

Sistemas Inteligentes

Escuela Técnica Superior de Informática

Universitat Politècnica de València

Introducción a la estimación del error en Reconocimiento de Formas

Evaluación de un clasificador

¿Cómo utilizo mis muestras etiquetadas para entrenar y evaluar un clasificador?

- Entrenamos un clasificador con $\underline{x_1}, \underline{x_2}, \underline{x_4}, \underline{x_5}, \underline{x_7}, \underline{x_8}$
 - Ejemplo: utilizando el algoritmo Perceptrón obtengo:

Iris Setosa	
$x_1 = (1.4 \ 0.2)^t$	
$x_2 = (1.3 \ 0.2)^t$	
$x_3 = (1.5 \ 0.2)^t$	←
Iris Versicolor	
$x_4 = (4.7 \ 1.4)^t$	
$x_5 = (4.5 \ 1.5)^t$	
$x_6 = (4.9 \ 1.5)^t$	←
Iris Virgínica	
$x_7 = (6.0 \ 2.5)^t$	
$x_8 = (5.1 \ 1.9)^t$	
$x_9 = (5.9 \ 2.1)^t$	←

$$\begin{array}{lcl}
 \text{Setosa} \rightarrow & w_1 = & (0.8 \quad -1.8 \quad -1)^t \\
 \text{Versicolor} \rightarrow & w_2 = & (0.6 \quad -1.7 \quad -1)^t \\
 \text{Virgínica} \rightarrow & w_3 = & (-2.5 \quad -1.5 \quad 0.6)^t
 \end{array}$$

- Evaluamos con $\boxed{x_3, x_6, x_9}$

- Ejemplo: un clasificador con los pesos del Perceptrón

$$\begin{array}{lcl}
 \text{Setosa} \rightarrow & c(x_3) = \text{argmax}(w_1^t x_3, w_2^t x_3, w_3^t x_3) = \textcircled{1} \rightarrow \text{Acierto} \\
 \text{Versicolor} \rightarrow & c(x_6) = \text{argmax}(w_1^t x_6, w_2^t x_6, w_3^t x_6) = \textcircled{3} \rightarrow \text{Error} \\
 & & \text{Virgínica} \rightarrow \text{Error} \\
 & & \text{Error} = \frac{1}{3} \\
 & c(x_9) = \text{argmax}(w_1^t x_9, w_2^t x_9, w_3^t x_9) = \underline{3} \rightarrow \text{Acierto}
 \end{array}$$

Diseño de experimentos

- Necesitamos dedicar muestras a entrenamiento y muestras a evaluación (test)
- **Partición**: Si los conjuntos de entrenamiento y test son disjuntos
 - Inconveniente: las muestras de test no se pueden usar para entrenar

Diseño de experimentos

- Necesitamos dedicar muestras a entrenamiento y muestras a evaluación (test)
- **Partición**: Si los conjuntos de entrenamiento y test son disjuntos
 - Inconveniente: las muestras de test no se pueden usar para entrenar
- **Resustitución**: Uso las mismas muestras para entrenar y para evaluar
 - Inconveniente: es muy optimista

Diseño de experimentos

- Necesitamos dedicar muestras a entrenamiento y muestras a evaluación (test)
- **Partición**: Si los conjuntos de entrenamiento y test son disjuntos
 - Inconveniente: las muestras de test no se pueden usar para entrenar
- **Resustitución**: Uso las mismas muestras para entrenar y para evaluar
 - Inconveniente: es muy optimista
- ¿Cómo aprovechar mejor mis muestras?
- **Validación Cruzada en B bloques (B -fold Cross Validation)**: Se definen B bloques. Iterativamente, un bloque para test y el resto para entrenamiento.
 - Inconvenientes: Reduce el número de datos de entrenamiento (sobre todo cuando B es pequeño) y el coste computacional se incrementa con B .
- **Exclusión individual (Leaving One Out)**: Iterativamente, una muestra para test y el resto para entrenamiento. Equivale a Validación Cruzada en N bloques.
 - Inconvenientes: el coste computacional.

