Statystyczna Eksploracja Danych

Wykład 4 - drzewa klasyfikacyjne

dr inż. Julian Sienkiewicz

22 marca 2019

Drzewa klasyfikacyjne: geneza

Drzewa klasyfikacyjne: geneza

Pierwszy raz pojawiły się kontekście badań socjologicznych dotyczących kwestionariuszy:

Drzewa klasyfikacyjne: geneza

Pierwszy raz pojawiły się kontekście badań socjologicznych dotyczących kwestionariuszy:

- Morgan, Sonquist, *Problems in the analysis of survey data and a proposal*, J. Am. Stat. Assoc. **58**, 415 (1963)

Drzewa klasyfikacyjne: geneza

Pierwszy raz pojawiły się kontekście badań socjologicznych dotyczących kwestionariuszy:

- Morgan, Sonquist, *Problems in the analysis of survey data and a proposal*, J. Am. Stat. Assoc. **58**, 415 (1963)
- autorzy zauważali, że większość metod analizy polega na założeniu *addytywności* składników

Drzewa klasyfikacyjne: geneza

Pierwszy raz pojawiły się kontekście badań socjologicznych dotyczących kwestionariuszy:

- Morgan, Sonquist, *Problems in the analysis of survey data and a proposal*, J. Am. Stat. Assoc. **58**, 415 (1963)
- autorzy zauważali, że większość metod analizy polega na założeniu *addytywności* składników
- ich podejście brało pod uwagę interakcje, działało sekwencyjnie i jako funcje celu stawiało sobie obniżenie błedów przewidywania

Drzewa klasyfikacyjne: geneza

Pierwszy raz pojawiły się kontekście badań socjologicznych dotyczących kwestionariuszy:

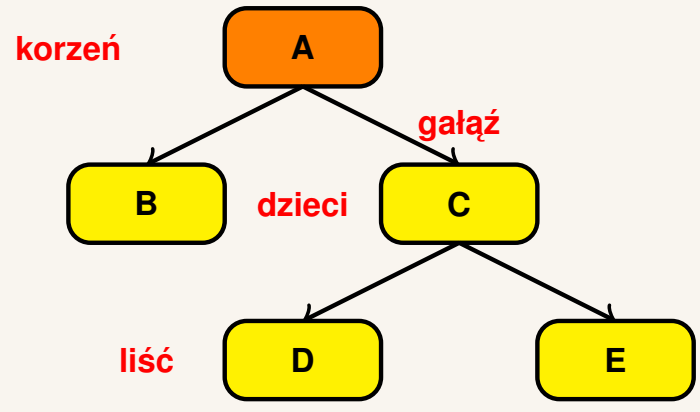
- Morgan, Sonquist, *Problems in the analysis of survey data and a proposal*, J. Am. Stat. Assoc. **58**, 415 (1963)
- autorzy zauważali, że większość metod analizy polega na założeniu *addytywności* składników
- ich podejście brało pod uwagę interakcje, działało sekwencyjnie i jako funcje celu stawiało sobie obniżenie błedów przewidywania

Drzewa klasyfikacyjne: definicja

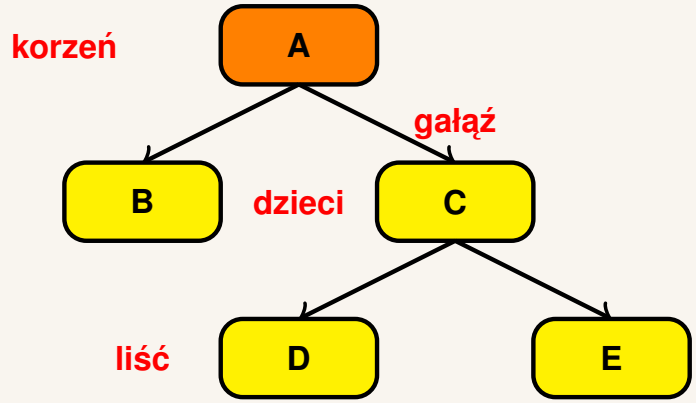
Drzewo skierowane (nieskierowany, acykliczny i spójny graf), posiadające jedyny dający się wyróżnić wierzchołek (**korzeń**), będący węzłem początkowym drzewa.

