Sztuczna Inteligencja: metody predyktywne. Analiza Bayesa.

Włodzisław Duch

Katedra Informatyki Stosowanej UMK

Google: Wlodzislaw Duch

Strona wykładów

Martwić się czy nie?

Załóżmy, 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.

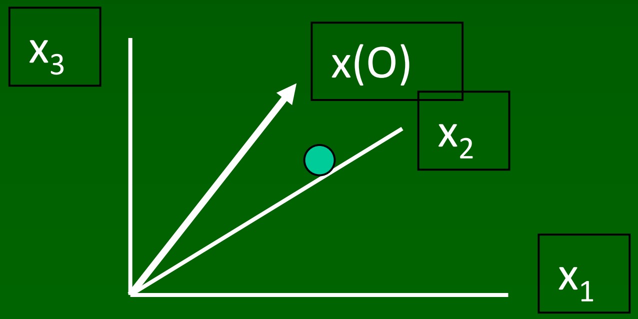
Obiekty w przestrzeni cech

- 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.

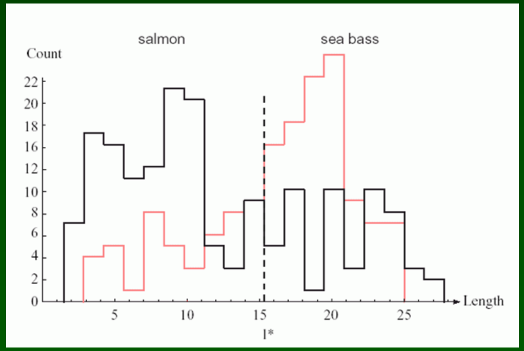
  Wektor cech $X = (x_1, x_2, x_3 \ldots x_d)$,
  o d-składowych wskazuje na punkt w przestrzeni cech.



Przykład: ryby

Przykład z: R. O. Duda, P. E. Hart and D. G. Stork, Wiley 2000,
Ch. 1.2, Pattern Classification (2nd ed)

Automatyzacja sortowania dwóch gatunków ryb, łososia i suma morskiego, które przesuwają się na pasie sortownika.
Czujniki oceniają różne cechy: długość, jasność, szerokość, liczbę płetw.



Zliczamy różne przypadki i tworzymy histogramy.

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

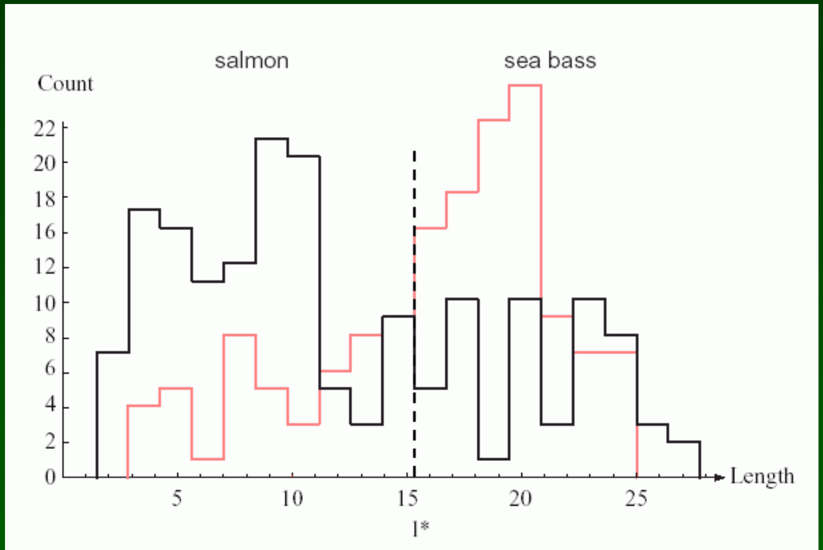
$$r_i = [x_{min}+(i-1)\Delta,\ x_{min}+i\Delta],\ i=1\ldots n$$

