

Metody oceny jakości klasyfikacji

Dr inż. Urszula Libal

Metody oceny jakości klasyfikacji

SPIS TREŚCI:

- ❑ Macierz pomyłek
- ❑ Miary oceny jakości klasyfikacji
- ❑ Krzywa ROC
- ❑ Krosvalidacja

1. Zbiór uczący i zbiór testowy

1. Zbiór uczący

- służy do konstrukcji (treningu) klasyfikatora w procesie uczenia

2. Zbiór testowy

- służy do weryfikacji jakości klasyfikatora

Podział zbioru danych na zbiór uczący i testowy zależy od liczności tego zbioru.

Zbiory uczący i testowy są rozłączne (zawierają inne obrazy).

Częsty podział dla dużych zbiorów danych to:

- $2/3$ (zbiór uczący), $1/3$ (zbiór testowy) lub
- 75% (zbiór uczący), 25% (zbiór testowy).

2. Metody oceny jakości klasyfikatora

Ocena klasyfikacji na podstawie zbioru testowego:

1. **Macierz pomyłek** (*confusion matrix*)

2. **Miary oceny jakości klasyfikacji:**

- ryzyko,
- błąd klasyfikacji,
- trafność klasyfikacji,
- współczynniki TP, TN, FP, FN, ...

3. **Krzywa ROC**

4. **Kroswalidacja** (*cross-validation*)

3. Macierz pomyłek

Przypadek klasyfikacji do wielu klas $\mathcal{M} = \{1, 2, \dots, m\}$:

	klasa wskazana przez klasyfikator			
klasa pochodzenia obrazu	C_1	C_2	\dots	C_m
C_1	r_{11}	r_{12}	\dots	r_{1m}
C_2	r_{21}	r_{22}	\dots	r_{2m}
\vdots	\vdots	\vdots	\ddots	\vdots
C_m	r_{m1}	r_{m2}	\dots	r_{mm}

r_{ij} - liczba obrazów testowych z klasy C_i , przypisana do klasy C_j ,

N_i - liczność obrazów z klasy C_i (zbiór testowy)

Liczność zbioru testowego:

$$\#test = \sum_{i=1}^m N_i$$

Łączna liczba **poprawnie** zaklasyfikowanych obrazów testowych:

$$\#correct = \sum_{i=1}^m r_{ii}$$

Łączna liczba **błędnie** zaklasyfikowanych obrazów testowych:

$$\#error = \#test - \#correct$$

Przypadek klasyfikatora binarnego:

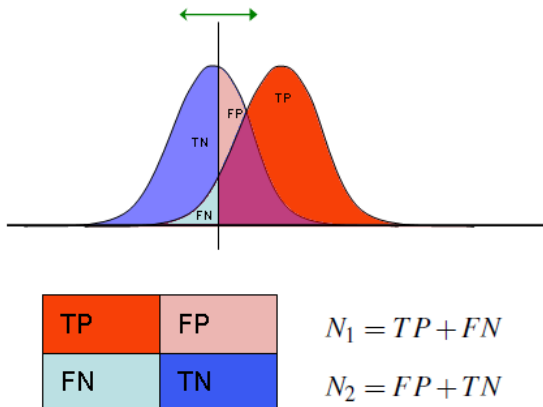
klasa pochodzenia obrazu	klasa wskazana przez klasyfikator	
	$C_1 (+)$	$C_2 (-)$
$C_1 (+)$	r_{11} (TP)	r_{12} (FN)
$C_2 (-)$	r_{21} (FP)	r_{22} (TN)

TP (*true positive*) - liczba poprawnie zaklasyfikowanych obrazów z klasy C_1

FP (*false positive*) - liczba błędnie zaklasyfikowanych obrazów z klasy C_2 do klasy C_1

FN (*false negative*) - liczba błędnie zaklasyfikowanych obrazów z klasy C_1 do klasy C_2

TN (*true negative*) - liczba poprawnie zaklasyfikowanych obrazów z klasy C_2



Rysunek 1. Współczynniki TP, FP, FN i TN dla klasyfikacji binarnej.

4. Miary oceny jakości klasyfikacji

Miary uniwersalne (dla dowolnej liczby klas):

— **trafność** (*accuracy*)

$$\frac{\#correct}{\#test}$$

— **błąd klasyfikacji** (*error rate*)

$$\frac{\#error}{\#test} = 1 - \frac{\#correct}{\#test}$$

Miary w przypadku klasyfikacji binarnej:

— **trafność**

$$\frac{TP + TN}{N_1 + N_2} = \frac{TP + TN}{TP + FN + FP + TN}$$

— **błąd klasyfikacji**

$$\frac{FP + FN}{N_1 + N_2} = \frac{FP + FN}{TP + FN + FP + TN}$$

— **Ryzyko klasyfikatora szacujemy za pomocą błędu klasyfikacji** (patrz rys. 1)

— **współczynnik TP** (*TP rate*, czułość)

$$TPR = \frac{TP}{N_1} = \frac{TP}{TP + FN}$$

— **współczynnik TN** (*TN rate*, specyficzność)

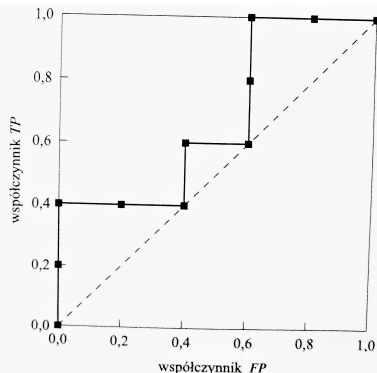
$$TNR = \frac{TN}{N_2} = \frac{TN}{FP + TN}$$

— **współczynnik FP** (*FP rate*)

$$FPR = \frac{FP}{N_2} = \frac{FP}{FP + TN}$$

5. Krzywa ROC

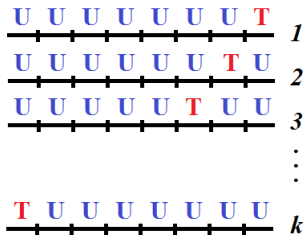
nr_rekordu	klasa	próg	TP	FP
		1,0	0	0
1	+	0,9	0,2	0
2	+	0,8	0,4	0
3	-	0,75	0,4	0,2
4	-	0,7	0,4	0,4
5	+	0,65	0,6	0,4
6	-	0,6	0,6	0,6
7	+	0,5	0,8	0,6
8	+	0,4	1,0	0,6
9	-	0,3	1,0	0,8
10	-	0,2	1,0	1,0



Rysunek 2. Krzywa ROC.

6. K-krotna krosvalidacja

Dzielimy dane na k możliwie równych, wzajemnie rozłącznych, części. Do uczenia wykorzystujemy $k - 1$ części, do testowania pozostałą jedną część. Procedurę powtarzamy k razy, za każdym razem zmieniając zbiór testowy na kolejną niewykorzystaną dotychczas część.



Rysunek 3. K-krotna krosvalidacja: U-uczenie, T-testowanie.

Sumaryczna liczba poprawnych klasyfikacji podzielona przez licznosc zbioru danych N stanowi **oszacowanie trafności klasyfikacji**.

— Szczególny przypadek:

N-krotna krosvalidacja (*leave-one-out cross-validation*),

gdzie N to licznosc zbioru danych przed podziałem na zbiór uczący i testowy.