

Sztuczna Inteligencja

Analiza Bayesa

Włodzisław Duch

Katedra Informatyki Stosowanej UMK

Google: Włodzisław Duch

[Strona wykładów](#)

Martwić się czy nie?

Założmy, że w Polsce 1 na 1000 osób ma wirusa grypy.

Nowy test polegający na badaniu śliny, o dokładności 99%, wprowadzono do obowiązkowych badań okresowych.

Test wypadł pozytywnie.

Jakie jest prawdopodobieństwo, że osoba ma grypę?

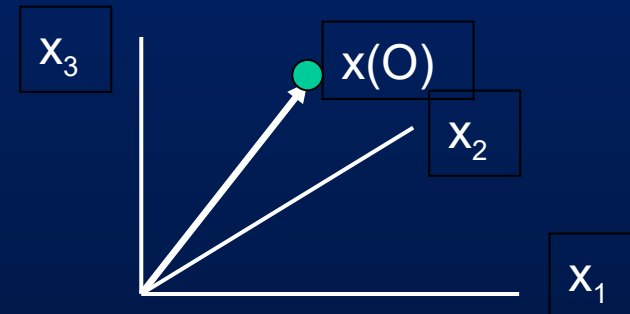
Przypomnijmy sobie podstawy rachunku prawdopodobieństwa.

Obiekty w przestrzeni cech



- Opis matematyczny reprezentuje obiekty O przy pomocy pomiarów, jakie na nich przeprowadzono, podając wartości cech $\{O_i\} \Rightarrow X(O_i)$, gdzie $X_j(O_i)$ jest wartością j -tej cechy opisującej O_i
- Atrybut i cecha są często traktowane jako synonimy, chociaż ściśle ujmując “wiek” jest atrybutem, a “młody” cechą, wartością atrybutu.
- Typy atrybutów:
 - kategoryczne: symboliczne, dyskretne – mogą mieć charakter nominalny (nieuporządkowany), np. “słodki, kwaśny, gorzki”, albo **porządkowy**, np. kolory w widmie światła, albo: mały < średni < duży (drink).
 - ciągłe: wartości numeryczne, np. wiek.

Wektor cech $X = (x_1, x_2, x_3 \dots x_d)$,
o d -składowych wskazuje na punkt
w przestrzeni cech.



Przykład: ryby



Chapter 1.2, Pattern Classification (2nd ed)
by R. O. Duda, P. E. Hart and D. G. Stork, Wiley 2000

Automatyzacja sortowania dwóch gatunków ryb, łososa i suma morskiego, które przesuwają się na pasie sortownika.

Czujniki oceniają różne cechy: długość, jasność, szerokość, liczbę płetw

Patrzemy na histogramy.

- Wybieramy liczbę przedziałów, np. $n=20$ (dyskretne dane)
- obliczamy szerokość przedziału $\Delta = (x_{max} - x_{min})/n$,
- obliczamy $N(C, r_i) = \# \text{sztuk } C \in \{\text{łosoś, sum}\}$ w każdym przedziale

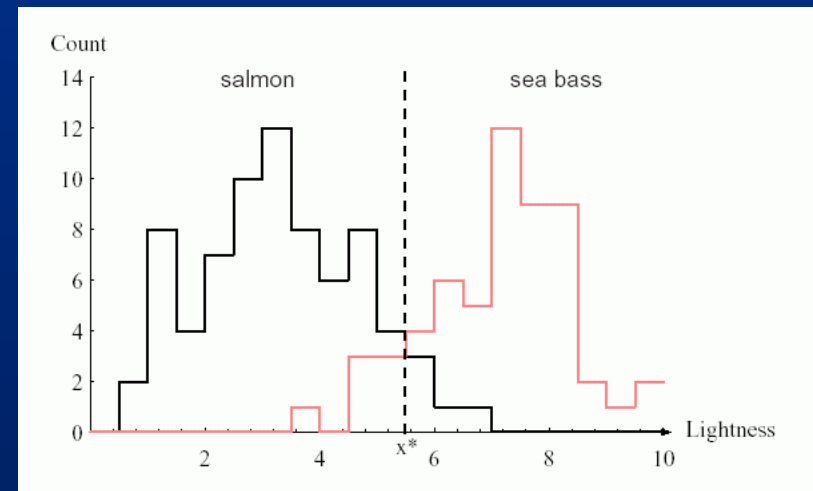
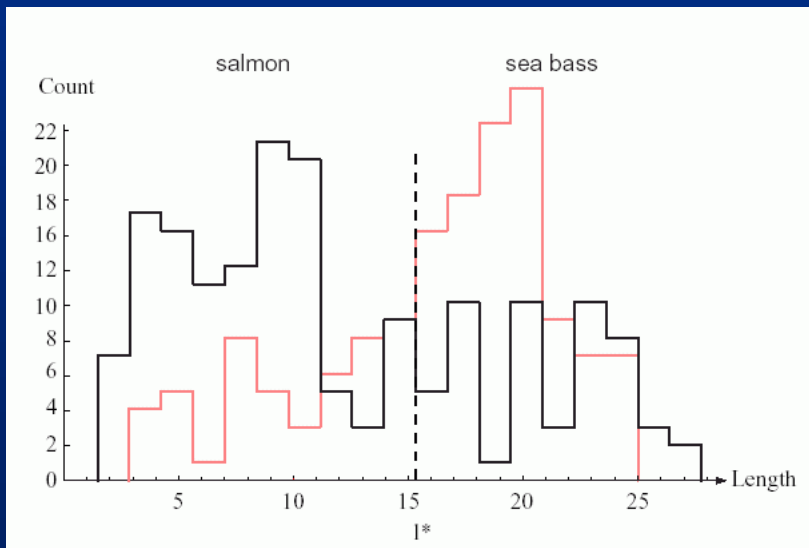
$$r_i = [x_{min} + (i-1)\Delta, x_{min} + i\Delta], i=1 \dots n$$

- prawdopodobieństwo łączne $P(C, r_i) = N(C, r_i)/N$, gdzie N = liczba ryb

łączne prawdopodobieństwo $P(C, r_i) = P(r_i|C)P(C)$

Przykład histogramów

Rozkład liczby ryb w dwóch wymiarach w 20 przedziałach: długość i jasność. Zaznaczono optymalne progi podziału.



