



Wykład 2: Metody oceny klasyfikatorów

Hung Son Nguyen

Institute of Mathematics, Warsaw University
son@mimuw.edu.pl

08/03/2016



1 Skuteczność predykcji



- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen



- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen
- 3 Metody walidacji danych



- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen
- 3 Metody walidacji danych
- 4 Krzywy Lift i ROC



- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen
- 3 Metody walidacji danych
- 4 Krzywy Lift i ROC



Błąd klasyfikacji

$$\text{Błąd klasyfikacji} = \frac{\text{liczba błędów}}{\text{liczba obiektów testowych}}$$

gdzie:



Błąd klasyfikacji

$$\text{Błąd klasyfikacji} = \frac{\text{liczba błędów}}{\text{liczba obiektów testowych}}$$

gdzie:

- *Sukces*: gdy obiekt jest prawidłowo klasyfikowany



Błąd klasyfikacji

$$\text{Błąd klasyfikacji} = \frac{\text{liczba błędów}}{\text{liczba obiektów testowych}}$$

gdzie:

- *Sukces*: gdy obiekt jest prawidłowo klasyfikowany
- *Błąd*: gdy obiekt jest źle klasyfikowany



Błąd klasyfikacji

$$\text{Błąd klasyfikacji} = \frac{\text{liczba błędów}}{\text{liczba obiektów testowych}}$$

gdzie:

- *Sukces*: gdy obiekt jest prawidłowo klasyfikowany
- *Błąd*: gdy obiekt jest źle klasyfikowany
- *Błąd klasyfikacji* (lub odsetka błędów podczas klasyfikacji) powinien być wyznaczony na losowych i nieznanymi danych.

Macierz błędu (ang. confusion matrix)

- Są dwa rodzaje błędu:



		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

Macierz błędu (ang. confusion matrix)

- Są dwa rodzaje błędu:



		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

- W “systemach uczących się”: minimalizujemy FP+FN lub miarę skuteczności (ang. Accuracy) (ACC):

$$ACC = (TP + TN) / (TP + FP + TN + FN)$$

Macierz błędu (ang. confusion matrix)

- Są dwa rodzaje błędu:



		Actual Value (as confirmed by experiment)	
		positives	negatives
Predicted Value (predicted by the test)	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

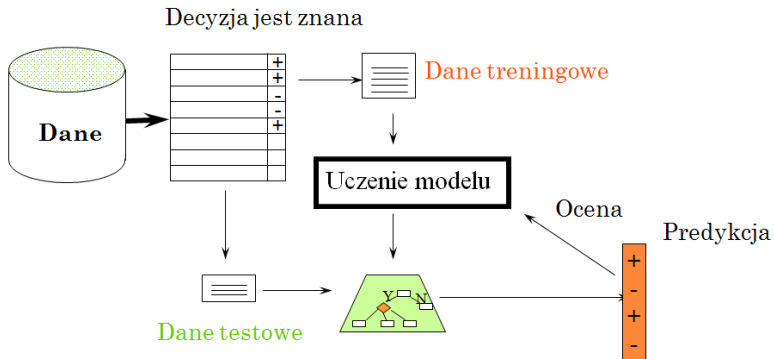
- W “systemach uczących się”: minimalizujemy FP+FN lub miarę skuteczności (ang. Accuracy) (ACC):

$$ACC = (TP + TN) / (TP + FP + TN + FN)$$

- W marketingu: maksymalizujemy TP.

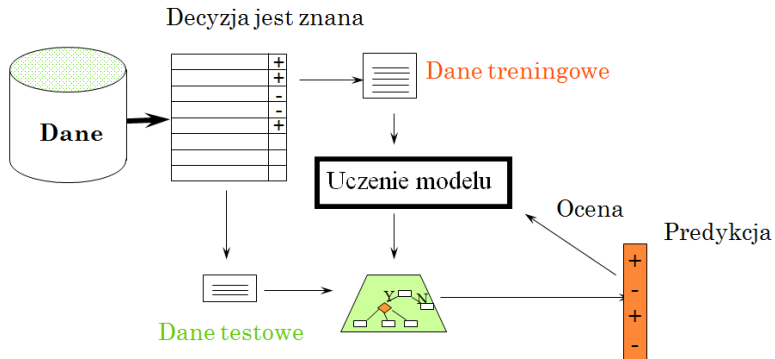
Proces klasyfikacji

- Podział zbioru danych na część treningową i testową;



