

# Uczenie maszynowe

Marcin Jukiewicz    Andrzej Gajda

Sztuczna Inteligencja

08.04.2020

# Uczenie maszynowe

Uczeniem maszynowym nazywamy dziedzinę nauki (i sztukę) programowania komputerów w sposób umożliwiający im **uczenie się z danych**.

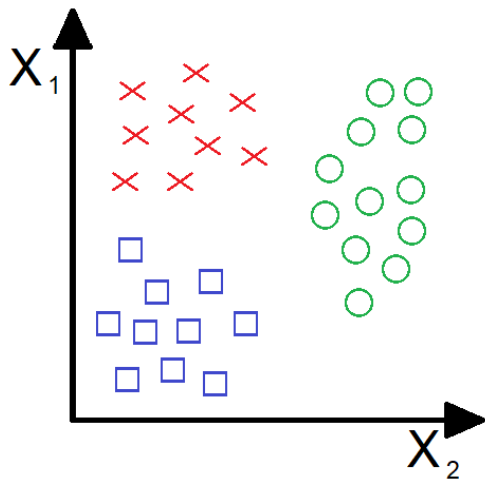
*Aurélien Géron, 2017*

[Uczenie maszynowe to] dziedzina nauki dająca komputerom możliwość uczenia się bez konieczności ich jawnego programowania.

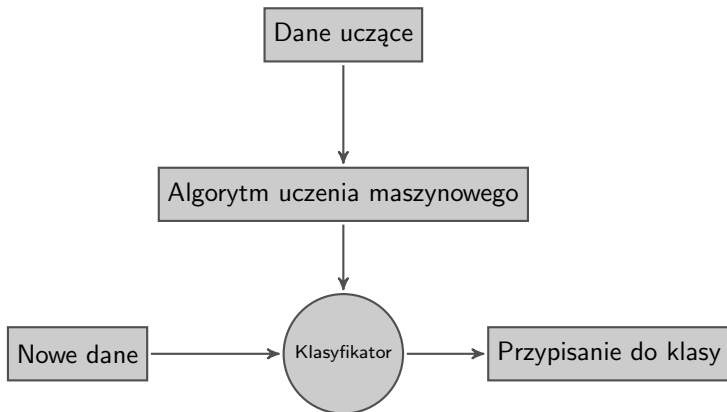
*Arthur Samuel, 1959*

- Do rozwiązywania problemów, które wymagają wielu „ręcznych” ustawień parametrów algorytmu lub składają się z wielu reguł. Algorytmy uczenia maszynowego potrafią „generalizować”.
- Do rozwiązywania problemów złożonych.
- Do zmiennych środowisk; algorytmy potrafią dostosowywać się do nowych danych.
- Do problemów składających się z olbrzymiej ilości danych.

- filtr antyspamowy
- diagnostyka medyczna
- rozpoznawanie obrazów
- podpowiedzi na różnych portalach, co jeszcze mogłoby się podobać
- translator Google
- badanie zdolności kredytowej
- rozpoznawanie mowy
- kierowanie pojazdem
- granie w gry
- włamania lub dziwne zachowania w sieci



## Schemat blokowy uczenia maszynowego



**Atrybuty** (bądź cechy) to pewne cechy charakteryzujące właściwości rozważanych obiektów.

**Zbiór uczący** to zbiór przykładów, obserwacji, próbek służący do wytrenowania klasyfikatora.

**Klasa** to zbiór obiektów charakteryzujących się pewnymi wspólnymi właściwościami.



- **Uczenie nadzorowane i nienadzorowane**
  - Uczenie nadzorowane (z nauczycielem)
  - Uczenie nienadzorowane (bez nauczyciela)
  - Uczenie częściowo nadzorowane
  - Uczenie przez wzmacnianie
- **Uczenie wsadowe/przyrostowe**
- **Uczenie z przykładów/modelu**

- Niedobór danych uczących
- Niereprezentatywne dane uczące
- Dane kiepskiej jakości
- Nieistotne cechy
- Przetrenowanie danych uczących
- Niedotrenowanie danych uczących

		Klasa predykowana — wynik testu	
		Klasyfikacja pozytywna	Klasyfikacja negatywna
Klasa rzeczywista	Stan pozytywny	Prawdziwie pozytywna, TP	Fałszywie negatywna (błąd drugiego rodzaju), FN
	Stan negatywny	Fałszywie pozytywna (błąd pierwszego rodzaju), FP	Prawdziwie negatywna, TN