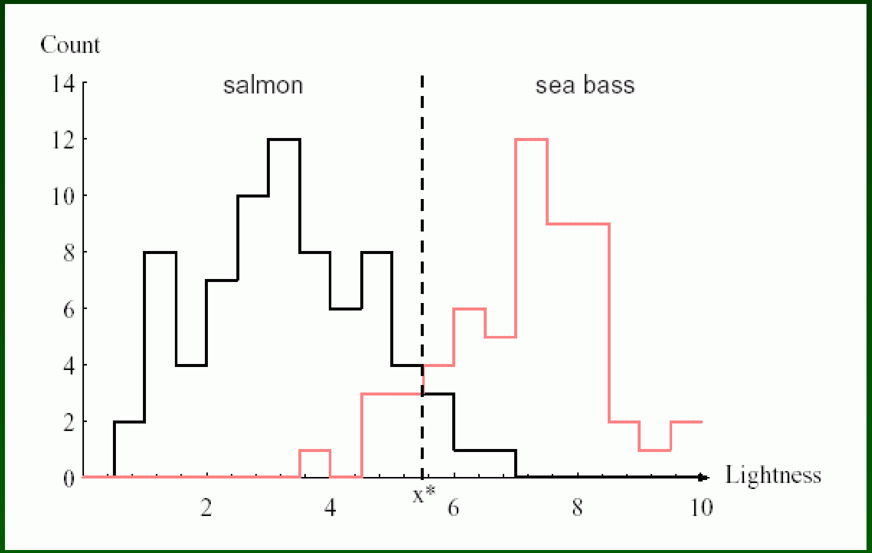
- prawdopodobieństwo łączne $P(C,r_i)=N(C,r_i)/N$, gdzie N = liczba ryb

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.

Dane

AI = algorytmy + hardware + dane.

Danologia, nauka o danych. Wiele problemów, Data Economy Congress.

Oczyszczanie danych – identyfikacja błędów, wartość danych.

Outliers – dane odstające.

Concept drift – zmiana ocenianych parametrów rozkładów w czasie.

Niepewność danych – słaba dokładność, zmienność danych medycznych.

Analiza starzenie się danych – coś nowego w dużych modelach.

Vela, D., Sharp, A., Zhang, R., Nguyen, T., Hoang, A., & Pianykh, O. S. (2022). Temporal quality degradation in AI models. Scientific Reports, 12(1), 11654.

Rodzaje prawdopodobieństw

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 $x\in r_i$

$$\begin{pmatrix} P(C_1,r_1) & P(C_1,r_2) & P(C_1,r_3) \\ P(C_2,r_1) & P(C_2,r_2) & P(C_2,r_3) \\ P(C_3,r_1) & P(C_3,r_2) & P(C_3,r_3) \\ P(C_4,r_1) & P(C_4,r_2) & P(C_4,r_3) \\ P(C_5,r_1) & P(C_5,r_2) & P(C_5,r_3) \end{pmatrix}$$

$P(C)$ 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 które 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:

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

i dla łącznego prawdopodobieństwa

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

Dlatego mamy:

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

$P_C(x)=P(x|C)$ 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 prawd. sumowało się do 1, czyli $P(C, x) = N(C,x)/N$

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

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

dlatego $P(x|C)=P(C, x)/P(C)$

sumuje się do

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

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

$$P(x_i) = \sum_{C} P(C, x_i);$$

dlatego $P(C|x)=P(C, x)/P(x)$

sumuje się do

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

Twierdzenie Bayesa

„Twierdzenie” to zwykła formułka Bayesa, pozwala na obliczenie prawdopodobieństwa *a posteriori* $P(C|x)$ (czyli po dokonaniu obserwacji), znając łatwy do oceny rozkład warunkowy $P(x|C)$.

Prawd. 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

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

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

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

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

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

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

Formułka Bayesa jest więc oczywista.

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

Inaczej: *H*=hipoteza, *E*=obserwacja

Prawd H mając E razy P(E) = prawd. E mając H razy P(H)

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

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):

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

Prawdopodobieństwa łączne różnych kwiatów Irysów: :

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

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?

Załóżmy, 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(HIV|T+)?

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 tylko jedna ma wirusa to prawdopodobieństwo poprawnej identyfikacji to tylko $\approx$ 1/10=10%.

Nawet przy takiej dokładności, jeśli choroba jest dość rzadka (a tak jest w początkowym stadium epidemii), będzie 10 razy więcej błędów niż poprawnych diagnoz.