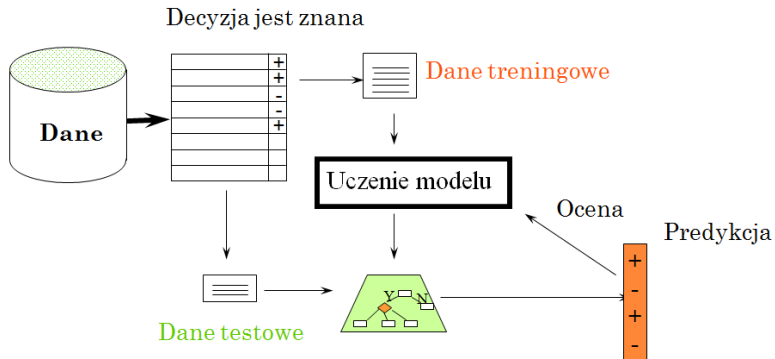
Proces klasyfikacji

- Podział zbioru danych na część treningową i testową;
- Uczenie lub poszukiwanie modelu



Proces klasyfikacji

- Podział zbioru danych na część treningową i testową;
- Uczenie lub poszukiwanie modelu
- Ocena klasyfikatora



- Niektóre metody uczenia działają w dwóch etapach:



Problem strojenia parametrów

- Niektóre metody uczenia działają w dwóch etapach:
- Etap 1: Buduje strukturę



- Niektóre metody uczenia działają w dwóch etapach:
- Etap 1: Buduje strukturę
- Etap 2: Optymalizuje parametry



- Niektóre metody uczenia działają w dwóch etapach:
- Etap 1: Buduje strukturę
- Etap 2: Optymalizuje parametry

Uwaga:

Nie używaj danych testowych do budowy klasyfikatorów!





- Niektóre metody uczenia działają w dwóch etapach:
- Etap 1: Buduje strukturę
- Etap 2: Optymalizuje parametry

Uwaga:

Nie używaj danych testowych do budowy klasyfikatorów!

- Właściwa procedura powinna zawierać 3 zbiory: treningowe, walidacyjne i testowe



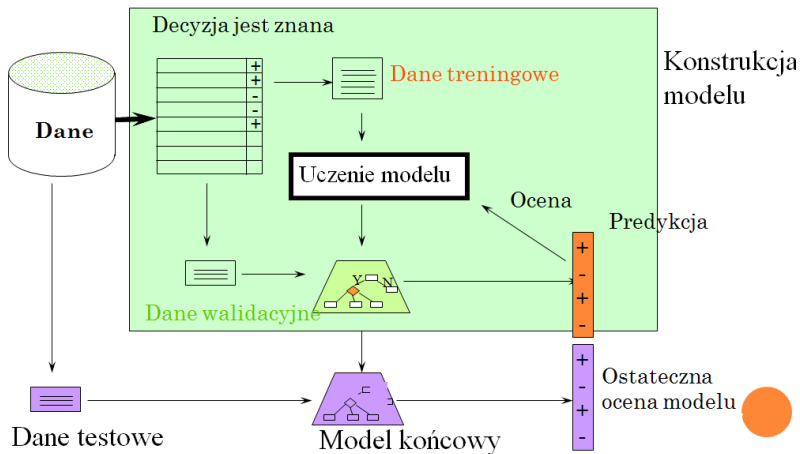
- Niektóre metody uczenia działają w dwóch etapach:
- Etap 1: Buduje strukturę
- Etap 2: Optymalizuje parametry

Uwaga:

Nie używaj danych testowych do budowy klasyfikatorów!

- Właściwa procedura powinna zawierać 3 zbiory: treningowe, walidacyjne i testowe
- Dane walidacyjne używane są do optymalizacji parametrów

Klasyfikacja: Zbiory treningowe, walidacyjne i testowe





- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen
- 3 Metody walidacji danych
- 4 Krzywy Lift i ROC



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach
 - Estymowana skuteczność: 75%



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach
 - Estymowana skuteczność: 75%
 - Jak bliska jest ta estymacja do prawdziwej skuteczności p ?



