Reconocimiento Estadístico de Formas

Teoría de la Decisión de Bayes.

Estimación del error

J. Salvador Sánchez

Base Teórica de Partida

- Suponemos conocidas todas las características estadísticas del problema. En particular, para cada clase, se conocen:
  - las probabilidades a priori.
  - las funciones densidad de probabilidad o probabilidad condicional.

Problema de las Frutas I

- Llamaremos ω al estado de naturaleza de cada objeto: (ω_1 : naranja, ω_2 : fresa).
- Llamaremos $P(\omega_i)$ a la probabilidad de que, al coger al azar una fruta, sea de tipo ω_i (probabilidad a priori).
- Problema:
  ¿Qué regla de decisión utilizar si no disponemos de ninguna información adicional?
- Solución:
  Diremos ω_1 si $P(\omega_1) > P(\omega_2)$, si no ω_2.

Uso de Información Adicional

- Supongamos que sabemos que la fresa tiene, en general, un perímetro más pequeño que la naranja.
- Llamaremos $p(x|\omega_i)$ a la densidad de probabilidad de que un objeto de la clase ω_i tenga un perímetro x.

Problema de las Frutas II

- Supongamos que hemos medido el perímetro (x) de una fruta, ¿cómo la clasificamos?
- Solución:

  Diremos que es ω_1 si la probabilidad de que sea ω_1 condicionada a haber medido un perímetro x, $P(\omega_1| x)$, es mayor que la probabilidad de que sea ω_2 condicionada a haber medido un perímetro x:

$$P(\omega_1| x) > P(\omega_2| x)$$

Teorema de Bayes

- Para calcular esta probabilidades condicionales (es decir, las probabilidades a posteriori), usaremos la *regla de Bayes*:

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{p(x)}$$

donde

$$p(x) = \sum_{j=1}^{c} p(x \mid \omega_j)P(\omega_j)$$

Regla de Decisión de Bayes

- A partir del Teorema de Bayes, se puede definir la ***Regla de Decisión de Bayes*** (δ^*) como sigue:

$$\delta^* \equiv \delta(x) = \omega_i \iff P(\omega_i | x) > P(\omega_j | x)$$

$$\nabla\, j \neq i,\ i, j = 1, \ldots, c$$

Función de Pérdida

- Un clasificador δ se considerará bueno si proporciona una estimación de clase (ω_j) que se aproxime a su verdadero valor (ω_i).

- Definamos $L(\omega_i, \omega_j)$ como la *pérdida* o *coste* de asignar una muestra x a la clase ω_j cuando su verdadera clase es ω_i ($i \neq j$, $i, j = 1, \ldots, c$).

Función de Pérdida

- Entonces, el *coste medio condicional a posteriori* o *riesgo condicional* para cada clase ω_i será:

$$r_i(x) = \sum_{j=1, j\neq i}^{c} L(\omega_i, \omega_j) P(\omega_i \mid x)$$

- Por tanto, dada una regla de clasificación, δ, el *riesgo condicional* se define como:

$$r_\delta(x) = \sum_{i=1}^{c} L(\omega_i, \delta(x)) P(\omega_i \mid x)$$

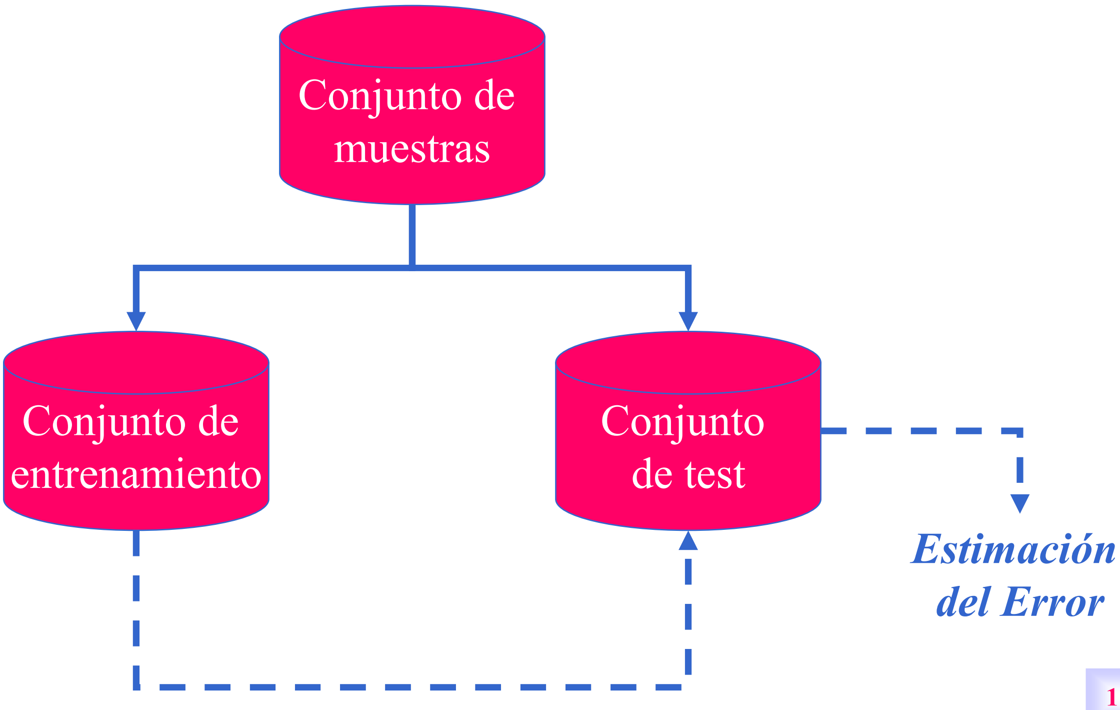
Función de Pérdida

- El *riesgo medio* se define como el valor esperado del riesgo condicional:

$$R_\delta = E[r_\delta(x)]$$

- La regla de Bayes (δ^*) será aquella que minimice el riesgo condicional de cada muestra a clasificar.