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 testu 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(Covid|Test) = 9.02%
Jeśli *a priori* mamy P(H+)=0.01, czyli 1:100, to P(Covid|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: $P(C|T)=P(T)P(T|C)\cdot P(C)$

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|¬C) to prawdopodobieństwo, że test jest pozytywny, jeśli osoba nie ma Covid-19. Jest to 1 minus dokładność testu, czyli 0.01.
P(¬C) to prawd że osoba nie ma Covid-19, które wynosi 1 - P(C) = 0.999.

Podstawiając te wartości do powyższego wzoru, otrzymujemy:
$P(C|T)=0.99\cdot 0.001/(0.99\cdot 0.001+0.01\cdot 0.999)$

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)=		
	$TP=P_{++}$	$FN=P_{+-}$
	$FP=P_{-+}$	$TN=P_{--}$

$TP+FN=P \quad TN+FP=N$

$P/M+N/M = P_+ + P_- = 1$

- 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 (Confusion matrix)

Dokładność (accuracy) nie mówi nam jakiego rodzaju są pomyłki systemu klasyfikującego.
Jest wiele miar, które to oceniają, np. 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**	**fałszywie ujemna** (błąd drugiego rodzaju, **FN**)	czułość, TPR $S_+ = \frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FN}}$	FNR $\frac{\sum \mathrm{FN}}{\sum \mathrm{TP} + \sum \mathrm{FN}}$
	Stan negatywny	**fałszywie dodatnia** (błąd pierwszego rodzaju, **FP**)	**prawdziwie ujemna**, **TN**	FPR $\frac{\sum \mathrm{FP}}{\sum \mathrm{FP} + \sum \mathrm{TN}}$	swoistość, SPC, TNR $S_- = \frac{\sum \mathrm{TN}}{\sum \mathrm{FP} + \sum \mathrm{TN}}$
	dokładność, ACC $\frac{\sum \mathrm{TP} + \sum \mathrm{TN}}{\sum \text{populacja}}$	precyzja, PPV $\frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FP}}$	FOR $\frac{\sum \mathrm{FN}}{\sum \mathrm{FN} + \sum \mathrm{TN}}$	LR+ $\frac{\mathrm{TPR}}{\mathrm{FPR}}$	DOR $\frac{\mathrm{LR+}}{\mathrm{LR-}}$
		FDR $\frac{\sum \mathrm{FP}}{\sum \mathrm{TP} + \sum \mathrm{FP}}$	NPV $\frac{\sum \mathrm{TN}}{\sum \mathrm{FN} + \sum \mathrm{TN}}$	LR- $\frac{\mathrm{FNR}}{\mathrm{TNR}}$	

Inne miary

Dłuższa lista miar jest na stronie Wiki Confusion Matrix.

Często stosowana jest miara F_1, czyli 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.

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

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?
Załóżmy, 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:

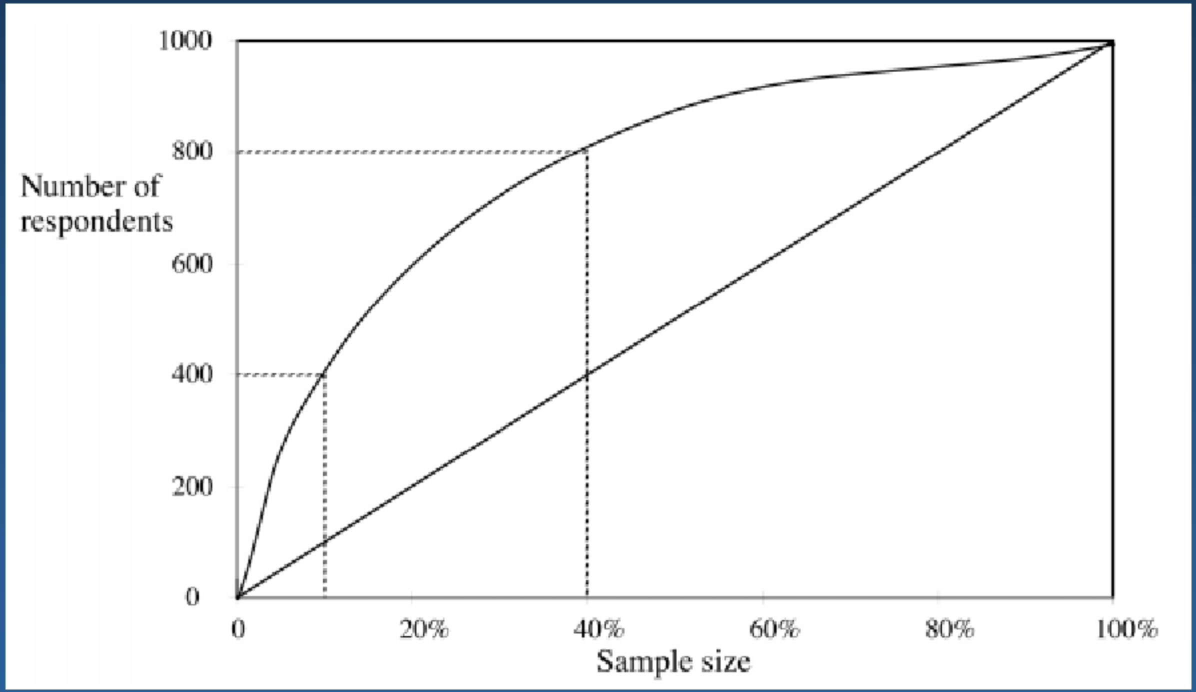
$$P(w_+|X^1;M) > P(w_+|X^2;M) \ldots > P(w_+|X^k;M)$$

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

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 $P(w_+|X;M)$ 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.



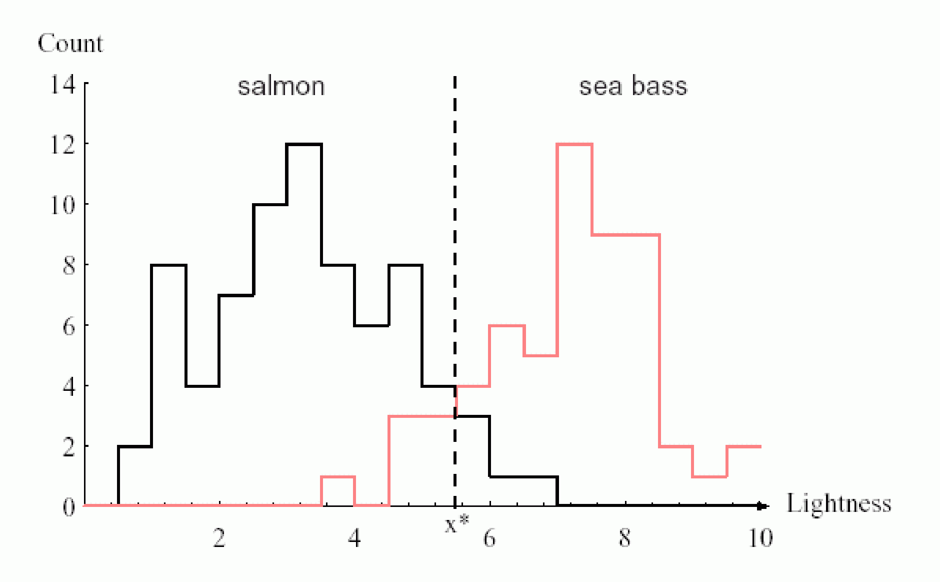
Rank	$P(\omega_+\|X)$	P/F
1	0.98	+
2	0.96	+
3	0.96	-
4	0.94	+
5	0.92	-
....		
1000	0.08	-

[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:

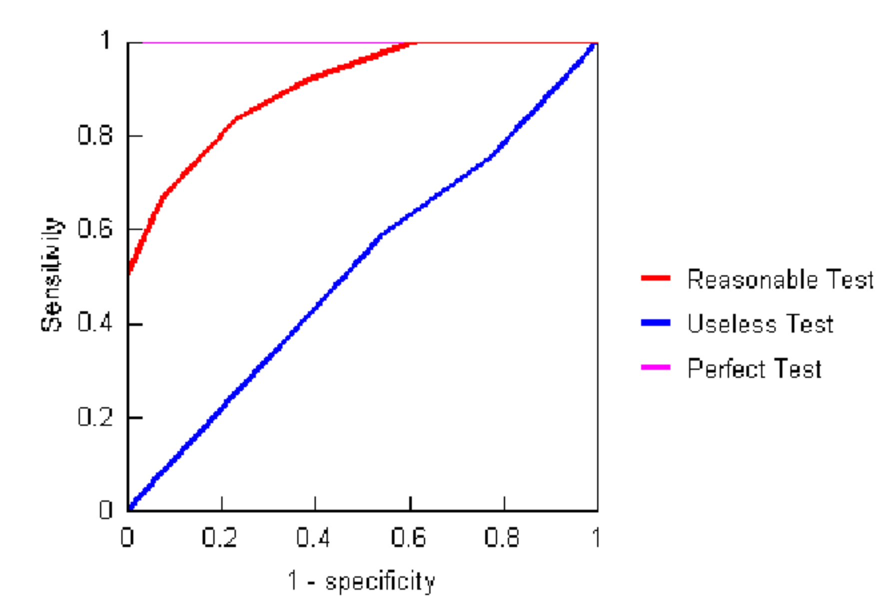
Próg może być traktowany jako zmienna mierząca wrażliwość testu TP/P; dla rozkładów 1D wysoki próg wykryje więcej przypadków pozytywnych (np. groźnej choroby), ale za to specyficzność testu TN/N maleje.



Jaki jest optymalny wybór? Zależy to od stosunku P_{-+} / P_{-} =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.



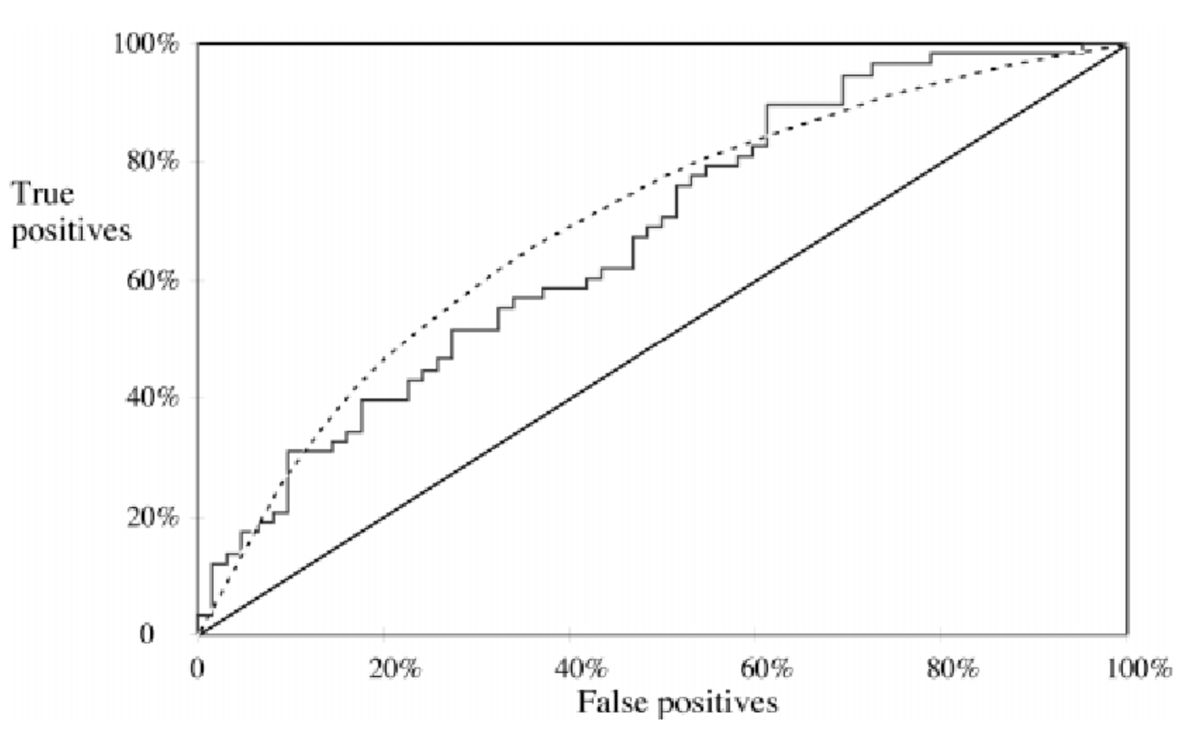
Idealny klasyfikator: poniżej pewnego progu $S_+ = 1$ (wszystkie przypadki pozytywne rozpoznane) dla $1-S_- = 0$ (brak fałszywych alarmów).