$P(r_i|C)$ przybliża rozkład prawdopodobieństwa dla klasy $P(x|C)$.

Możemy go dokładnie obliczyć tylko w granicy nieskończenie wielu przykładów i podziału na nieskończenie wiele przedziałów.

W praktyce zawsze dzielimy na niewielką liczbę przedziałów.

Rodzaje prawdopodobieństwa

Tablica współwystępowania klasa-cecha: $P(C, r_i) = N(C, r_i) / N$

$N(C, r_i)$ = macierz,
rzędy = klasy, kolumny = cechy r_i

$P(C, r_i)$ – prawdopodobieństwo
łączne, P obserwacji obiektu z klasy
 C dla którego cecha $x \in r_i$

$P(C_1, r_1)$	$P(C_1, r_2)$	$P(C_1, r_3)$
$P(C_2, r_1)$	$P(C_2, r_2)$	$P(C_2, r_3)$
$P(C_3, r_1)$	$P(C_3, r_2)$	$P(C_3, r_3)$
$P(C_4, r_1)$	$P(C_4, r_2)$	$P(C_4, r_3)$
$P(C_5, r_1)$	$P(C_5, r_2)$	$P(C_5, r_3)$

$P(C)$ to prawd. *a priori* pojawienia się obiektów z danej klasy, przed wykonaniem pomiarów i określeniem, że $x \in r_i$ ma jakąś wartość.

To suma w danym rzędzie:

$$\sum_i P(C, x \in r_i) = P(C)$$

$P(x \in r_i)$ to prawd że znajdujemy jakąś obserwację dla której cecha $x \in r_i$ czyli suma dla danej kolumny.

$$\sum_j P(C_j, x \in r_i) = P(x \in r_i)$$

Prawdopodobieństwa warunkowe

Jeśli znana jest klasa C (rodzaj obiektu) to jakie jest prawdopodobieństwo że ma on własność $x \in r_i$?

$P(x \in r_i | C)$ oznacza warunkowe prawdopodobieństwo, że znając C cecha x będzie leżała w przedziale r_i .

Suma po wszystkich wartościach cech:

i dla łącznego
prawdopodobieństwa

$$\sum_i P(x \in r_i | C) = 1$$

Dlatego mamy:

$$\sum_i P(C, x \in r_i) = P(C)$$

$$P(x \in r_i | C) = P(C, x \in r_i) / P(C)$$

$P_C(x) = P(x | C)$ rozkład prawd. warunkowego to po prostu przeskalowane prawdopodobieństwo łączne, trzeba podzielić $P(C, x) / P(C)$

Formuły probabilistyczne

Relacje probabilistyczne wynikają z prostych reguł sumowania!

Macierz rozkładu łącznych prawdopodobieństw: $P(C, x)$ dla dyskretnych wartości obserwacji x , liczymy ile razy zaobserwowano łącznie $N(C, x)$, skalujemy tak by prawdop. sumowało się do 1, czyli $P(C, x) = N(C, x)/N$

Rząd macierzy $P(C, x)$ sumuje się do:

dlatego $P(x|C) = P(C, x)/P(C)$
sumuje się do

Kolumna macierzy $P(C, x)$ sumuje się do:

dlatego $P(C|x) = P(C, x)/P(x)$
sumuje się do

$$P(C) = \sum_{i=1}^n P(C, x_i);$$

$$\sum_{i=1}^n P(x_i | C) = 1;$$

$$P(x_i) = \sum_C P(C, x_i);$$

$$\sum_C P(C | x_i) = 1;$$

Twierdzenie Bayesa

Formuła Bayesa pozwala na obliczenie prawdopodobieństwa *a posteriori* $P(C|x)$ (czyli po dokonaniu obserwacji) znając łatwy do zmierzenia rozkład warunkowy $P(x|C)$.

Sumują się do 1 bo wiemy, że jeśli obserwujemy x_i to musi to być jedna z C klas, jak też wiemy, że jeśli obiekt jest z klasy C to x musi mieć jedną z wartości x_i

Obydwa prawdopodobieństwa są wynikiem podzielenia $P(C, x_i)$.

Formułka Bayesa jest więc oczywista.

