdot B) = log(A) + log(B)$$

Entonces nuestra probabilidad será:

$$P(A) = \sum_{i=1}^{N} log(P(F_i \mid A))$$

Para llevar a cabo el proceso de la toma de decisiones recurrimos al teorema de decisión de bayes:

$$\frac{P\ (B\mid W)}{P\ (\neg B\mid W\ )} > 1$$

Desarrollando:

$$\frac{\frac{P\ (W\mid B)\cdot P\ (B\ )}{P\ (W\ )}}{\frac{P\ (W\mid \neg B)\cdot P\ (\neg B\ )}{P\ (W\ )}} > 1$$

Sin embargo podemos eliminar ese factor repetido (tiene valor 1 dentro del producto) y ya con eso solo nos concentramos en las otras probabilidades. Si la probabilidad es mayor en un elemento a comparación de la otra, este pertenece a esa clase, en caso contrario, a la otra clase.

**PointWise Mutual Information (PMI)**

Está basado en el bayesiano y es una medida de asociación de dos palabras usado en la teoría de información y estadística. A diferencia de Mutual Information (MI), este refiere a eventos únicos mientras que MI refiere a el promedio de todos los posibles eventos.

Partiremos de

$$P\ (A,B)\ =\ P(A|B)$$

$$\frac{P\ (A\ ,B)}{P\ (\ A\mid B\ \ )\ P\ (B)} = 1 \quad \textit{A y B son independientes}$$

$$log(\frac{P\ (A\ ,B)}{P\ (\ A\ )\ P\ (B)})\ =\ 1 \text{ y } log(1)\ =\ 0$$

Sabemos que las probabilidades de las palabras dentro del texto son:

$$P\ (A,B)\ =\ \frac{freq\ (A\ ,B)}{N}$$

$$P\ (A)\ =\ \frac{freq\ (A)}{N}$$

$$P\ (B)\ =\ \frac{freq\ (B)}{N}$$

Si sustituimos estas probabilidades en la fórmula de logaritmo anterior conseguimos lo siguiente:

$$log(\frac{\frac{freq\,(A\,,\,B)}{N}}{\frac{freq\,(A)}{N}\,.\,\frac{freq\,(B)}{N}}) \;=\; 1$$

Simplificando tenemos lo siguiente:

$$log(\frac{N \times freq\,(A\,,\,B)}{freq\,(A)\;\times\;freq\,(B)})$$

Esto es la fórmula de PMI para determinar la relación de dos palabras.

$$PMI(A,B) \;=\; log(\frac{N \times freq\,(A\,,\,B)}{freq\,(A)\;\times\;freq\,(B)})$$