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.

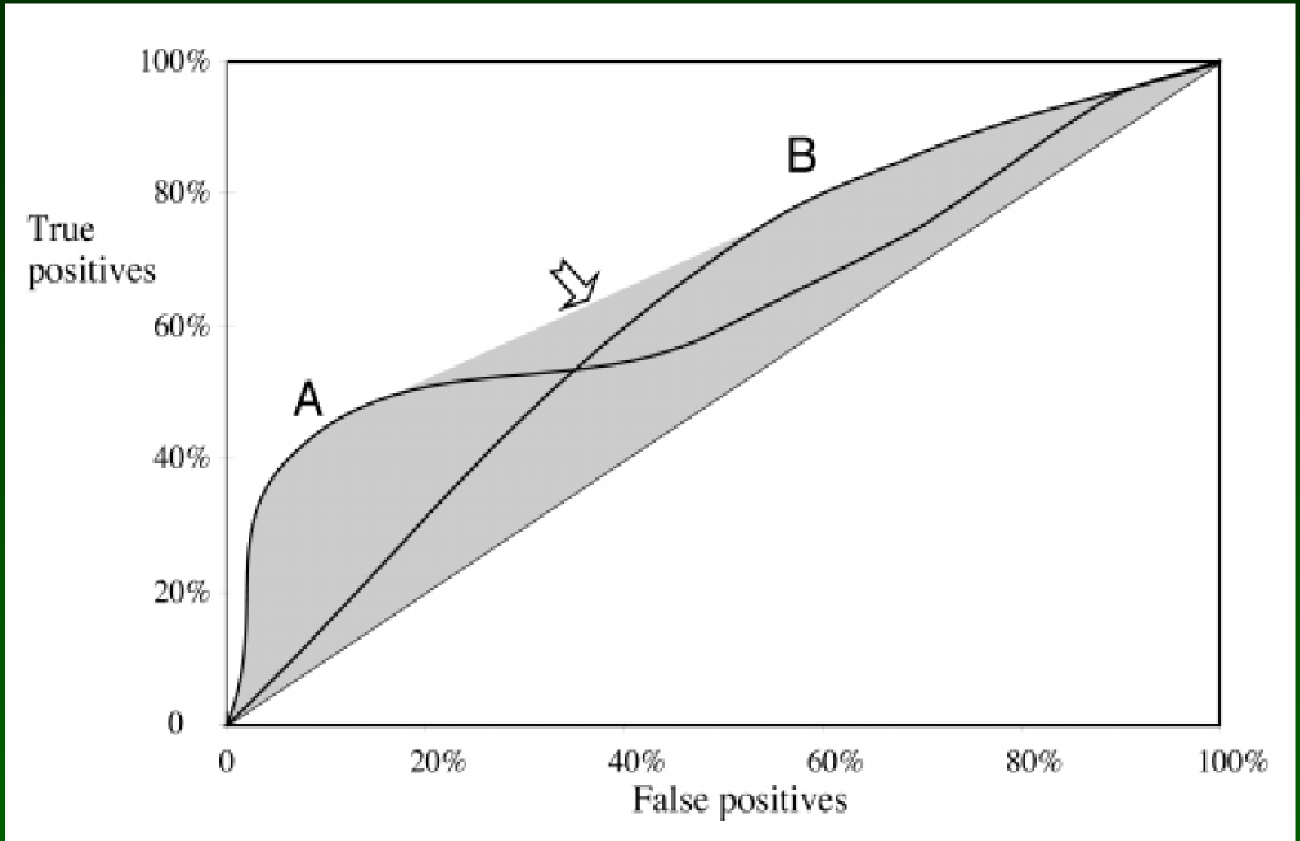
Dobra miara jakości: wysoki współczynnik AUC (Area Under ROC Curve).
AUC = 0.5 oznacza przypadkowe zgadywanie,
AUC = 1 doskonałe przewidywanie.