Inaczej: H =hipoteza, E =obserwacja

$$\sum_C P(C) = 1;$$

$$\sum_i P(x_i) = 1$$

$$\sum_i P(x_i | C) = \sum_C P(C | x) = 1;$$

$$P(x_i | C) = P(C, x_i) / P(C)$$

$$P(C | x_i) = P(C, x_i) / P(x_i)$$

$$P(C | x_i) P(x_i) = P(x_i | C) P(C)$$

$$P(H | E) = P(E | H) P(H) / P(E)$$

Kwiatki

Mamy dwa rodzaje Irysów:
Irys Setosa oraz Irys Virginica



Długość liści określamy w dwóch przedziałach, $r_1=[0,3]$ cm i $r_2=[3,6]$ cm.
Dla 100 kwiatów dostajemy następujące rozkłady (Setosa, Virginica):

$$N(C, r) = \begin{pmatrix} 36 & 4 \\ 8 & 52 \end{pmatrix} \quad \begin{aligned} N(C_1) &= 40, N(C_2) = 60 \\ N(r_1) &= 44, N(r_2) = 56 \end{aligned}$$

Prawdopodobieństwa łączne różnych kwiatów Irysów: :

$$P(C, r) = \begin{pmatrix} 0.36 & 0.04 \\ 0.08 & 0.52 \end{pmatrix} \quad \begin{aligned} P(C_1) &= 0.4; P(r_1) = 0.44 \\ P(C_2) &= 0.6; P(r_2) = 0.56 \end{aligned}$$

Stąd
$$P(C | r) = \begin{pmatrix} 0.82 & 0.07 \\ 0.18 & 0.93 \end{pmatrix}; P(r | C) = \begin{pmatrix} 0.90 & 0.10 \\ 0.13 & 0.87 \end{pmatrix}$$

Martwić się czy nie?

Założmy, że w Polsce 1 na 1000 osób ma wirusa HIV.

Nowy test polegający na badaniu śliny, o dokładności 99%, wprowadzono do obowiązkowych badań okresowych.

Test wypadł pozytywnie. Jakie jest $P(\text{HIV})$?

Naiwne oszacowanie:

1 na 1000 osób ma wirusa HIV, czyli jeśli test ma dokładność 99%, to na 1000 osób wykaże 10 z HIV, a ponieważ jest tylko 1 ma wirusa to prawdopodobieństwo poprawnej identyfikacji $\sim 1/10=10\%$.

Widać więc, że nawet przy takiej dokładności, jeśli choroba jest dość rzadka (a tak jest w początkowym stadium epidemii), będzie 10 razy więcej błędów niż poprawnych identyfikacji.

Dokładniej to obliczymy korzystając z formułki Bayesa.

Mamy dwie klasy $H+$, $H-$ i dwie obserwacje $T+$, $T-$

Interesuje nas $P(H+ | T+)$, czyli: wynik testu był pozytywny, czy to HIV?

Co powie Bayes?

Mamy dwie klasy H+, H-, dwie obserwacje T+, T-

Interesuje nas prawdopodobieństw *a posteriori* $P(H+|T+)$,
jeśli wynik testu był pozytywny, jakie jest P(Hiv)? Oznaczmy je przez P(H+).

Znamy: $P(H+)=0.001$, $P(H-)=0.999$ (a priori, czyli bez innych informacji)

$P(T+|H+)=0.99= P(T-|H-)$ to dokładność testu,

zakładamy tu dla uproszczenia takie same błędy dla “fałszywie pozytywnych”
i “fałszywie negatywnych” przypadków (tak oczywiście być nie musi).

$P(H+|T+) P(T+) = P(H+,T+) = P(T+|H+) P(H+)$ (Bayes)

$P(T+)=P(T+, H+) + P(T+, H-)= P(T+|H+)*P(H+)+P(T+,H-);$

$P(T+, H-)=P(H-)-P(T-,H-)=P(H-)-P(T-|H-)*P(H-)=0.999(1-0.99)=0.00999$

$P(T+)=0.99*0.001+0.00999=0.01098$

$P(H+|T+) = P(T+|H+) P(H+)/P(T+) = 0.99*0.001/0.01098= 0.0902$

Czyli $P(\text{Covid}|\text{Test}) = 9.02\%$

Jeśli *a priori* mamy $P(H+)=0.01$, czyli 1:100, to $P(\text{Covid}|\text{Test}) = 50.0\%$,

[Kalkulator Bayesa](#)

Co na to Copilot MS?

$P(C)$ to prawd. że osoba ma Covid-19, które wynosi $1/1000 = 0.001$.

$P(T)$ to prawd. że test jest pozytywny.

$P(T|C)$ to prawd. że test jest pozytywny, jeśli osoba ma Covid-19, które wynosi 0.99 (dokładność testu).

Chcemy obliczyć $P(C|T)$, prawd. że to Covid-19, jeśli test jest pozytywny.

Zgodnie z twierdzeniem Bayesa, mamy: $P(C|T) = \frac{P(T|C) \cdot P(C)}{P(T|C) \cdot P(C) + P(T|\neg C) \cdot P(\neg C)}$

Prawd. że test jest pozytywny, $P(T)$, można obliczyć jako sumę prawd. że test jest pozytywny i osoba ma Covid-19, oraz prawd. że test jest pozytywny i osoba nie ma Covid-19. Czyli: $P(T) = P(T|C) \cdot P(C) + P(T|\neg C) \cdot P(\neg C)$

Gdzie $P(T|\neg C)$ to prawdopodobieństwo, że test jest pozytywny, jeśli osoba nie ma Covid-19. Jest to 1 minus dokładność testu, czyli 0.01.

$P(\neg C)$ to prawd. że osoba nie ma Covid-19, które wynosi $1 - P(C) = 0.999$.

Podstawiając te wartości do powyższego wzoru, otrzymujemy:

$$P(C|T) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.01 \cdot 0.999}$$

Po uproszczeniu, otrzymujemy: $P(C|T) \approx 0.09$

Prawdopodobieństwo, że osoba rzeczywiście ma wirusa wynosi około 9%.

Oceny dokładności

Jak oceniać dokładność modelu predykcyjnego?