		Rozpoznany jako	
		pies	kot
Obiekt	pies	9	5
	kot	6	20

### Skuteczność

$$\frac{TP + TN}{\text{Liczebność populacji}}$$

ang.: *Accuracy, ACC*

### Czułość, pełność

$$\frac{TP}{TP + FN}$$

ang.: *True positive rate (TPR), Recall, Sensitivity*

:)

$$\frac{FP}{FP + TN}$$

ang.: *False positive rate (FPR), Fall-out*

### Precyzja

$$\frac{TP}{TP + FP}$$

ang.: *Precision, Positive predictive value (PPV)*

		Rozpoznany jako	
		pies	$\sim$ pies
Obiekt	pies	9	5
	$\sim$ pies	6	20

$$TPR = \frac{TP}{TP + FN} = \frac{9}{9 + 5} = 0.64$$

Jak dużo wystąpień psa zakwalifikowaliśmy do klasy 'pies'? Jest to zdolność klasyfikatora do prawidłowego określania klasy pozytywnej.

$$FPR = \frac{FP}{FP + TN} = \frac{6}{6 + 20} = 0.23$$

Jak dużo wystąpień innych zwierząt zakwalifikowaliśmy do klasy 'pies'? Jest to zdolność klasyfikatora do prawidłowego określania klasy negatywnej.

$$PPV = \frac{TP}{TP + FP} = \frac{9}{9 + 6} = 0.60$$

Jak często klasę 'pies' udało nam się zakwalifikować poprawnie?

Współczynnik kappa informuje nas, gdzie w skali pomiędzy klasyfikacją losową a idealną, znajduje się dany klasyfikator.

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

gdzie:

- $p_o$  — trafność obserwowana
- $p_e$  — trafność oczekiwana



## Współczynnik kappa — jak wyznaczyć?

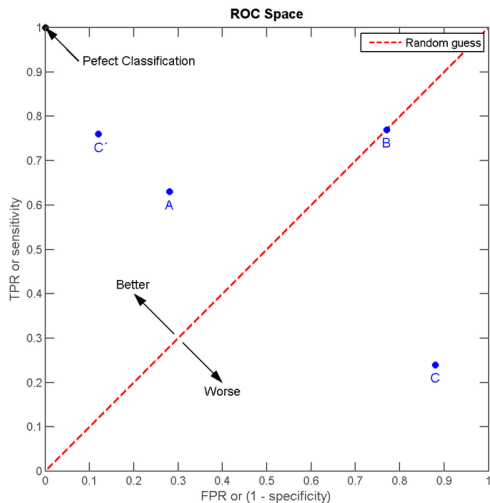
		Rozpoznany jako	
		pies	~pies
Obiekt	pies	a	b
	~pies	c	d

$$p_o = \frac{9 + 20}{9 + 5 + 6 + 20} = \frac{29}{40} = 0.725$$

$$\begin{aligned}
 p_e &= \frac{a + b}{a + b + c + d} \frac{a + c}{a + b + c + d} + \frac{c + d}{a + b + c + d} \frac{b + d}{a + b + c + d} = \\
 &= \frac{9 + 5}{9 + 5 + 6 + 20} \frac{9 + 6}{9 + 5 + 6 + 20} + \frac{6 + 20}{9 + 5 + 6 + 20} \frac{5 + 20}{9 + 5 + 6 + 20} = \\
 &= 0.350 * 0.375 + 0.650 * 0.625 = 0.131 + 0.406 = 0.538
 \end{aligned}$$

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{0.725 - 0.538}{1 - 0.538} = 0.405$$

# Krzywa ROC i pole pod krzywą



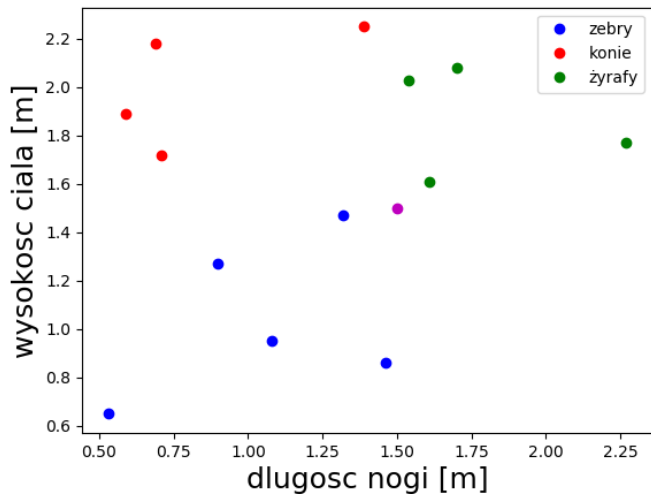
<b>Outlook</b>	<b>Temp</b>	<b>Humidity</b>	<b>Windy</b>	<b>Play Golf</b>
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Algorytm  $k$  najbliższych sąsiadów ( $k$ -NN, *k Nearest Neighbours*)