Ejemplo: diseño de experimentos

- Validación cruzada en 3 bloques:

$$\hat{p}_e = \frac{n_{\text{err}B1} + n_{\text{err}B2} + n_{\text{err}B3}}{N}$$

	Test	Entrenamiento	
$B_1 = \{x_1, x_4, x_7\}$	B_1	B_2, B_3	$\leftarrow i=1$
$B_2 = \{x_2, x_5, x_8\}$	B_2	B_1, B_3	$\leftarrow i=2$
$B_3 = \{x_3, x_6, x_9\}$	B_3	B_1, B_2	$\leftarrow i=3$

$N=9$

Iris Setosa

$$x_1 = (1.4 \ 0.2)^t$$

$$x_2 = (1.3 \ 0.2)^t$$

$$x_3 = (1.5 \ 0.2)^t$$

Iris Versicolor

$$x_4 = (4.7 \ 1.4)^t$$

$$x_5 = (4.5 \ 1.5)^t$$

$$x_6 = (4.9 \ 1.5)^t$$

Iris Virgínica

$$x_7 = (6.0 \ 2.5)^t$$

$$x_8 = (5.1 \ 1.9)^t$$

$$x_9 = (5.9 \ 2.1)^t$$

3 ejecuciones del algoritmo de entrenamiento y evaluación

$$B=2 \quad B_1 = \{x_1, x_4, x_7, x_2, x_5\} \quad B_2 = \{x_8, x_3, x_6, x_9\}$$

- Exclusión individual (Leaving One Out):

Test	Entrenamiento	
x_1	$x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$	$\leftarrow i=1$
x_2	$x_1, x_3, x_4, x_5, x_6, x_7, x_8, x_9$	$\leftarrow i=2$
x_3	$x_1, x_2, x_4, x_5, x_6, x_7, x_8, x_9$	
...	...	
x_9	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	$\leftarrow i=9$

9 ejecuciones del algoritmo de entrenamiento y evaluación

Estimación empírica de la probabilidad de error

- La estimación empírica de la probabilidad de error es:

$$\hat{p}_{\text{error}} = \frac{N_e}{N} \quad \leftarrow \begin{array}{l} N^{\circ} \text{ errores} \\ N^{\circ} \text{ de muestras de test} \end{array}$$

siendo N_e el número de errores al clasificar N muestras.

- ¿Cuál sería la estimación de la verdadera probabilidad de error p_{error} ?

- Si N es muy grande, podemos asumir que $\hat{p}_{\text{error}} \sim \mathcal{N}\left(p_{\text{error}}, \frac{p_{\text{error}}(1 - p_{\text{error}})}{N}\right)$

- Intervalo de confianza al 95 %:

$$p(\hat{p}_{\text{error}} - \epsilon \leq p_{\text{error}} \leq \hat{p}_{\text{error}} + \epsilon) = 0.95;$$

$$\epsilon = 1.96 \sqrt{\frac{\hat{p}_{\text{error}}(1 - \hat{p}_{\text{error}})}{N}}$$

- Ejemplo:* Tenemos 50 errores en 1000 muestras de test. Con una confianza del 95 % podemos afirmar que p_{error} es:

$$p_{\text{error}} = \underbrace{0.05}_{\hat{p}_{\text{error}}} \pm \underbrace{1.96 \sqrt{\frac{0.05 \cdot 0.95}{1000}}}_{\epsilon} = 0.05 \pm 0.014 \quad (5\% \pm 1.4\%)$$

$\hat{p}_{\text{error}} = \frac{50}{1000} = 5\%$

$[3.6\% - 6.4\%]$

- Ejemplo:* Si hay 5 errores en 100 muestras de test p_{error} es:

$$p_{\text{error}} = \dots = 0.05 \pm \underline{\underline{0.043}} \quad (5\% \pm 4.3\%)$$

$$\hat{p}_{\text{error}} = \frac{5}{100} = 0.05 \quad 5\%$$