ROC dla porównania różnych modeli



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.

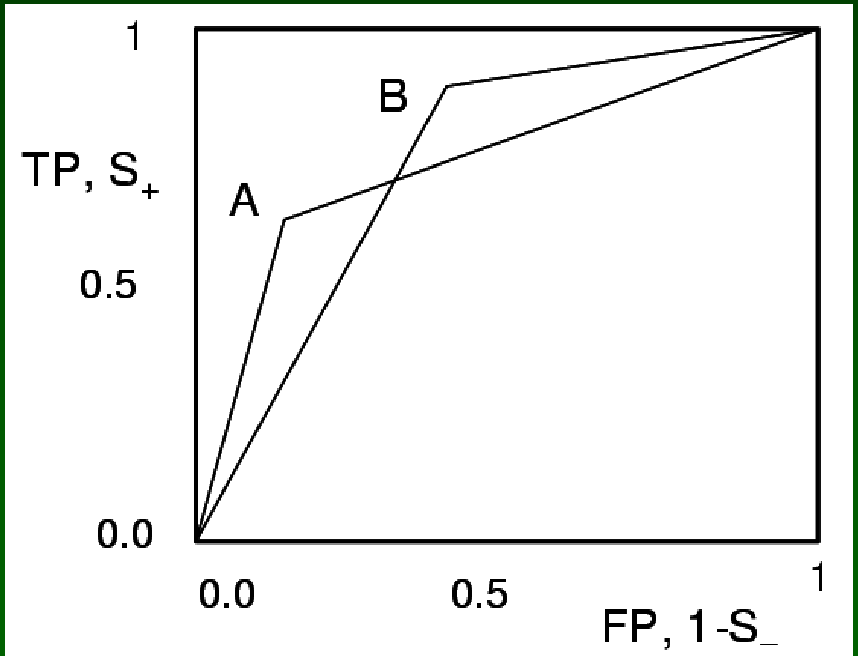
Kilka modeli

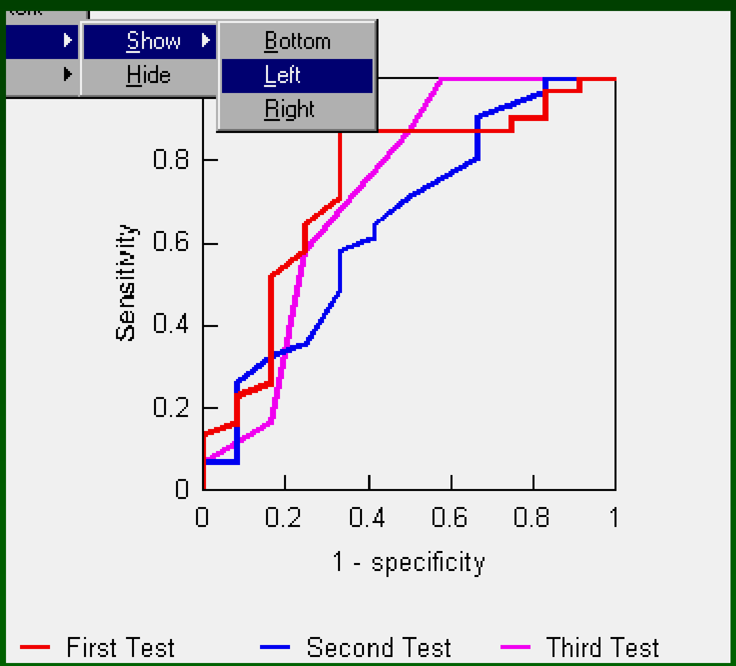


Idealna krzywa to 100% TP przy 0% FP.
Bardziej wypukłe krzywe ROC wskazują na wyższość modeli dla wszystkich progów.
Krzywe ROC dla dwóch modeli pokażą mocne i słabe obszary: połączenie wyników dwóch modeli może dać krzywą ROC pokrywającą szary obszar.