Ogólny schemat:

- 1 Poszukaj  $k$  najbliższych punktów (sąsiadów) dla analizowanego przypadku.
- 2 Przypisz rozważany przypadek do występującej najczęściej wśród sąsiadów klas.

<b>długość nogi [m]</b>	<b>wysokość ciała [m]</b>	<b>zwierzę</b>
1.46	0.86	zebry
1.32	1.47	zebry
0.9	1.27	zebry
1.08	0.95	zebry
0.53	0.65	zebry
1.39	2.25	konie
0.69	2.18	konie
0.59	1.89	konie
0.71	1.72	konie
2.27	1.77	żyrafy
1.61	1.61	żyrafy
1.7	2.08	żyrafy
1.54	2.03	żyrafy



Odległość pomiędzy dwoma punktami  $p_1 = (x_1, y_1)$  oraz  $p_2 = (x_2, y_2)$ .

### Odległość euklidesowa

$$d(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

### Odległość taksówkowa (Manhattan)

$$d(A, B) = |x_1 - x_2| + |y_1 - y_2|$$

<b>długość nogi [m]</b>	<b>wysokość ciała [m]</b>	<b>zwierzę</b>	<b>odległość</b>
1.46	0.86	zebry	0.64
1.32	1.47	zebry	0.18
0.9	1.27	zebry	0.64
1.08	0.95	zebry	0.69
0.53	0.65	zebry	1.29
1.39	2.25	konie	0.76
0.69	2.18	konie	1.06
0.59	1.89	konie	0.99
0.71	1.72	konie	0.82
2.27	1.77	żyrafy	0.82
1.61	1.61	żyrafy	0.16
1.7	2.08	żyrafy	0.61
1.54	2.03	żyrafy	0.53



Dla  $k = 1$  dominująca klasa to: żyrafy

Dla  $k = 3$  dominująca klasa to: żyrafy

Dla  $k = 5$  dominująca klasa to: żyrafy

Dla  $k = 7$  dominująca klasa to: zebry



- nie trzeba uczyć klasyfikatora
- „kosztowny” proces klasyfikacji
- wrażliwość na dane zaszumione lub błędne
- dobrze radzi sobie z dużą liczbą klas

Algorytm k-średnich (centroidów, *k-means*)

**Centroid** to punkt związany z wielokątem, leżący wewnątrz niego, reprezentujący jego geometryczne „środek”.

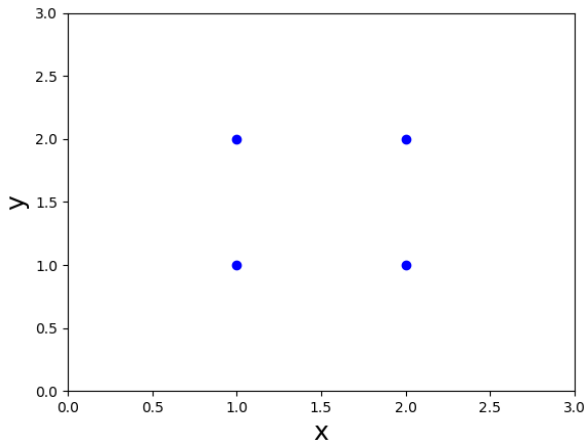
Schemat:

- 1 wybieramy  $k$  początkowych punktów centroidów (np. na drodze losowania)
- 2 obliczamy odległość pomiędzy każdym punktem danych i każdym z  $k$  punktów centroidów
- 3 dzielimy punkty danych na  $k$  skupień, dany punkt należy do tego skupienia, do którego centroidu ma najbliższą
- 4 uaktualniamy położenie centroidu — liczymy średnią wszystkich współrzędnych punktów należących do danego skupienia
- 5 jeśli nie jest spełnione kryterium końca wracamy do pkt 2