Struktura drzewa

Struktura drzewa

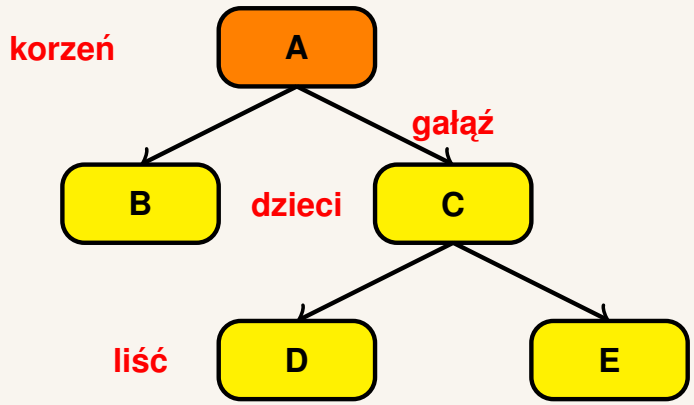


Struktura drzewa



Nazewnictwo

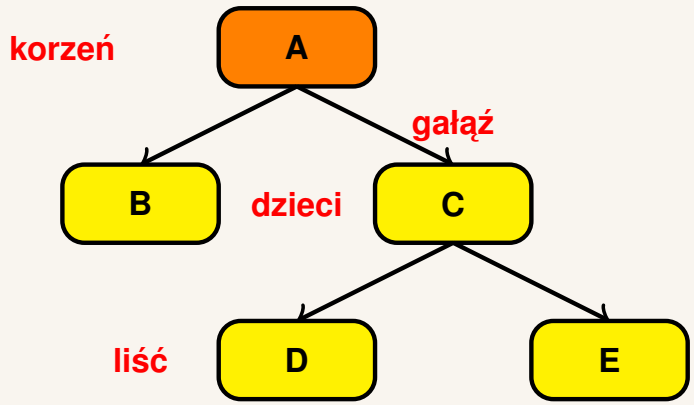
Struktura drzewa



Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,

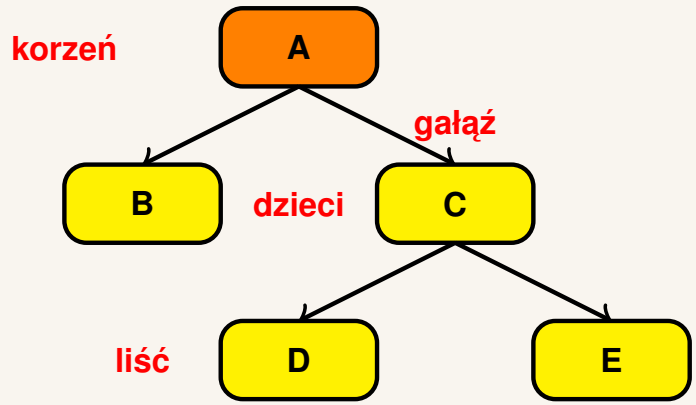
Struktura drzewa



Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,

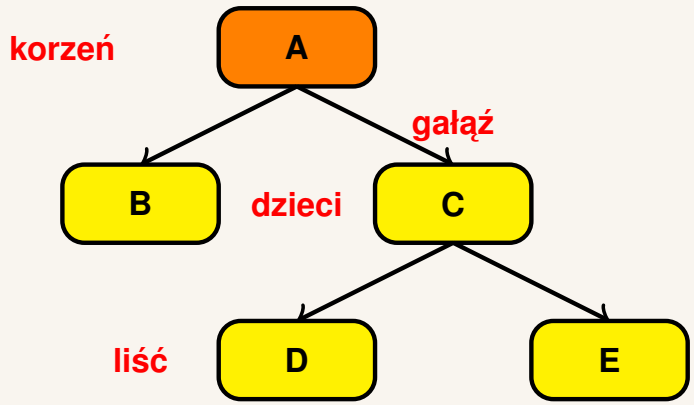
Struktura drzewa



Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,
- węzeł bez dzieci, to **liść**