Estimación del Error



Estimación del Error

- La tasa de error será el cociente entre el número de muestras mal clasificadas y el número total de muestras disponibles.
- Se utiliza para conocer la "bondad" de un clasificador y para poder compararlo con otros diseños.
- **<u>Nunca</u>** deberíamos utilizar el mismo conjunto de muestras para diseño y test: independencia estadística.
- En general, se proporciona también la desviación típica (o la varianza) o un intervalo de confianza (95%).

Estimación del Error

- Problema:
  Normalmente, se tiene un conjunto reducido de muestras para hacer el entrenamiento y el test, ¿cuántas utilizo para cada función?
  - Si utilizo muchas para test, dispondré de pocas para entrenamiento y, por tanto, el clasificador no estará bien diseñado.
  - Si utilizo muchas para diseño, tendré pocas para test y, en consecuencia, la tasa de error no será significativa.

Estimación del Error

- *Resustitución*:
  - Diseño y test del clasificador con la totalidad de las muestras originales.
  - Dependencia estadística: dará lugar a estimaciones demasiado optimistas.
  - Las estimaciones obtenidas con este método suponen una cota inferior de las probabilidades de error.

Estimación del Error

- *Partición (Holdout)*:
  - El conjunto original se divide en dos subconjuntos mutuamente exclusivos: uno de entrenamiento y uno de test.
  - Independencia estadística.
  - Sin embargo, un clasificador diseñado a partir de la totalidad de las muestras tendrá, en general, un mejor comportamiento.
  - Dependiente de la partición realizada.
  - Estimaciones relativamente pesimistas.

Estimación del Error

- *Validación Cruzada (Cross Validation)*:
  - Se realizan diversas particiones aleatorias del conjunto original. Aplicación sucesiva del método de partición, intercambiando las funciones de diseño y test.
  - La estimación del error será la media de la estimación calculada sobre cada una de las particiones.
  - Suele calcularse también la desviación típica de la estimación media del error: indicativo de la robustez del clasificador.
  - Menor dependencia de la partición.
  - Estimaciones no tan pesimistas como en "*holdout*".

Estimación del Error

- *Leaving-one-out*:
  - Si disponemos de n muestras, se realizarán n particiones, dejando en cada una de ellas 1 única muestra para test y utilizando las restantes $n - 1$ para entrenamiento.
  - La estimación del error corresponderá a la media de las n estimaciones.
  - Supone una cota superior de las probabilidades del error, al igual que con el método "*holdout*".

Estimación del Error

- *Rotación o m-Validación Cruzada (Leaving-k-out)*:
  - Supone un compromiso entre los métodos "*holdout*" y "*leaving-one-out*".
  - Se realizarán $m = n / k$ particiones distintas, dejando en cada una de ellas k muestras para test y utilizando las restantes $n - k$ para diseñar el clasificador.
  - La estimación del error corresponderá a la media de las m estimaciones.
  - Si $k = 1$, se reduce al método "*leaving-one-out*". Si $m = n / 2$, esencialmente corresponde al "*holdout*", intercambiando las funciones de test y entrenamiento.

Ejemplo de la Regla de Bayes

- La meningitis causa rigidez de cuello en un 50% de los casos:

$$P(S \mid M) = 0'5$$

- Se conoce también la probabilidad a priori de que un paciente tenga meningitis:

$$P(M) = 1 / 50000$$

- Se conoce la probabilidad a priori de que un paciente tenga rigidez de cuello:

$$P(S) = 1 / 20$$

Ejemplo de la Regla de Bayes

- Por tanto, podemos calcular:

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = 0'0002$$

- Obsérvese que, a pesar de que la meningitis causa (con alta probabilidad 0’5) dolor de cuello, la probabilidad de meningitis en el paciente sigue siendo pequeña, debido a que la probabilidad a priori de la rigidez de cuello es mucho mayor que la de la meningitis

Ejemplo de la Regla de Bayes

- El médico podría saber que 1 de cada 5000 pacientes con rigidez de cuello tienen meningitis y, por tanto, no tendría que usar la regla de *Bayes*:
  - supongamos que hay una epidemia de meningitis y la probabilidad a priori $P(M)$ sube.
  - el médico que computa la probabilidad de meningitis condicionada a rigidez de cuello basándose en la observación estadística de los pacientes antes de la epidemia no sabe como actualizar el valor.
  - el médico que usa la regla de *Bayes* sabrá que $P(M \mid S)$ aumenta proporcionalmente a $P(M)$, y que $P(S \mid M)$ no se ve afectada por la epidemia.