Dla dwóch klas + i – tablica pomyłek (confusion matrix) dla M danych:

P(prawda model)=	TP=P ₊₊	FN=P ₊₋	TP+FN=P	TN+FP=N
	FP=P ₋₊	TN=P ₋₋		

TP = True positive, było C₊ a model przewiduje klasę C₊

TN = True negative, było C₋ a model przewiduje klasę C₋

FN = False Negative, było C₊ a model przewiduje klasę C₋

FP = False Positive, było C₋ i model przewiduje klasę C₊

Dokładność $A=(TP+TN)/M$ nie mówi jakiego rodzaju są pomyłki.

Dlatego używa się kombinacji elementów macierzy pomyłek.

Czułość, wrażliwość (sensitivity, recall) $S_+=TPR=TP/P=TP/(TP+FN)$

Swoistość, specyficzność (specificity, SPC) $S_-=TNR=TN/N=TN/(FP+TN)$

Precyzja (precision, positive predictive value, PPV) $PPV=TP/(TP+FP)$

Tablica pomyłek

Dokładność (accuracy) nie mówi nam jakiego rodzaju są pomyłki systemu klasyfikującego. Jest wiele miar, które to oceniają.

Program PyCM do analizy dokładności [na Githubie](#)

		Klasa predykowana – wynik testu			
Populacja		Klasyfikacja pozytywna	Klasyfikacja negatywna	Częstość występowania, chorobowość	
				$\frac{\sum \text{stan pozytywny}}{\sum \text{populacja}}$	
Klasa rzeczywista	Stan pozytywny	prawdziwie dodatnia, TP	falszywie ujemna (błąd drugiego rodzaju, FN)	czułość, TPR $S_+ = \frac{\sum TP}{\sum TP + \sum FN}$	FNR $\frac{\sum FN}{\sum TP + \sum FN}$
	Stan negatywny	falszywie dodatnia (błąd pierwszego rodzaju, FP)	prawdziwie ujemna, TN	FPR $\frac{\sum FP}{\sum FP + \sum TN}$	swoistość, SPC, TNR $S_- = \frac{\sum TN}{\sum FP + \sum TN}$
dokładność, ACC $\frac{\sum TP + \sum TN}{\sum \text{populacja}}$		precyzja, PPV $\frac{\sum TP}{\sum TP + \sum FP}$	FOR $\frac{\sum FN}{\sum FN + \sum TN}$	LR+ $\frac{TPR}{FPR}$	DOR $\frac{LR+}{LR-}$
		FDR $\frac{\sum FP}{\sum TP + \sum FP}$	NPV $\frac{\sum TN}{\sum FN + \sum TN}$	LR- $\frac{FNR}{TNR}$	

Inne miary

Dłuższa lista jest na stronie Wiki [Confusion Matrix](#).

Często stosowana jest miara F_1 , średnia harmoniczna precyzji i czułości:

$$F_1 = 2 / (1/TPR + 1/PPV) = 2 TP / (2 TP + FP + FN)$$

Współczynnik Phi lub Matthews Correlation Coefficient (MCC) stosowany jest w przypadku dużej różnicy liczebności klas.

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

Macierze pomyłek mogą być oczywiście zdefiniowane dla wielu klas, np. wielu chorób, ale wówczas trudno jest stosować różne miary.

Stąd często podajemy rozróżnienie dwóch stanów, wybrana klasa C i coś innego. W ten sposób testy statystyczne można stosować dla każdej z podejrzanych klas.

Lifts, czyli zyski kumulacyjne

Technika popularna w marketingu, gdzie skumulowane zyski i "wzrosty" są przedstawiane graficznie: wzrost jest miarą skuteczności przewidywań modelu = (wyniki uzyskane z)/(bez modelu predykcyjnego).

Np: czy X^i odpowie? Czy powinienem wysłać mu ofertę?
Założmy, że 20% osób odpowiada na Twoją ofertę.

Wysłanie tej oferty losowo do N osób daje $Y_0 = 0.2 * N$ odpowiedzi.

Model predykcyjny (w marketingu nazywany "modelem odpowiedzi").

$P(w_+ | X; M)$ wykorzystuje informacje X do przewidywania, kto odpowie na ofertę.

Uporządkuj przewidywania od najbardziej do najmniej prawdopodobnego:

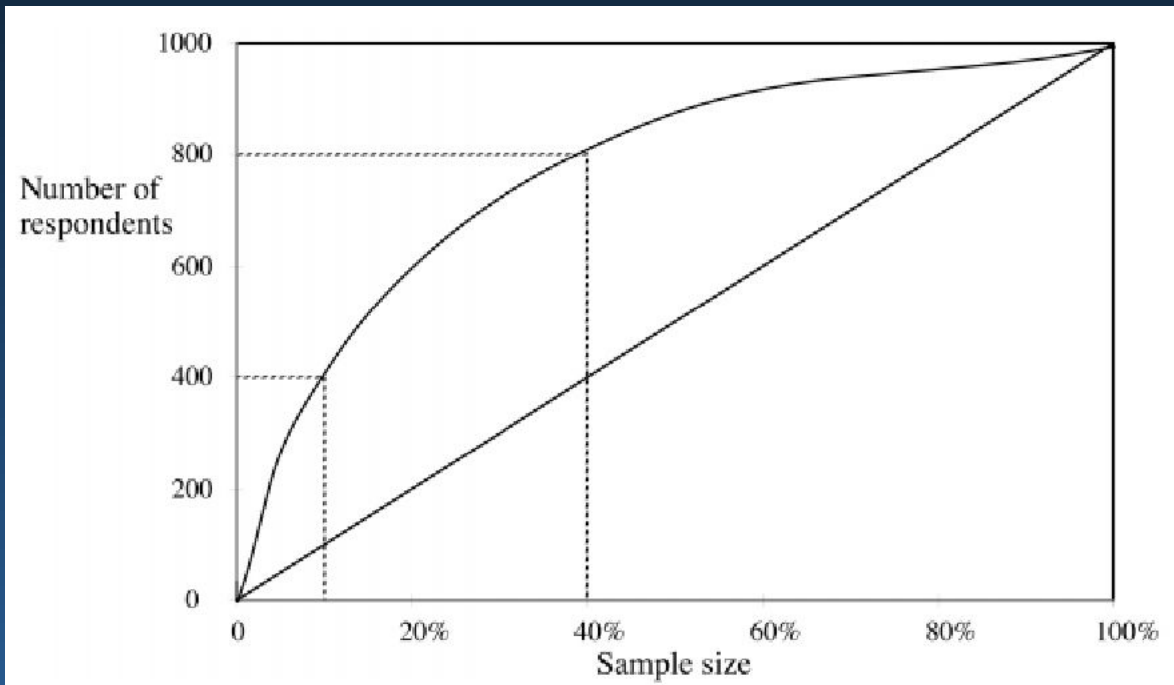
$$P(w_+ | X^1; M) > P(w_+ | X^2; M) \dots > P(w_+ | X^k; M)$$