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach
 - Estymowana skuteczność: 75%
 - Jak bliska jest ta estymacja do prawdziwej skuteczności p ?
 - Odp: z 80% pewnością możemy twierdzić, że $p \in [73.2, 76.7]$



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach
 - Estymowana skuteczność: 75%
 - Jak bliska jest ta estymacja do prawdziwej skuteczności p ?
 - Odp: z 80% pewnością możemy twierdzić, że $p \in [73.2, 76.7]$
- Inny przykład: $S=75$ i $N=100$



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach
 - Estymowana skuteczność: 75%
 - Jak bliska jest ta estymacja do prawdziwej skuteczności p ?
 - Odp: z 80% pewnością możemy twierdzić, że $p \in [73.2, 76.7]$

- Inny przykład: $S=75$ i $N=100$
 - Estymowana skuteczność: 75%;



- Przykład: $S = 750$ sukcesów w $N = 1000$ próbach
 - Estymowana skuteczność: 75%
 - Jak bliska jest ta estymacja do prawdziwej skuteczności p ?
 - Odp: z 80% pewnością możemy twierdzić, że $p \in [73.2, 76.7]$
- Inny przykład: $S=75$ i $N=100$
 - Estymowana skuteczność: 75%;
 - $p \in [69.1, 80.1]$ z 80% pewnością.

Przypomnienia ze statystyki

- Rozpatrujemy rozkład Bernoulliego: $p, p(1 - p)$



Przypomnienia ze statystyki

- Rozpatrujemy rozkład Bernoulliego: $p, p(1 - p)$
- Oczekiwany odsetek sukcesu w N próbach:
 $f = S/N$



Przypomnienia ze statystyki

- Rozpatrujemy rozkład Bernoulliego: $p, p(1 - p)$
- Oczekiwany odsetek sukcesu w N próbach:
 $f = S/N$
- Wartość oczekiwana i wariancja dla f :

$$p, p(1 - p)/N$$



Przypomnienia ze statystyki

- Rozpatrujemy rozkład Bernoulliego: $p, p(1 - p)$
- Oczekiwany odsetek sukcesu w N próbach:
 $f = S/N$
- Wartość oczekiwana i wariancja dla f :

$$p, p(1 - p)/N$$

- Dla dużych N , zm.i. f ma rozkład zbliżony do rozkładu normalnego;



Przypomnienia ze statystyki

- Rozpatrujemy rozkład Bernoulliego: $p, p(1 - p)$
- Oczekiwany odsetek sukcesu w N próbach:
 $f = S/N$
- Wartość oczekiwana i wariancja dla f :

$$p, p(1 - p)/N$$

- Dla dużych N , zm.l. f ma rozkład zbliżony do rozkładu normalnego;
- $[-z \leq X \leq z]$ nazywamy *przedziałem ufności na poziomie $c\%$* dla zm.l. X o zerowej wartości oczekiwanej wtw:

$$P[-z \leq X \leq z] = c$$



Przypomnienia ze statystyki

- Rozpatrujemy rozkład Bernoulliego: $p, p(1 - p)$
- Oczekiwany odsetek sukcesu w N próbach:
 $f = S/N$
- Wartość oczekiwana i wariancja dla f :

$$p, p(1 - p)/N$$

- Dla dużych N , zm.l. f ma rozkład zbliżony do rozkładu normalnego;
- $[z \leq X \leq z]$ nazywamy *przedziałem ufności na poziomie $c\%$* dla zm.l. X o zerowej wartości oczekiwanej wtw:

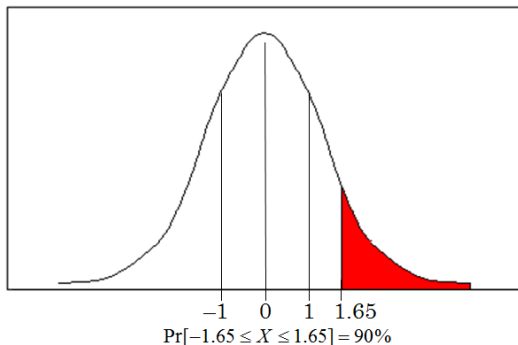
$$P[-z \leq X \leq z] = c$$

- Dla rozkładu symetrycznego mamy:

$$P[-z \leq X \leq z] = 1 - 2P[X \geq z]$$



Granice ufności



$\Pr[X \geq z]$	z
0.1%	3.09
0.5%	2.58
1%	2.33
5%	1.65
10%	1.28
20%	0.84
40%	0.25

Dla zmiennej X o rozkładzie $N(0,1)$



Granice ufności dla rozkładu Bernoulliego

- Wartość oczekiwana i wariancję dla f : $p, p(1 - p)/N$



Granice ufności dla rozkładu Bernoulliego

- Wartość oczekiwana i wariancję dla f : $p, p(1 - p)/N$
- Normalizacja zm. f :

$$\frac{f - p}{\sqrt{p(1 - p)/N}}$$



Granice ufności dla rozkładu Bernoulliego

- Wartość oczekiwana i wariancję dla f : $p, p(1-p)/N$
- Normalizacja zm. f :

$$\frac{f - p}{\sqrt{p(1-p)/N}}$$

- Mamy równanie na p :

$$Pr \left(-z \leq \frac{f - p}{\sqrt{p(1-p)/N}} \leq z \right) = c$$



Granice ufności dla rozkładu Bernoulliego

- Wartość oczekiwana i wariancję dla f : $p, p(1 - p)/N$
- Normalizacja zm. f :

$$\frac{f - p}{\sqrt{p(1 - p)/N}}$$

- Mamy równanie na p :

$$Pr \left(-z \leq \frac{f - p}{\sqrt{p(1 - p)/N}} \leq z \right) = c$$

- Rozwiązanie dla p : $p \in [p_1, p_2]$, gdzie

$$p_{1,2} = \frac{f + \frac{z^2}{2N} \pm z \sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}}$$





- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen
- 3 Metody walidacji danych
- 4 Krzywy Lift i ROC



Ogólna zasada:

- Im większy zbiór treningowy, tym lepszy jest klasyfikator



Ogólna zasada:

- Im większy zbiór treningowy, tym lepszy jest klasyfikator
- Im większy jest zbiór testowy, tym lepiej można aproksymować błąd klasyfikacji.



Ogólna zasada:

- Im większy zbiór treningowy, tym lepszy jest klasyfikator
- Im większy jest zbiór testowy, tym lepiej można aproksymować błąd klasyfikacji.

Praktyczna rada:

Kiedy proces oceniania się zakończy, wszystkie dane mogą być wykorzystywane do skonstruowania ostatecznego klasyfikatora

Walidacja krzyżowa (ang. CV = Cross-Validation)



- Walidacja krzyżowa nie pozwala na wielokrotne testowanie tego samego obiektu

Walidacja krzyżowa (ang. CV = Cross-Validation)



- Walidacja krzyżowa nie pozwala na wielokrotne testowanie tego samego obiektu
 - Krok 1: Podział zbioru danych na k równych podzbiorów

Walidacja krzyżowa (ang. CV = Cross-Validation)



- Walidacja krzyżowa nie pozwala na wielokrotne testowanie tego samego obiektu
 - Krok 1: Podział zbioru danych na k równych podzbiorów
 - Krok 2: Testowanie każdego podzbioru używając pozostałych jako zbiór treningowy

Walidacja krzyżowa (ang. CV = Cross-Validation)



- Walidacja krzyżowa nie pozwala na wielokrotne testowanie tego samego obiektu
 - Krok 1: Podział zbioru danych na k równych podzbiorów
 - Krok 2: Testowanie każdego podzbioru używając pozostałych jako zbiór treningowy
- To się nazywa k -CV = k -fold cross-validation

Walidacja krzyżowa (ang. CV = Cross-Validation)



- Walidacja krzyżowa nie pozwala na wielokrotne testowanie tego samego obiektu
 - Krok 1: Podział zbioru danych na k równych podzbiorów
 - Krok 2: Testowanie każdego podzbioru używając pozostałych jako zbiór treningowy
- To się nazywa k -CV = k -fold cross-validation
- Zwykle obiekty są przetasowane przed dokonaniem podziału.

Walidacja krzyżowa (ang. CV = Cross-Validation)



- Walidacja krzyżowa nie pozwala na wielokrotne testowanie tego samego obiektu
 - Krok 1: Podział zbioru danych na k równych podzbiorów
 - Krok 2: Testowanie każdego podzbioru używając pozostałych jako zbiór treningowy
- To się nazywa k -CV = k -fold cross-validation
- Zwykle obiekty są przetasowane przed dokonaniem podziału.
- Błędy wszystkich iteracji są uśrednione, aby otrzymać błąd globalny.