## Dane punkty

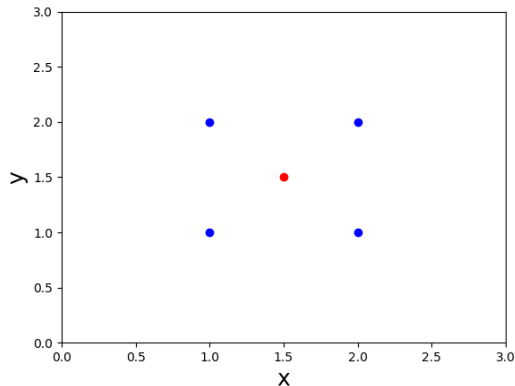
---

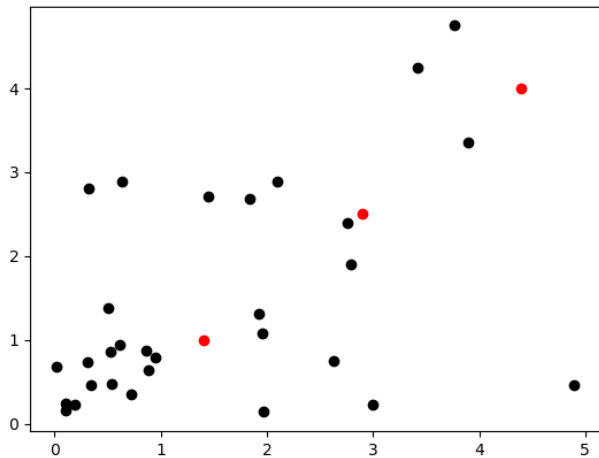
(1,1)  
(1,2)  
(2,1)  
(2,2)

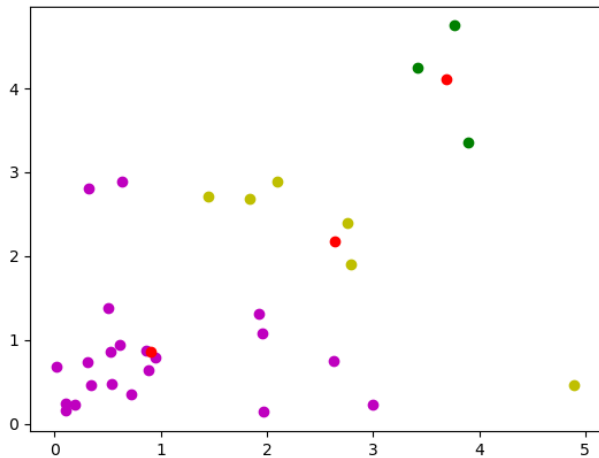


$$x_c = \frac{1 + 1 + 2 + 2}{4} = 1.5$$

$$y_c = \frac{1 + 2 + 1 + 2}{4} = 1.5$$









### Zalety:

- niewielka złożoność obliczeniowa
- prosta idea
- wydajny dla dużych zbiorów i niedużej liczbie grup

### Wady:

- mało odporny na zaszumione dane
- problem z ustaleniem liczby skupień
- czasem różne wartości początkowe prowadzą do różnych wyników
- działa dobrze tylko dla „sferycznych” skupisk o podobnej gęstości

### Algorytm k-mendoidów (*k-medoids*)

Jest to bliźniaczy dla k-średnich algorytm, w którym nowy punkt centralny powstaje przez wybranie z danego skupienia punktu, który ma najmniejszą sumę odległości do wszystkich pozostałych punktów w tym skupieniu.

Pozostały schemat działania pozostaje bez zmian.

### Zalety:

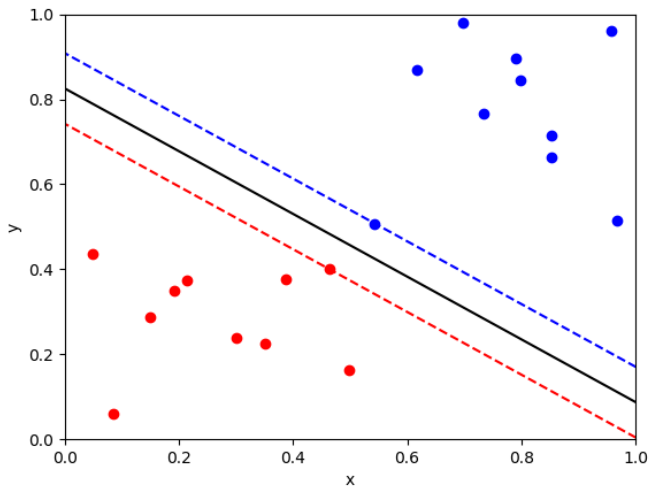
- lepiej radzi sobie z dalekimi obiektami (punktami), bardziej odporny na szum
- prosta idea
- początkowy zbiór nie wpływa na wyniki