Rok dla logicznych reguł

Klasyfikatory binarne 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 $S_+ + (1-S_-) = const$.





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.

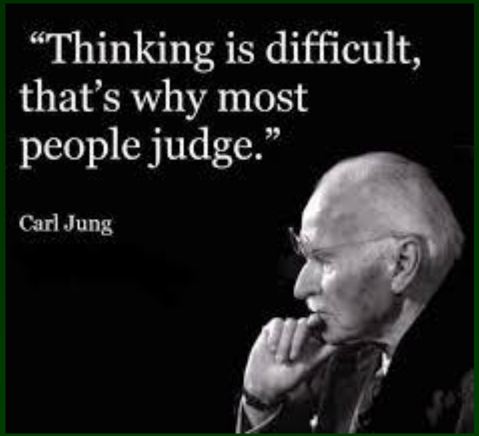
Czy naprawdę? Dopiero w 2008 roku zauważono, że:

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 w teorii kategoryzacji.

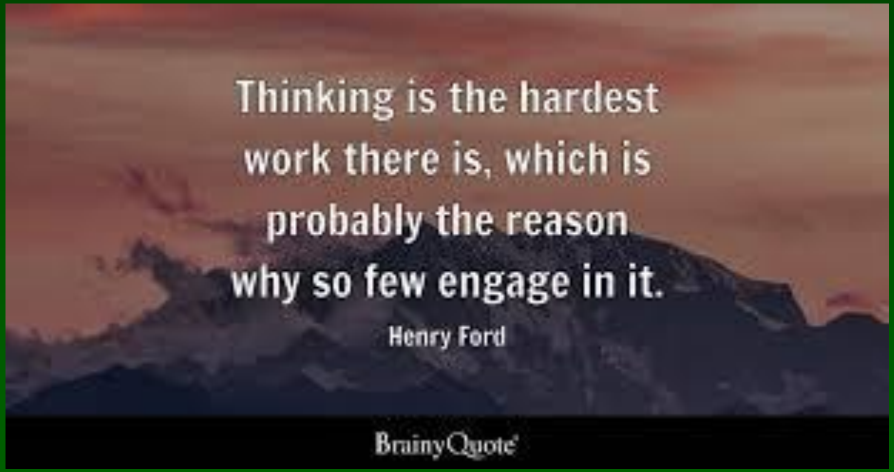
Wnioski



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

Myślenie probabilistyczne przychodzi nam z trudnością.
Popełniamy wiele błędów poznawczych. Ludziom wybaczamy, modelom AI nie.

Rozumowanie wymaga edukacji, strategii, systematycznych rozważań.



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: Przemyslaw Biecek and Tomasz Burzykowski,
  Explanatory Model Analysis. Explore, Explain, and Examine Predictive Models.
  With examples in R and Python. 2020

- Google LLM Bayesian training: YouTube summary.

- Paper: Qiu, L., Sha, F., Allen, K., Kim, Y., Linzen, T., & Steenkiste, S. van. (2026). Bayesian Teaching Enables Probabilistic Reasoning in Large Language Models. *Nature Communications*, *17*(1), 1238. https://doi.org/10.1038/s41467-025-67998-6

Pytania-probabilistyka

- Czym różnią się atrybuty od cech w kontekście opisu obiektów?
- Podaj przykład atrybutu kategorycznego nominalnego.
- Jak oblicza się łączną macierz prawdopodobieństw $P(C, x)$ dla wartości dyskretnych?
- Wyjaśnij, co oznacza prawdopodobieństwo a priori $P(C)$.
- Czym jest prawdopodobieństwo warunkowe $P(x|C)$?
- W jaki sposób twierdzenie Bayesa pozwala obliczyć prawdopodobieństwo a posteriori $P(C|x)$?
- Co oznacza wartość True Positive (TP) w tablicy pomyłek?
- Czym różni się czułość (sensitivity) od swoistości (specificity)?
- Jaka jest interpretacja współczynnika AUC (Area Under ROC Curve)?
- Wyjaśnij w kontekście paradoksu Monty Halla, dlaczego zmiana wyboru po odsłonięciu pustego kubka jest korzystniejsza.
- Wyjaśnić czemu analiza eksperymentów psychologicznych jest trudna.