Idealny model powinien umieścić te 20% osób, które odpowiedzą, na pierwszym miejscu. Wtedy liczba odpowiedzi to $Y_0 = 0.2 * N$ dla $j=1 \dots Y_0$.

W idealnym przypadku skumulowany przyrost będzie miał postać krzywej liniowej, która osiągnie Y_0 a następnie pozostanie stała;
wzrost będzie równy stosunkowi $Y(X^j) / 0.2 * j$.

Wykres kumulacyjnych zysków

Nie ma idealnego modelu, więc sprawdź, czy Twoje przewidywania $P(\omega_+ | X; M)$ są zgodne z rzeczywistością. Jeśli przewidywania były prawdziwe, wykreśl następny punkt o jedną jednostkę w prawo i jedną jednostkę w górę; jeśli przewidywania były fałszywe, wykreśl go o jedną jednostkę w prawo. Na osi pionowej znajduje się część P_+ wszystkich próbek danych $Y_0 =$ liczba wszystkich osób, które odpowiedziały (w tym przypadku $Y_0=1000$) z $N=5000$ osób.



	Rank $P(\omega_+ X)$	P/F
1	0.98	+
2	0.96	+
3	0.96	-
4	0.94	+
5	0.92	-
.....		
1000	0.08	-

[Przykłady są na tej stronie](#)

Wykresy ROC

Receiver Operator Characteristic (ROC): oceń TP (czyli P_{++}) jako funkcję pewnego progu, na przykład przy użyciu współczynnika prawdopodobieństwa:

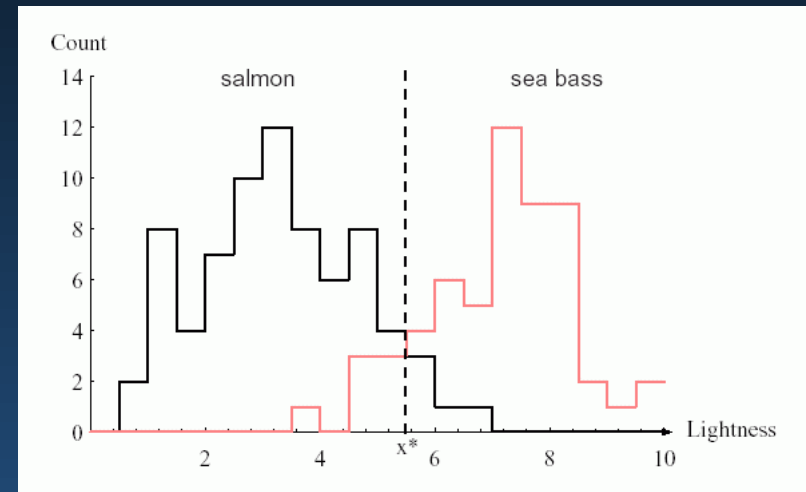
$$P(\mathbf{X} | \omega_1)P(\omega_1) > P(\mathbf{X} | \omega_2)P(\omega_2)$$

$$\Lambda(\mathbf{X}) = \frac{P(\mathbf{X} | \omega_1)}{P(\mathbf{X} | \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} = \theta$$

Próg może być traktowany jako zmienna mierząca wrażliwość testu TP/P; dla rozkładów 1D wysoki próg wykryje więcej przypadków pozytywnych (np. groźnej choroby), ale za to specyficzność testu TN/N maleje.

Jaki jest optymalny wybór? Zależy to od stosunku

$P_{-+} / P_{-} = \text{FP}/N = 1 - S$, czyli liczby fałszywych alarmów (FP, false positives) P_{-+} który jesteśmy skłonni zaakceptować.



Przykład ROC

Krzywe ROC przedstawiają $(S_+, 1-S_-)$, czyli błąd klasy C_- w stosunku do dokładności klasy C_+ dla różnych progów.

Idealny klasyfikator: poniżej pewnego progu $S_+ = 1$ (wszystkie przypadki pozytywne rozpoznane) dla $1-S_- = 0$ (brak fałszywych alarmów).