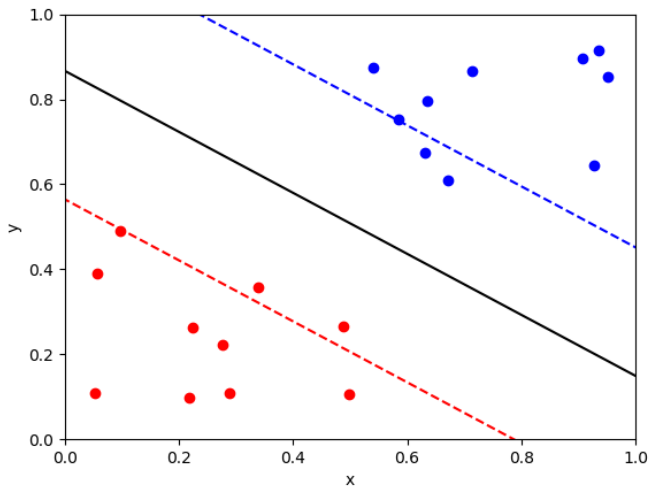
### Wady:

- gorzej radzi sobie z dużymi zbiorami danych
- problem z ustaleniem liczby skupień
- potrzebna więcej zasobów dla dużej liczby skupień i obiektów

**Maszyna wektorów nośnych** lub **maszyna wektorów wspierających** (*support vector machine*) do klasyfikacji wykorzystuje proces nazywany „klasyfikowaniem maksymalnego marginesu”. Proces ten ma dwa warianty:

- klasyfikacja twardego marginesu
- klasyfikacja miękkiego marginesu





## Zalety:

- skuteczność
- działa dobrze z małymi zestawami danych uczących
- radzi sobie z nadmiernym dopasowaniem danych

## Wady:

- czas szkolenia może być długi dla dużych zbiorów
- ma problemy z nakładającymi się zbiorami

**Nadmierne dopasowanie, przeuczenie, przetrenowanie, overfitting**  
jeśli klasyfikator budowany jest na niewystarczającej ilości danych, będzie on dawał gorsze wyniki dla danych, z którymi wcześniej się nie zetknął.

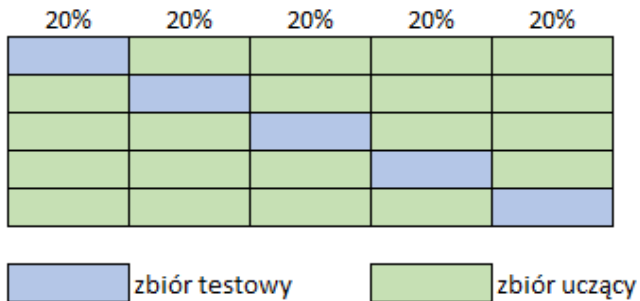
Krótko mówiąc, zjawisko to zachodzi, kiedy::

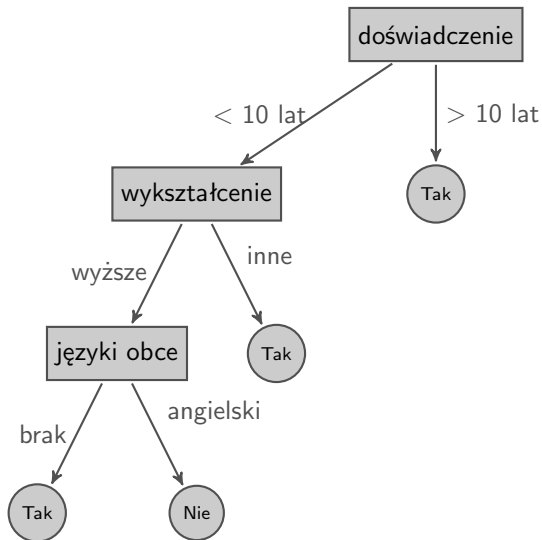
$$ACC(dane\_uczace) > ACC(dane\_testowe)$$



- test na zbiorze uczącym
- test na wydzielonym zbiorze testowym
- walidacja krzyżowa

**Walidacja krzyżowa, krosvalidacja, sprawdzanie krzyżowe, k-fold cross-validation** metoda polegająca na podziale analizowanego zbioru na podzbiory, a następnie trenowanie klasyfikatora na części z nich (zbiór uczący), podczas gdy pozostałe służą do sprawdzenia wiarygodności jej wyników (zbiór testowy).