Struktura drzewa

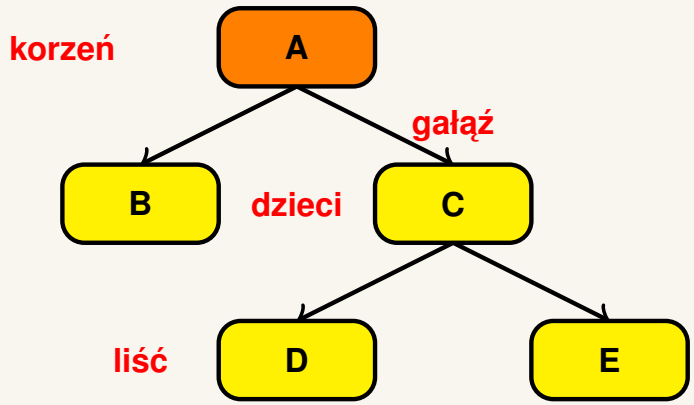


Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,
- węzeł bez dzieci, to **liść**

- konwencja rysowania drzew rosnących od góry w dół [sic!]: korzeń na górze, na dole liście

Struktura drzewa

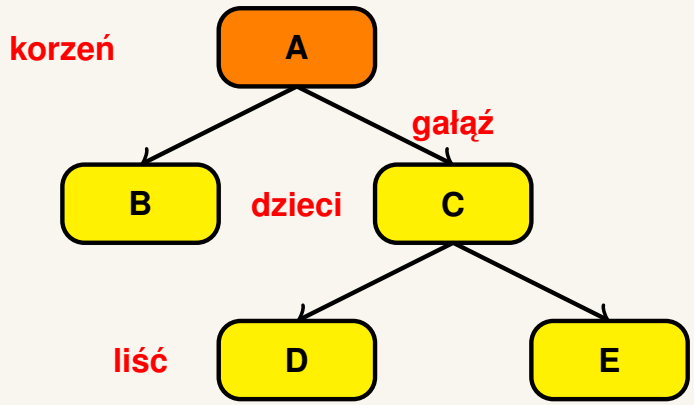


Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,
- węzeł bez dzieci, to **liść**

- konwencja rysowania drzew rosnących od góry w dół [sic!]: korzeń na górze, na dole liście
- od korzenia do **każdego** liścia prowadzi tylko jedna droga,

Struktura drzewa

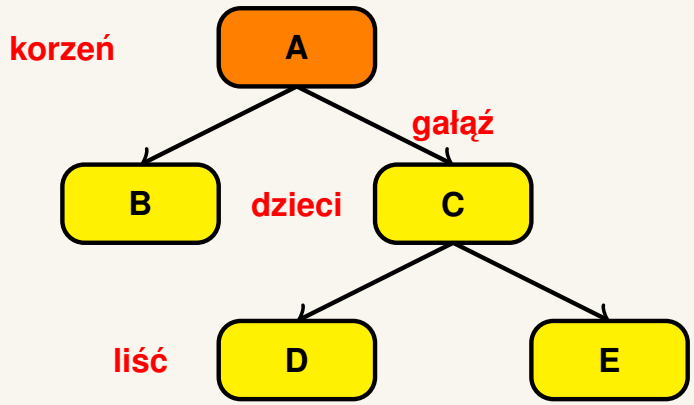


Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,
- węzeł bez dzieci, to **liść**

- konwencja rysowania drzew rosnących od góry w dół [sic!]: korzeń na górze, na dole liście
- od korzenia do **każdego** liścia prowadzi tylko jedna droga,
- w korzeniach jest skupiona cała **próba ucząca**, kolejne elementy PU są przesuwane wzdłuż gałęzi, z góry w dół,

Struktura drzewa

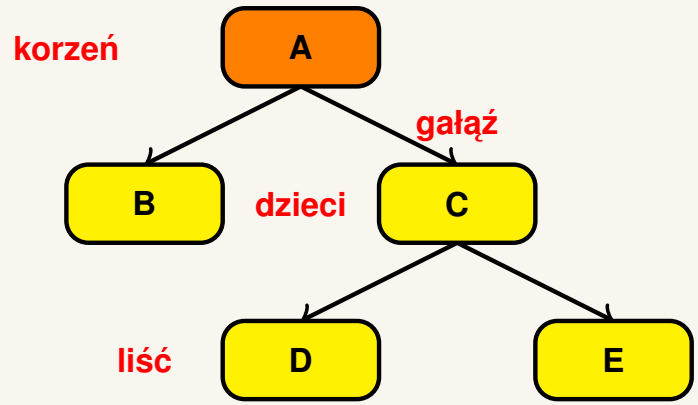


Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,
- węzeł bez dzieci, to **liść**