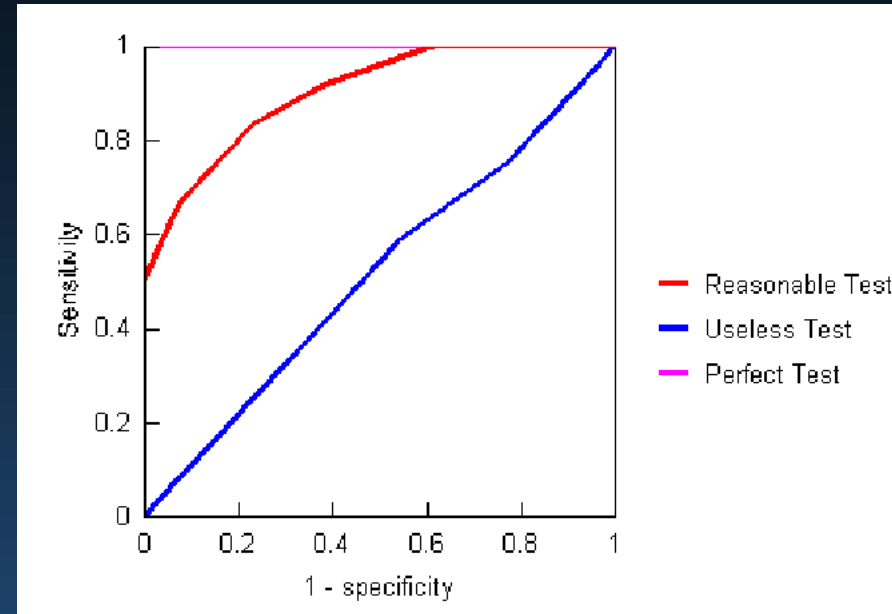
Bezużyteczny (niebieski): tyle samo TP jak i fałszywych alarmów FP dla dowolnego progu.

Rozsądny klasyfikator (czerwony), np: brak błędów do progu, który pozwala na rozpoznanie 0.5 pozytywnych przypadków, powoli rosnące błędy do $P_{+}=1-S_-=0.6$ dla 100% pozytywnych.

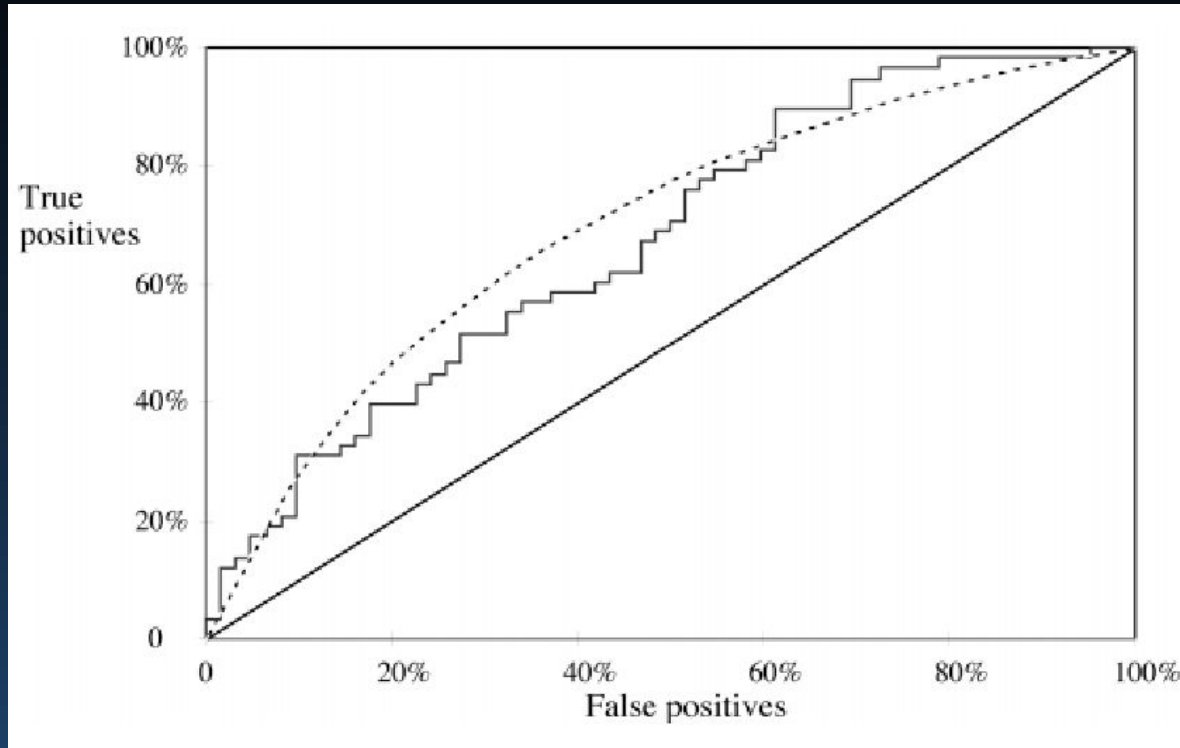
Dobra miara jakości: wysoki współczynnik **AUC (Area Under ROC Curve)**.

AUC = 0.5 oznacza przypadkowe zgadywanie,

AUC = 1 doskonałe przewidywanie.