- **Węzeł** — w węźle znajduje się zawsze albo test na wartości pewnego atrybutu, albo klasa decyzyjna.
- **Krawędź** — łączy dwa węzły, przy czym strzałka określa kierunek połączenia. Z każdą krawędzią związana jest pewna etykieta, np. niskie, średnie, tak, nie. Etykiety to możliwe wartości atrybutu związanego z węzłem, z którego wyszła ta krawędź.
- **Liść** — węzeł, z którego nie wychodzi żadna krawędź. W liściu zawsze znajdować się będzie przypisanie do jakiejś klasy decyzyjnej, w tym wypadku do TAK albo do NIE.
- **Korzeń** — od korzenia drzewo zaczyna rosnąć. Jest to węzeł, do którego nie dochodzi żadna krawędź.

## Entropia (funkcja przyrostu informacji)

Średnia ilość informacji, przypadająca na pojedynczą wiadomość ze źródła informacji. Innymi słowy jest to średnia ważona ilości informacji niesionej przez pojedynczą wiadomość, gdzie wagami są prawdopodobieństwa nadania poszczególnych wiadomości.

Funkcja oceny testu, nazywana przyrostem informacji, jest obliczana następująco:

$$g_t(P) = I(P) - E_t(P)$$

gdzie:

$$I(P) = \sum_{v \in V} -\frac{|P^v|}{|P|} \log \frac{|P^v|}{|P|}$$

$$E_t(P) = \sum_{r \in R_t} \frac{|P_{tr}|}{|P|} E_{tr}(P)$$

$$E_{tr}(P) = \sum_{v \in V} -\frac{|P_{tr}^v|}{|P_{tr}|} \log \frac{|P_{tr}^v|}{|P_{tr}|}$$

	<b>Outlook</b>	<b>Temp</b>	<b>Humidity</b>	<b>Windy</b>	<b>Play Golf</b>
<b>1</b>	Rainy	Hot	High	False	No
<b>2</b>	Rainy	Hot	High	True	No
<b>3</b>	Overcast	Hot	High	False	Yes
<b>4</b>	Sunny	Mild	High	False	Yes
<b>5</b>	Sunny	Cool	Normal	False	Yes
<b>6</b>	Sunny	Cool	Normal	True	No
<b>7</b>	Overcast	Cool	Normal	True	Yes
<b>8</b>	Rainy	Mild	High	False	No
<b>9</b>	Rainy	Cool	Normal	False	Yes
<b>10</b>	Sunny	Mild	Normal	False	Yes
<b>11</b>	Rainy	Mild	Normal	True	Yes
<b>12</b>	Overcast	Mild	High	True	Yes
<b>13</b>	Overcast	Hot	Normal	False	Yes
<b>14</b>	Sunny	Mild	High	True	No

$$|T^{Yes}| = |\{3, 4, 5, 7, 9, 10, 11, 12, 13\}| = 9$$

$$|T^{No}| = |\{1, 2, 6, 8, 14\}| = 5$$

$$|T_{Outlook,Sunny}| = |\{1, 2, 8, 9, 11\}| = 5$$

$$|T_{Outlook,Sunny}^{Yes}| = |\{9, 11\}| = 2$$

$$|T_{Outlook,Sunny}^{No}| = |\{1, 2, 8\}| = 3$$

$$|T_{Outlook,Overcast}| = |\{3, 7, 12, 13\}| = 4$$

$$|T_{Outlook,Overcast}^{Yes}| = |\{3, 7, 12, 13\}| = 4$$

$$|T_{Outlook,Overcast}^{No}| = |\emptyset| = 0$$

$$|T_{Outlook,Rainy}| = |\{4, 5, 6, 10, 14\}| = 5$$

$$|T_{Outlook,Rainy}^{Yes}| = |\{4, 5, 10\}| = 3$$

$$|T_{Outlook,Rainy}^{No}| = |\{6, 14\}| = 2$$



Entropia dla kolejnych elementów:

$$E_{Outlook,Sunny}(P) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.971$$

$$E_{Outlook,Overcast}(P) = -\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4} = 0$$

$$E_{Outlook,Rainy}(P) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$$

Entropia ważona:

$$E_{Outlook}(T) = \frac{5}{14} \cdot 0.971 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0.971 = 0.694$$