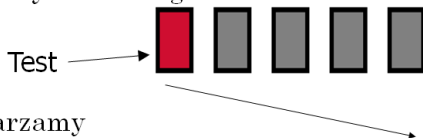
Walidacja krzyżowa



- Losowy podział zbioru danych na k grup



- Zatrzymamy jedną grupę do testu a reszty używamy do treningu



- Powtarzamy





- Standardowa metoda ocena klasyfikatorów:
10-krotna walidacja krzyżowa



- Standardowa metoda ocena klasyfikatorów:
10-krotna walidacja krzyżowa
- Liczba 10 została wyznaczona w wyniku wielu doświadczeń.



- Standardowa metoda ocena klasyfikatorów:
10-krotna walidacja krzyżowa
- Liczba 10 została wyznaczona w wyniku wielu doświadczeń.
- Walidacja pozwala na zmniejszenie długości przedziału ufności



- Standardowa metoda ocena klasyfikatorów:
10-krotna walidacja krzyżowa
- Liczba 10 została wyznaczona w wyniku wielu doświadczeń.
- Walidacja pozwala na zmniejszenie długości przedziału ufności
- Jeszcze lepsza metoda oszacowania parametrów:
Walidacja z powtórzeniami!

- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów





- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy



- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy
- Najlepiej ocenia klasyfikatora



- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy
- Najlepiej ocenia klasyfikatora
- Obliczeniowo kosztowna metoda (wyjątek: kNN)



- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy
 - Najlepiej ocenia klasyfikatora
 - Obliczeniowo kosztowna metoda (wyjątek: kNN)
- **Bootstrapping:** próbkuje ze zwracaniem, żeby stworzyć różne zbiory treningowe i testowe



- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy
 - Najlepiej ocenia klasyfikatora
 - Obliczeniowo kosztowna metoda (wyjątek: kNN)
- **Bootstrapping:** próbkuje ze zwracaniem, żeby stworzyć różne zbiory treningowe i testowe
 - Próbkuje ze zwracaniem n razy



- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy
 - Najlepiej ocenia klasyfikatora
 - Obliczeniowo kosztowna metoda (wyjątek: kNN)
- **Bootstrapping:** próbkuje ze zwracaniem, żeby stworzyć różne zbiory treningowe i testowe
 - Próbkuje ze zwracaniem n razy
 - Wybrane obiekty tworzą zbiór treningowy



- **Leave-one-out:** przypadek szczególny walidacji krzyżowej

Liczba grup = liczba przykładów

- Dla n obiektów budujemy klasyfikator n razy
 - Najlepiej ocenia klasyfikatora
 - Obliczeniowo kosztowna metoda (wyjątek: kNN)
- **Bootstrapping:** próbkuje ze zwracaniem, żeby stworzyć różne zbiory treningowe i testowe
 - Próbkuje ze zwracaniem n razy
 - Wybrane obiekty tworzą zbiór treningowy
 - Reszta – zbiór testowy.



- 1 Skuteczność predykcji
- 2 Przedział ufności miar ocen
- 3 Metody walidacji danych
- 4 Krzywy Lift i ROC

Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$
- false positive rate (FPR): $FPR = \frac{FP}{(FP+TN)} = 1 - FPC$



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$
- false positive rate (FPR): $FPR = \frac{FP}{(FP+TN)} = 1 - FPC$
- positive predictive value lub precision: $PPV = \frac{TP}{(TP+FP)}$



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$
- false positive rate (FPR): $FPR = \frac{FP}{(FP+TN)} = 1 - FPC$
- positive predictive value lub precision: $PPV = \frac{TP}{(TP+FP)}$
- negative predictive value (NPV): $NPV = TN / (TN + FN)$



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$
- false positive rate (FPR): $FPR = \frac{FP}{(FP+TN)} = 1 - FPC$
- positive predictive value lub precision: $PPV = \frac{TP}{(TP+FP)}$
- negative predictive value (NPV): $NPV = TN / (TN + FN)$
- false discovery rate (FDR): $FDR = FP / (FP + TP)$



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$
- false positive rate (FPR): $FPR = \frac{FP}{(FP+TN)} = 1 - FPC$
- positive predictive value lub precision: $PPV = \frac{TP}{(TP+FP)}$
- negative predictive value (NPV): $NPV = TN / (TN + FN)$
- false discovery rate (FDR): $FDR = FP / (FP + TP)$
- Matthew's correlation coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(FN + TN)(FP + TN)}}$$



Inne miary ocen klasyfikatorów

- Miara "sensitivity" lub "true positive rate" (TPR)

$$TPR = TP / (TP + FN)$$

czasem nazywa się też "recall" lub "hit rate".