- konwencja rysowania drzew rosnących od góry w dół [sic!]: korzeń na górze, na dole liście
- od korzenia do **każdego** liścia prowadzi tylko jedna droga,
- w korzeniach jest skupiona cała **próba ucząca**, kolejne elementy PU są przesuwane wzdłuż gałęzi, z góry w dół,
- w każdym węźle jest podejmowana o wyborze gałęzi, wzdłuż której będzie trwać przesuwanie próby

Struktura drzewa

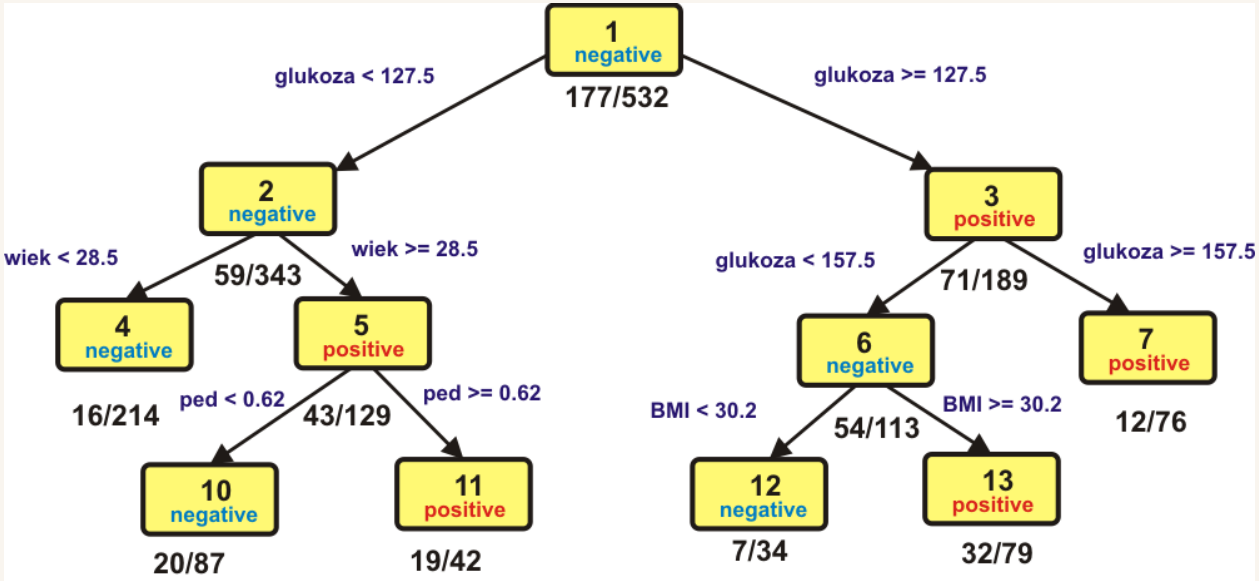


Nazewnictwo

- jeżeli z węzła wychodzą gałęzie do innych węzłów (**dzieci**), to jest to **rodzic** węzła,
- dzieci oraz ich dzieci, to **potomkowie** węzła-rodzica,
- węzeł bez dzieci, to **liść**

- konwencja rysowania drzew rosnących od góry w dół [sic!]: korzeń na górze, na dole liście
- od korzenia do **każdego** liścia prowadzi tylko jedna droga,
- w korzeniach jest skupiona cała **próba ucząca**, kolejne elementy PU są przesuwane wzdłuż gałęzi, z góry w dół,
- w każdym węźle jest podejmowana o wyborze gałęzi, wzdłuż której będzie trwać przesuwanie próby
- czyli w w każdym węźle (oprócz liści) -> podział na podgrupy