Informacja zawarta w zbiorze dla atrybutu *Outlook*:

$$I(T) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} = 0.940$$

Przyrost informacji:

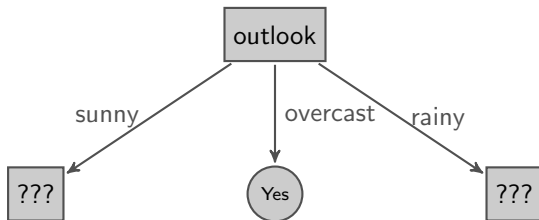
$$g_{Outlook}(T) = 0.940 - 0.694 = 0.246$$

W ten sam sposób wyznaczono kolejno:

$$g_{Temperature}(T) = 0.940 - 0.911 = 0.029$$

$$g_{Humidity}(T) = 0.940 - 0.788 = 0.152$$

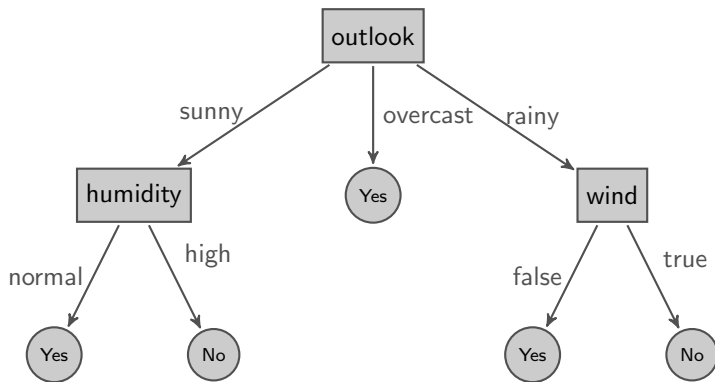
$$g_{Windy}(T) = 0.940 - 0.892 = 0.048$$



Dla gałęzi drzewa możemy analizować skrócone tabele:

<b>Outlook</b>	<b>Temp</b>	<b>Humidity</b>	<b>Windy</b>	<b>Play Golf</b>
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Sunny	Mild	Normal	False	Yes
Sunny	Mild	High	True	No

<b>Outlook</b>	<b>Temp</b>	<b>Humidity</b>	<b>Windy</b>	<b>Play Golf</b>
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Rainy	Mild	Normal	True	Yes



outlook = sunny  $\wedge$  humidity = high  $\rightarrow$  No

outlook = sunny  $\wedge$  humidity = normal  $\rightarrow$  No

outlook = overcast  $\rightarrow$  Yes

outlook = rainy  $\wedge$  wind = false  $\rightarrow$  Yes

outlook = rainy  $\wedge$  wind = true  $\rightarrow$  No

### Zalety:

- jest to efektywna metoda reprezentacji pojęć
- wykorzystuje dane ilościowe i jakościowe
- odporna na szumy w danych
- metoda czytelna dla człowieka
- możliwe łatwe przejście do reguł

### Wady:

- drzewa mogą się zbyt szybko rozrastać
- każda zmiana w drzewie powoduje, że algorytm trzeba uruchomić na nowo
- może popełniać wiele błędów, jeśli coś blisko korzenia „pójdzie nie tak”
- atrybuty decyzyjne powinny być dyskretne

Naiwny klasyfikator bayesowski jest prostym probabilistycznym (czyli wykorzystującym prawdopodobieństwo) klasyfikatorem. Jego „naiwność” polega na założeniu o niezależności atrybutów, tj. wzajemną niezależność zmiennych niezależnych. Często nie mają one żadnego związku z rzeczywistością i właśnie z tego powodu nazywa się je naiwnymi.