- Specificity (SPC) lub True Negative Rate $SPC = \frac{TN}{(FP+TN)}$
- false positive rate (FPR): $FPR = \frac{FP}{(FP+TN)} = 1 - FPC$
- positive predictive value lub precision: $PPV = \frac{TP}{(TP+FP)}$
- negative predictive value (NPV): $NPV = \frac{TN}{(TN + FN)}$
- false discovery rate (FDR): $FDR = FP / (FP + TP)$
- Matthew's correlation coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(FN + TN)(FP + TN)}}$$

- F1 score: $F_1 = 2TP / [(TP + FN) + (TP + FP)]$ lub

$$\frac{1}{F_1} = \frac{1}{2} \left(\frac{1}{recall} + \frac{1}{precision} \right)$$



Lista rankingowa



Wynik (przynależność do klasy Y)
zwracany przez (niektóre) klasyfikator

No	Score	Target	CustID	Age	
1	0.97	Y	1746	...	
2	0.95	N	1024	...	
3	0.94	Y	2478	...	
4	0.93	Y	3820	...	
5	0.92	N	4897	...	
...	
99	0.11	N	2734	...	
100	0.06	N	2422		

3 trafienia wśród
pierwszych 5% listy

Jeśli klasa Y liczy
15 obiektów, to
pierwsza 5 już
obejmuje $3/15=20\%$
tej klasy

Wykres trafienia (Gain chart)



Funkcje



Funkcje

- p - parametr określający początek listy rankingowej



Funkcje

- p - parametr określający początek listy rankingowej
- CPH - (ang. Cumulative Percentage Hit)
 $CPH(p)$ = część klasy docelowej znajdująca się wśród $p\%$ pierwszych obiektów z listy rankingowej.



Funkcje

- p - parametr określający początek listy rankingowej
- CPH - (ang. Cumulative Percentage Hit)
 $CPH(p)$ = część klasy docelowej znajdująca się wśród $p\%$ pierwszych obiektów z listy rankingowej.
- zysk (ang. lift): $Lift(p) = CPH(p)/p$



Funkcje

- p - parametr określający początek listy rankingowej
- CPH - (ang. Cumulative Percentage Hit)
 $CPH(p)$ = część klasy docelowej znajdująca się wśród $p\%$ pierwszych obiektów z listy rankingowej.
- zysk (ang. lift): $Lift(p) = CPH(p)/p$
- Trafienie lub true positive rate:
 $TPR(p) = TP/(TP + FN)$



Funkcje

- p - parametr określający początek listy rankingowej
- CPH - (ang. Cumulative Percentage Hit)
 $CPH(p)$ = część klasy docelowej znajdująca się wśród $p\%$ pierwszych obiektów z listy rankingowej.
- zysk (ang. lift): $Lift(p) = CPH(p)/p$
- Trafienie lub true positive rate:
 $TPR(p) = TP/(TP + FN)$
- Odsetek fałszywych alarmów
 $FPR(p) = FP/(FP + TN)$



Wyróżnione krzywy



Wyróżnione krzywy

- Gain chart:

$Ox : p$

$Oy : CPH(p)$



Wyróżnione krzywy

- Gain chart:

$Ox : p$

$Oy : CPH(p)$

- Lift chart:

$Ox : p$

$Oy : Lift(p)$



Wyróżnione krzywy

- Gain chart:

$Ox : p$

$Oy : CPH(p)$

- Lift chart:

$Ox : p$

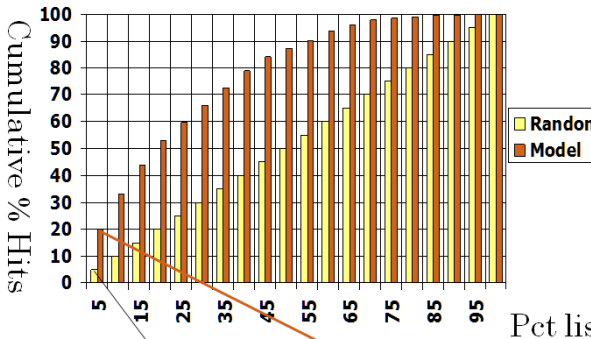
$Oy : Lift(p)$

- ROC (receiver operating characteristic):

$Ox : FPR(p)$

$Oy : TPR(p)$

CPH: LOSOWA LISTA VS LISTA RANGKINGOWA



5% losowej listy obejmuje 5% klasy docelowej,
ale 5% listy rankingowej obejmuje 20% klasy docelowej $CPH(5\%,model)=20\%$.