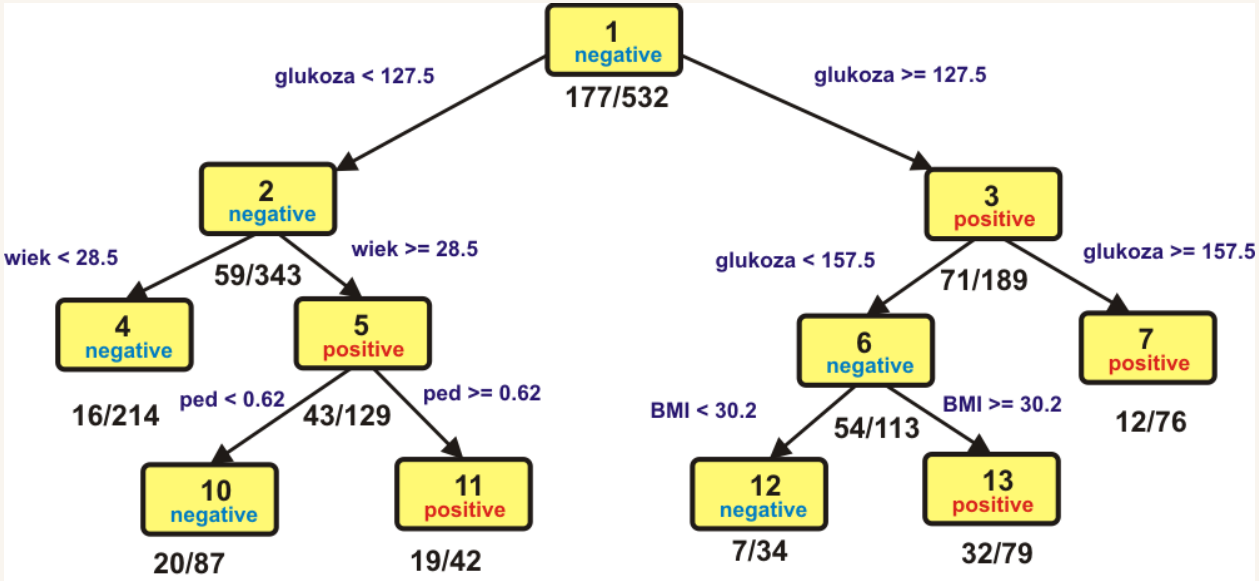
Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

- klasyfikacja **532** Indianek w wieku powyżej 20 lat ze szczepu Pima (Phoenix) na osoby chore na cukrzycę (positive) i zdrowe (negative)

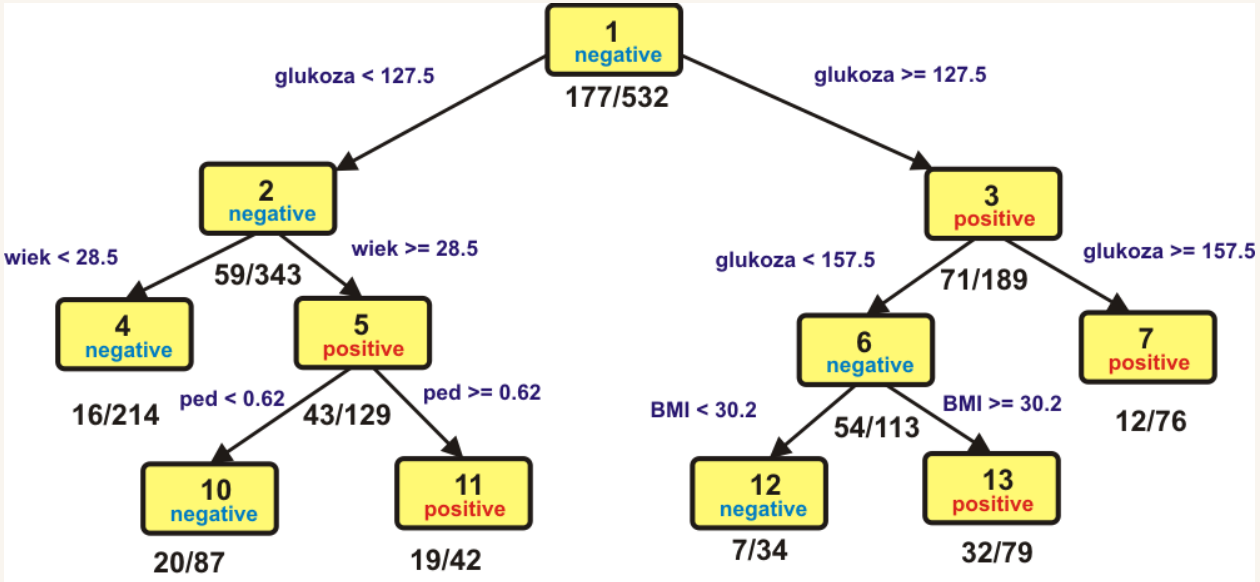
Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