*Model cech niezależnych* można wyprowadzić korzystając z twierdzenia Bayesa.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

gdzie:

- $P(A|B)$  — prawdopodobieństwo warunkowe zajścia zdarzenia  $A$ , pod warunkiem zajścia zdarzenia  $B$
- $P(B|A)$  — prawdopodobieństwo warunkowe zajścia zdarzenia  $B$ , pod warunkiem zajścia zdarzenia  $A$
- $P(A)$  — prawdopodobieństwo zajścia zdarzenia  $A$
- $P(B)$  — prawdopodobieństwo zajścia zdarzenia  $B$



	Chorzy	Zdrowi	
Test +	99	99	198
Test -	1	9801	9802
	100	9900	10000

- $P(A) = \frac{100}{10000} = 0.01$  — prawdopodobieństwo choroby
- $P(B) = \frac{198}{10000} = 0.0198$  — prawdopodobieństwo testu +
- $P(B|A) = 0.99$  — prawdopodobieństwo testu +, jeżeli ma się chorobę

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.99 \cdot 0.01}{0.0198} = 0.5$$

$$P(A|B) = P(A_1 \wedge A_2 \wedge \dots \wedge A_n|B) = P(A_1|B) \cdot P(A_2|B) \cdots P(A_n|B)$$

$$P(X|B) = \frac{P(B|X)P(X)}{P(B)}$$

$$P(Y|B) = \frac{P(B|Y)P(Y)}{P(B)}$$

$$P(X|B) = P(B_1|X) \cdot P(B_2|X) \cdots P(B_n|X) \cdot P(X) \cdot C$$

$$P(Y|B) = P(B_1|Y) \cdot P(B_2|Y) \cdots P(B_n|Y) \cdot P(Y) \cdot C$$

<b>Outlook</b>	<b>Temp</b>	<b>Humidity</b>	<b>Windy</b>	<b>Play Golf</b>
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	No
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Mamy zdecydować, czy grać, czy nie grać w golfa, jeśli:  
 jest pochmurno, chłodno, wilgotność normalna i nie wieje wiatr?

$$P(\text{YES}) = \frac{8}{14}$$

$$P(\text{NO}) = \frac{6}{14}$$

$$P(\text{Overcast}|\text{YES}) = \frac{3}{8}$$

$$P(\text{Cool}|\text{YES}) = \frac{3}{8}$$

$$P(\text{Normal}|\text{YES}) = \frac{6}{8}$$

$$P(\text{False}|\text{YES}) = \frac{5}{8}$$

$$P(\text{Overcast}|\text{NO}) = \frac{1}{6}$$

$$P(\text{Cool}|\text{NO}) = \frac{1}{6}$$

$$P(\text{Normal}|\text{NO}) = \frac{1}{6}$$

$$P(\text{False}|\text{NO}) = \frac{3}{6}$$

$$P(\text{Example}|\text{YES}) = \frac{8}{14} \cdot \frac{3}{8} \cdot \frac{3}{8} \cdot \frac{6}{8} \cdot \frac{5}{8} = 0.038$$

$$P(\text{Example}|\text{NO}) = \frac{6}{14} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{3}{6} = 0.001$$

$$P(X) = \frac{\text{liczba\_przykladow}(X)}{\text{laczna\_liczba\_przykladow}} \rightarrow$$

$$P(X) = \frac{\text{liczba\_przykladow}(X) + 1}{\text{laczna\_liczba\_przykladow} + k \cdot \text{liczba\_klas}}$$

$$P(X|Y) = \frac{\text{liczba\_przykladow}(X \cap Y)}{\text{laczna\_liczba\_przykladow}(Y)} \rightarrow$$

$$P(X|Y) = \frac{\text{liczba\_przykladow}(X \cap Y) + 1}{\text{laczna\_liczba\_przykladow}(Y) + k \cdot \text{liczba\_klas}}$$

Celem selekcji cech jest wybrane ze zbioru cech takich, które zapewniają nam możliwie najlepszą klasyfikację. W zbiorze cech mogą wystąpić cechy redundantne, czyli niosące identyczną informację jak istniejące już cechy lub cechy wprowadzające szum, obniżające skuteczność klasyfikacji.

Selekcja ma wpływ na:

- poprawę wyników predykcji,
- zmniejszenie wymagań obliczeniowych,
- zmniejszenie wymagań odnośnie gromadzenia danych,
- redukcję kosztów przyszłych pomiarów,
- poprawę jakości danych.

Przykłady metod selekcji cech:

- Metody rankingowe,
- Selekcja podzbioru cech,
- Metody wyspecjalizowane:
  - Relief,
  - mRMR,
  - SBS, SFS
  - itp.
- Algorytmy genetyczne

**Dziękujemy za uwagę**