

Clasificador Basado en el Teorema de Bayes



26 MARZO

Pattern Recognition
María Elena Cruz Meza



Introducción a los Clasificadores estadísticos probabilísticos

Temas relacionados con las Unidades I y II

“Selección de rasgos y los Clasificadores Estadísticos-probabilísticos”

a) Clasificación de un lote de frutas mediante el Clasificador Bayesiano Simple

Dado el ejemplo visto en clases (consultar notas), consideremos ahora que se agrega una clase más al CMA que deberá reconocer el clasificador que se desea diseñar.

Supongamos ahora que deseamos clasificar automáticamente un lote de frutas de dos tipos: manzanas, mangos y peras, con una muestra de entrenamiento que consiste de 12000 frutas. De este lote de frutas, se disponen 5000 piezas de manzanas, 4000 piezas de mango y 2000 piezas correspondientes a las peras. Si deseamos iniciar el entrenamiento o aprendizaje para estos dos tipos de clases mediante el clasificador bayesiano simple. Si tomamos como rasgo al peso de una fruta a clasificar, considerar que la longitud de la fruta varía entre i – gramos, $i = 100$ cm a 200 grs, $\{100, 120, 150, 170, 190, 200\}$

Objetivo: Diseñar un clasificador basado en la teoría de la probabilidad y un clasificador bayesiano simple para resolver el problema de clasificación, comparar los resultados del ejemplo manejado en clases y este diseño. Particularmente porque ahora se presenta una clase más.

Solución: Partiendo del análisis del diseño de un Algoritmo basado en la teoría de la probabilidad

Llamaremos C_1 a la clase que representa a las manzanas, C_2 a la clase mango y C_3 a la que representa a las peras. Ahora, calculemos los siguientes datos a priori necesarios para diseñar el clasificador.

Información a priori:

- ✓ El espacio muestral $S = \{x \mid x \text{ es una fruta}\}$ y $|S| = 12,000$
- ✓ Llamaremos C_1 al evento $\{x \mid x \text{ es un manzana}\}$ y $|C_1| = 5,000$
- ✓ Llamaremos C_2 al evento $\{x \mid x \text{ es un mango}\}$ y $|C_2| = 4,000$
- ✓ Llamaremos C_3 al evento $\{x \mid x \text{ es una pera}\}$ y $|C_3| = 2,000$
- ✓ Si suponemos que $|S|$ es un espacio equiprobable,

$$P(C_1) = \frac{|C_1|}{|S|} ; \quad P(C_2) = \frac{|C_2|}{|S|} \quad P(C_3) = \frac{|C_3|}{|S|}$$

- ✓ Al momento de clasificar un patrón debemos considerar los rasgos que son fueron seleccionados y probar el clasificador para determinar la eficiencia y eficacia de este con respecto a su clasificación utilizando ese rasgo. En este sentido, Bayes nos propone construir el clasificador considerando la distribución de las muestras conforme los rasgos, en este caso tendríamos que considerar los cinco tipos de i gramos distribuidos, {100, 120, 150, 170, 190, 200} que fueron encontrados en el muestreo.

Resolver ahora diseñando el clasificador basado en el teorema de Bayes

Algoritmo

- 1.- Obtener una muestra representativa S de los objetos a clasificar
- 2.- Determinar cada una de las clases C_k que formarán parte del sistema
- 3.- Determinar, con base en la muestra y en la cardinalidad de cada clase, las probabilidades $P(C_k)$
4. Determinar los rasgos útiles que se van a utilizar para clasificar, y elaborar cada distribución de probabilidad $p(C_i|X)$ la cual va a ser dependiente del número y naturaleza de cada rasgo de la variable aleatoria vectorial X
5. Para clasificar un patrón desconocido de entrada X, aplicar la regla de aprendizaje:

$$X \in C_i, \text{ si } d_i > d_j \forall i \neq j, \text{ con } d_k = \ln(P(C_k)) + \ln(p(X|C_k))$$