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?

Iris Setosa

$\boldsymbol{x}_1$=(1.4 0.2)^t
$\boldsymbol{x}_2$=(1.3 0.2)^t
$\boldsymbol{x}_3$=(1.5 0.2)^t

Iris Versicolor

$\boldsymbol{x}_4$=(4.7 1.4)^t
$\boldsymbol{x}_5$=(4.5 1.5)^t
$\boldsymbol{x}_6$=(4.9 1.5)^t

Iris Virgínica

$\boldsymbol{x}_7$=(6.0 2.5)^t
$\boldsymbol{x}_8$=(5.1 1.9)^t
$\boldsymbol{x}_9$=(5.9 2.1)^t

- Entrenamos un clasificador con $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_7, \boldsymbol{x}_8$
  - Ejemplo: utilizando el algoritmo Perceptrón obtengo:

$$\begin{aligned} \mathbf{w}_1 &= (0.8 \quad -1.8 \quad -1)^t \\ \mathbf{w}_2 &= (0.6 \quad -1.7 \quad -1)^t \\ \mathbf{w}_3 &= (-2.5 \quad -1.5 \quad 0.6)^t \end{aligned}$$

- Evaluamos con $\boldsymbol{x}_3, \boldsymbol{x}_6, \boldsymbol{x}_9$
  - Ejemplo: un clasificador con los pesos del Perceptron

$$\begin{aligned} c(\mathbf{x}_3) &= \text{argmax}(\mathbf{w}_1{}^t\mathbf{x}_3, \mathbf{w}_2{}^t\mathbf{x}_3, \mathbf{w}_3{}^t\mathbf{x}_3) = 1 \rightarrow \text{Acierto} \\ c(\mathbf{x}_6) &= \text{argmax}(\mathbf{w}_1{}^t\mathbf{x}_6, \mathbf{w}_2{}^t\mathbf{x}_6, \mathbf{w}_3{}^t\mathbf{x}_6) = 3 \rightarrow \text{Error} \\ c(\mathbf{x}_9) &= \text{argmax}(\mathbf{w}_1{}^t\mathbf{x}_9, \mathbf{w}_2{}^t\mathbf{x}_9, \mathbf{w}_3{}^t\mathbf{x}_9) = 3 \rightarrow \text{Acierto} \end{aligned}$$

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

Iris Setosa

$\boldsymbol{x}_1$=(1.4 0.2)^t
$\boldsymbol{x}_2$=(1.3 0.2)^t
$\boldsymbol{x}_3$=(1.5 0.2)^t

Iris Versicolor

$\boldsymbol{x}_4$=(4.7 1.4)^t
$\boldsymbol{x}_5$=(4.5 1.5)^t
$\boldsymbol{x}_6$=(4.9 1.5)^t

Iris Virgínica

$\boldsymbol{x}_7$=(6.0 2.5)^t
$\boldsymbol{x}_8$=(5.1 1.9)^t
$\boldsymbol{x}_9$=(5.9 2.1)^t

- Validación cruzada en 3 bloques:

$$B_1 = \{\boldsymbol{x}_1, \boldsymbol{x}_4, \boldsymbol{x}_7\}$$
$$B_2 = \{\boldsymbol{x}_2, \boldsymbol{x}_5, \boldsymbol{x}_8\}$$
$$B_3 = \{\boldsymbol{x}_3, \boldsymbol{x}_6, \boldsymbol{x}_9\}$$

Test	Entrenamiento
B_1	B_2, B_3
B_2	B_1, B_3
B_3	B_1, B_2

3 ejecuciones del algoritmo de entrenamiento y evaluación

- Exclusión individual (Leaving One Out):

Test	Entrenamiento
$\boldsymbol{x}_1$	$\boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_8, \boldsymbol{x}_9$
$\boldsymbol{x}_2$	$\boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_8, \boldsymbol{x}_9$
$\boldsymbol{x}_3$	$\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_8, \boldsymbol{x}_9$
$\cdots$	$\cdots$
$\boldsymbol{x}_9$	$\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_8$

9 ejecuciones del algoritmo de entrenamiento y evaluación

Estimación empírica de la probabilidad de error

- La estimación empírica de la probabilidad de error es:

$$\hat{p}_{\text{error}} = \frac{N_e}{N}$$

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 $95\,\%$:

$$p(\hat{p}_{\text{error}} - \epsilon \leq p_{\text{error}} \leq \hat{p}_{\text{error}} + \epsilon) \;=\; 0.95; \qquad \epsilon \;=\; 1.96\sqrt{\frac{\hat{p}_{\text{error}}(1 - \hat{p}_{\text{error}})}{N}}$$

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

$$p_{\text{error}} \;=\; 0.05 \pm 1.96\sqrt{\frac{0.05 \cdot 0.95}{1000}} \;=\; 0.05 \pm 0.014 \;\; (5\,\% \pm 1.4\,\%)$$

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

$$p_{\text{error}} \;=\; \cdots \;=\; 0.05 \pm 0.043 \;\; (5\,\% \pm 4.3\,\%)$$