- klasyfikacja **532** Indianek w wieku powyżej 20 lat ze szczepu Pima (Phoenix) na osoby chore na cukrzycę (positive) i zdrowe (negative)
- wektor cech: (1) liczba ciąż, (2) poziom testu glukozowgo, (3) ciśnienie tętnicze, (4) grubość fałdu skóry na tricepsie, (5) BMI, (6) wiek, (7) pedigree

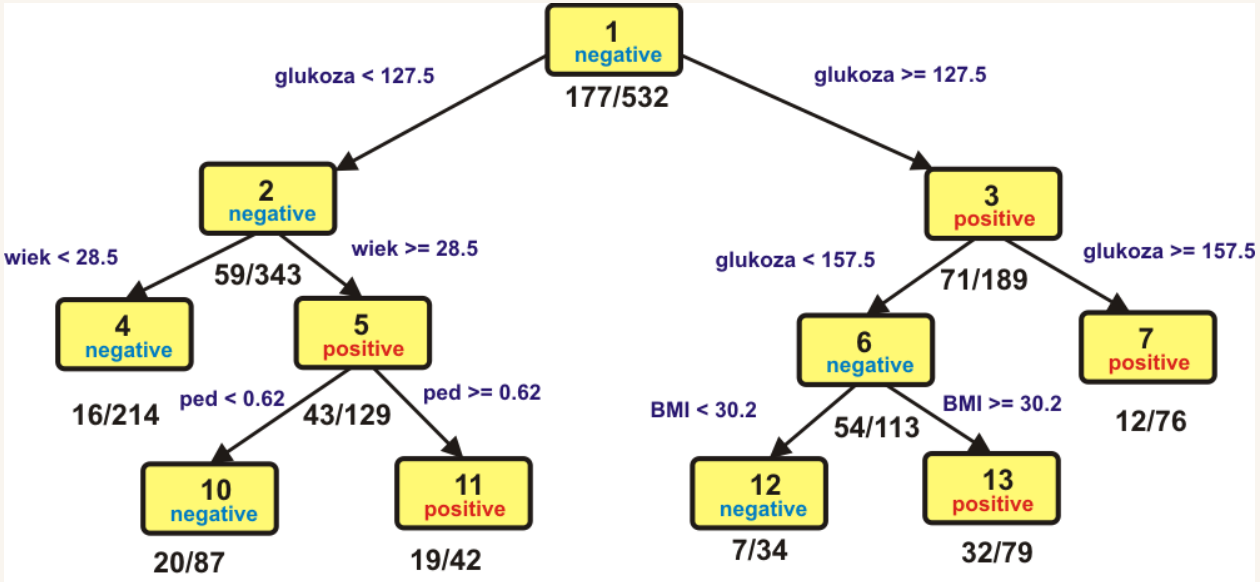
Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

- obok gałęzi $\rightarrow$ warunek podziału (np. liść nr. 4: poziom glukozy $<$ 127.5 i wiek $<$ 28.5 to osoba zdrowa