Realistic ROC



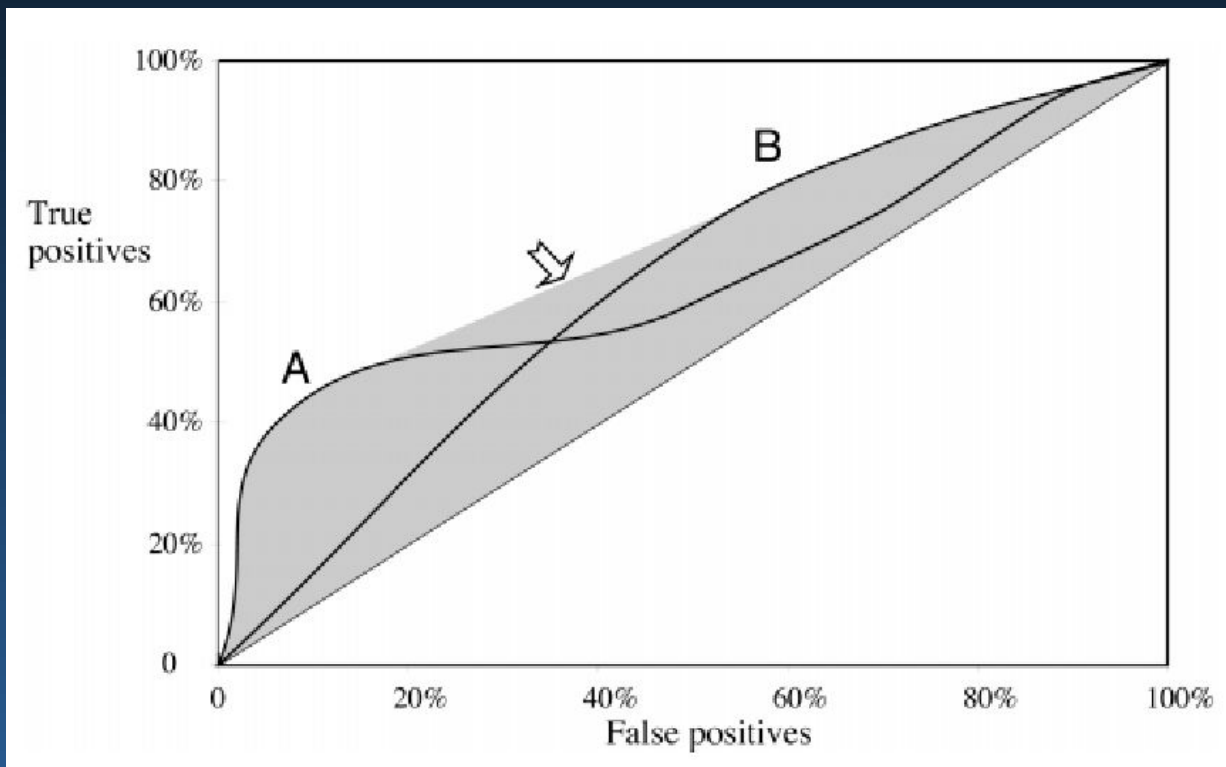
Krzywa ROC dla przypadku realistycznego, gdy jest skończona liczba progów i punktów. Zamiast procentu wyników prawdziwie pozytywnych (S_+ , recall, sensitivity) czasami używa się precyzji PPV.

Każdy punkt ROC zawiera wszystkie informacje zawarte w macierzy pomyłek dla pewnych parametrów (progów) modelu i pokazuje dla różnych prawdopodobieństw zaufanie do przewidywań klasyfikatora.

ROC dla porównania różnych modeli

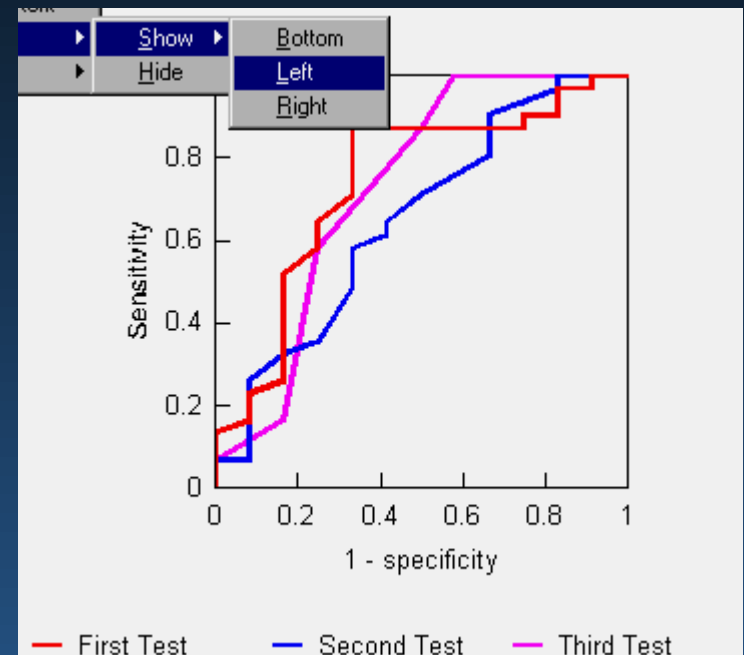
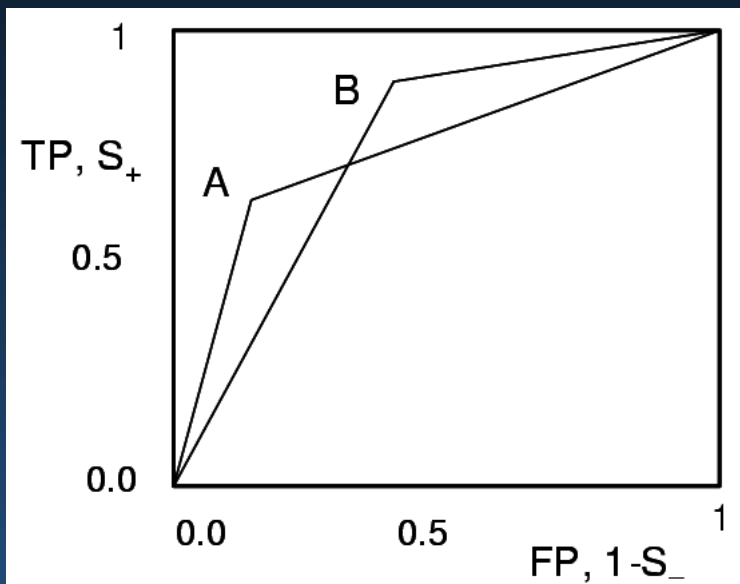
Idealna krzywa to 100% TP przy 0% FP. Bardziej wypukłe krzywe ROC wskazują na wyższość modeli dla wszystkich progów.