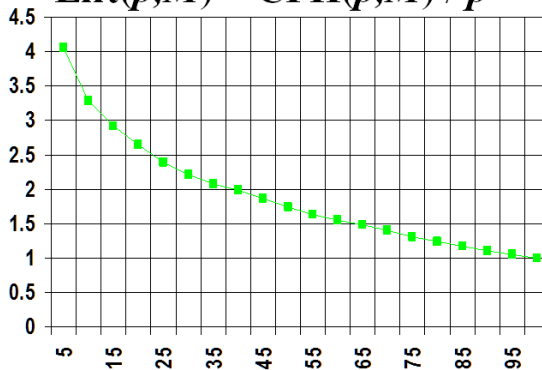
LIFT CHART (KRZYWA PODNOSZENIA)

Lift (5%)

= 20% / 5%

= 4 razy lepiej
niż losowo

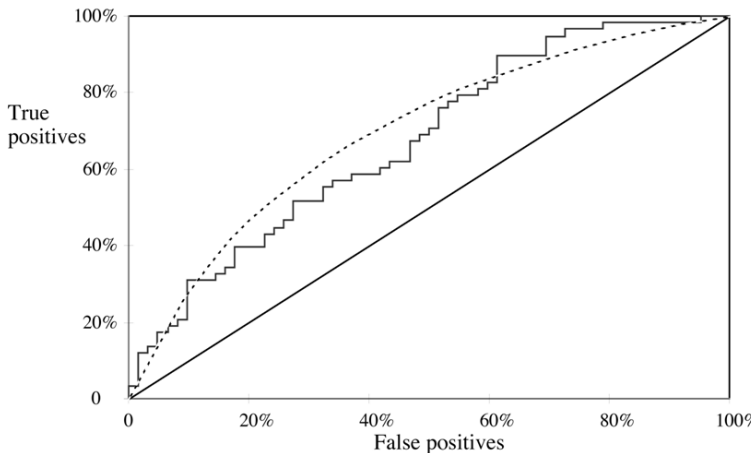
$$\text{Lift}(p, M) = \text{CPH}(p, M) / p$$



P -- percent of the list



PRZYKŁADY KRZYWYCH ROC



- Jagged curve—one set of test data
- Smooth curve—use cross-validation



- Krzywy ROC są podobne do wykresu pokrycia

Własności krzywej ROC



- Krzywy ROC są podobne do wykresu pokrycia
- Jest to skrót od: "receiver operating characteristic"

Własności krzywej ROC



- Krzywy ROC są podobne do wykresu pokrycia
- Jest to skrót od: "receiver operating characteristic"
- Pokazuje zależność między stopniem trafienia a stopniem fałszywych alarmów

Własności krzywej ROC



- Krzywy ROC są podobne do wykresu pokrycia
- Jest to skrót od: "receiver operating characteristic"
- Pokazuje zależność między stopniem trafienia a stopniem fałszywych alarmów
- Jednak różni się od krzywy pokrycia (gain chart):

Własności krzywej ROC



- Krzywy ROC są podobne do wykresu pokrycia
- Jest to skrót od: "receiver operating characteristic"
- Pokazuje zależność między stopniem trafienia a stopniem fałszywych alarmów
- Jednak różni się od krzywy pokrycia (gain chart):
 - Oś Oy : pokazuje TP rate w danej próbce zamiast pokrycia: $TPrate = TP / (TP + FN)$

Własności krzywej ROC



- Krzywy ROC są podobne do wykresu pokrycia
- Jest to skrót od: "receiver operating characteristic"
- Pokazuje zależność między stopniem trafienia a stopniem fałszywych alarmów
- Jednak różni się od krzywy pokrycia (gain chart):
 - Oś O_y : pokazuje TP rate w danej próbce zamiast pokrycia: $TPrate = TP / (TP + FN)$
 - Oś O_x : pokazuje FP rate w danej próbce zamiast wielkości próbki $FPrate = FP / (FP + TN)$