Krzywe ROC dla dwóch modeli pokażą mocne i słabe obszary: połączenie wyników dwóch modeli może dać krzywą ROC pokrywającą szary obszar.



ROC for logical rules

Klasyfikatory, które przewidują tylko tak/nie, np. reguły logiczne, dają tylko jeden punkt na krzywej ROC, w punkcie $(S_+, 1-S_-)$. Pole pod krzywą ROC $AUC = (S_+ + S_-)/2$ jest identyczne wzdłuż linii $S_+ + (1-S_-) = \text{const.}$



Jest wiele programów do analizy ROC np. [ROC Analysis for Windows](#).

Paradoks Monty Hall

Monty Hall Paradox, czyli przykład złudzenia kognitywnego.
Stosowany np. w teleturnieju „idź na całość”.

Reguły zabawy:

Mamy 3 kubki i złota monetę.

Wychodzisz z pokoju, ja pod jednym z kubków ukrywam monetę. Wracasz i wybierasz jeden z kubków.

Ja, wiedząc, pod którym jest moneta, odkrywam jeden z **pustych** kubków.

Masz teraz szansę zmienić swoją decyzję i pozostać przy już wybranym kubku lub wybrać pozostały.

Czy najlepszą strategią jest:

1. zawsze trzymanie się pierwotnego wyboru,
2. zawsze zmiana,
3. czy przypadkowy wybór?

[Zajrzyj tu by zagrać samemu.](#)

Wideo na ten temat – wielu profesorów się nabrało ...

Swobodny wybór



Eksperymenty psychologiczne:

Wybieramy cukierki różnych kolorów, wydaje się, że kolory R, G, B wybierane są równie często, więc zakładamy równe preferencje.

Dajemy do wyboru R i G, wybierane jest np. R

Dajemy do wyboru G i B, wybierane jest zwykle B.

Wnioski psychologów: mamy tu dysonans poznawczy, wybieramy B bo jak się raz decydujemy że nie chcemy G to później też nie wybieramy G.

Czy naprawdę? Dopiero w 2008 roku zauważono, że:

Jeśli początkowo były słabe preferencje $R > G$ to są 3 możliwości:

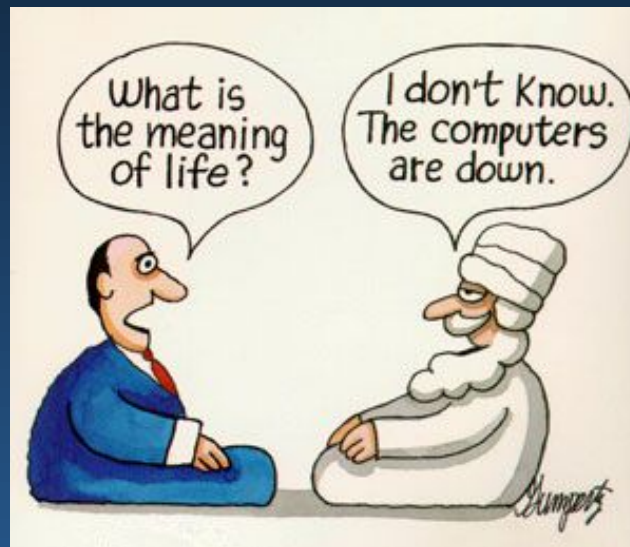
$R > G > B$, $R > B > G$, lub $B > R > G$, czyli 2/3 szans na wybór B zamiast G.

Być może wszystkie podobne psychologiczne eksperymenty były źle przeanalizowane? **Inverse base rates** i inne dziwne zjawiska?

Wnioski

Myślenie jest rzeczą trudną ...
prościej jest używać schematów.

Tylko w kontekście naturalnych sytuacji myślenie
przychodzi nam łatwo.



Przykładowe pytania

- Jak działa Teoretyk Logiki?
- Jakie były cele GPS? Czego nas nauczył GPS?
- Jaka jest kolejność ocen węzłów grafu w strategii minimaksu?
- Podać przykładowe funkcje oceny dla szachów.
- Do czego służy technika alfa-beta?
- Co umożliwia sprawne działanie w grach pomimo niewielkiej pojemności pamięci roboczej?
- Jaka jest pojemność pamięci roboczej człowieka i jakie inspiracje dla AI z tego wynikają?
- Jaką strategię stosują ludzie w grze w szachy?
- Wszyscy A to B. Żaden C nie jest A. Jaki stąd wniosek?
- Oszacuj jaka jest szansa choroby mając częstość jej występowania i dokładność testu, który wypadł pozytywnie.
- Oszacuj liczbę operacji wykonywanych przez mózg Kasparowa i wytłumacz, dlaczego przegrał z systemem Deep Blue.
- Narysować zależność stopnia kompetencji programu od szybkości szukania i wielkości jego bazy wiedzy.

Ciekawe linki

Modelowanie danych jest tu tylko wspomniane, ale warto zajrzeć do:

- Komiksu: P. Biecek, A. Kozak, A. Zawada, **Mini Wprowadzenie do Modelowania Predykcyjnego**. Warszawa 2022
- Książki: Przemysław Biecek and Tomasz Burzykowski, **Explanatory Model Analysis**. Explore, Explain, and Examine Predictive Models. With examples in R and Python. 2020