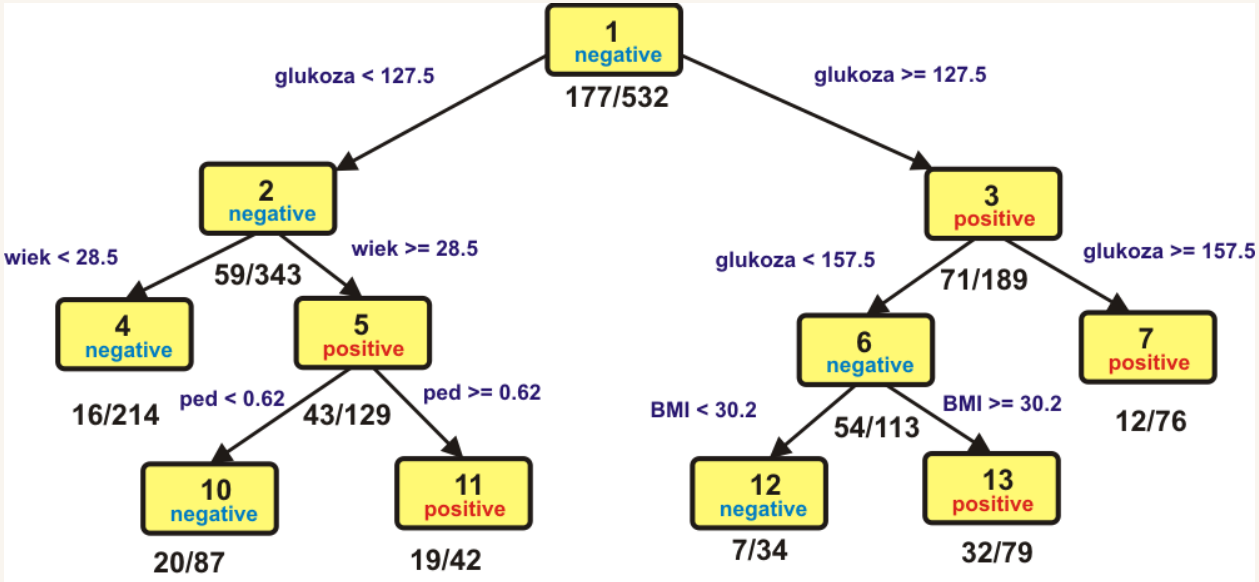
Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

- obok gałęzi $\rightarrow$ warunek podziału (np. liść nr. 4: poziom glukozy $<$ 127.5 i wiek $<$ 28.5 to osoba zdrowa
- w węzłach podana jest klasa większościowa oraz numer węzła

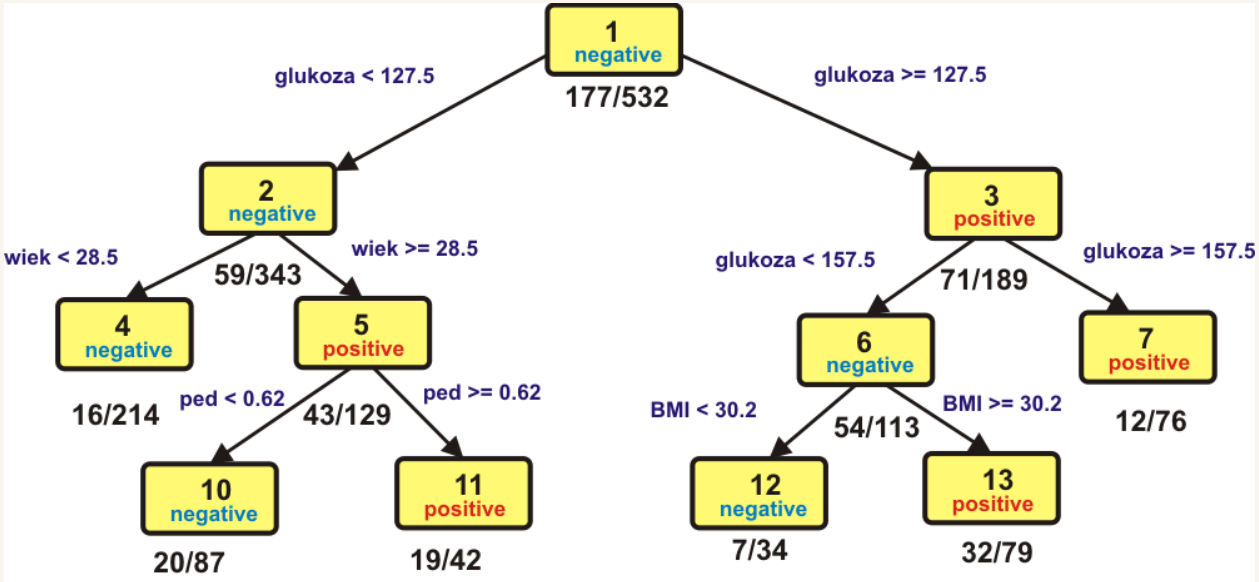
Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

- obok gałęzi $\rightarrow$ warunek podziału (np. liść nr. 4: poziom glukozy $<$ 127.5 i wiek $<$ 28.5 to osoba zdrowa
- w węzłach podana jest klasa większościowa oraz numer węzła
- ułamek błędów podany jest poniżej węzła

Przykładowe drzewo klasyfikacyjne

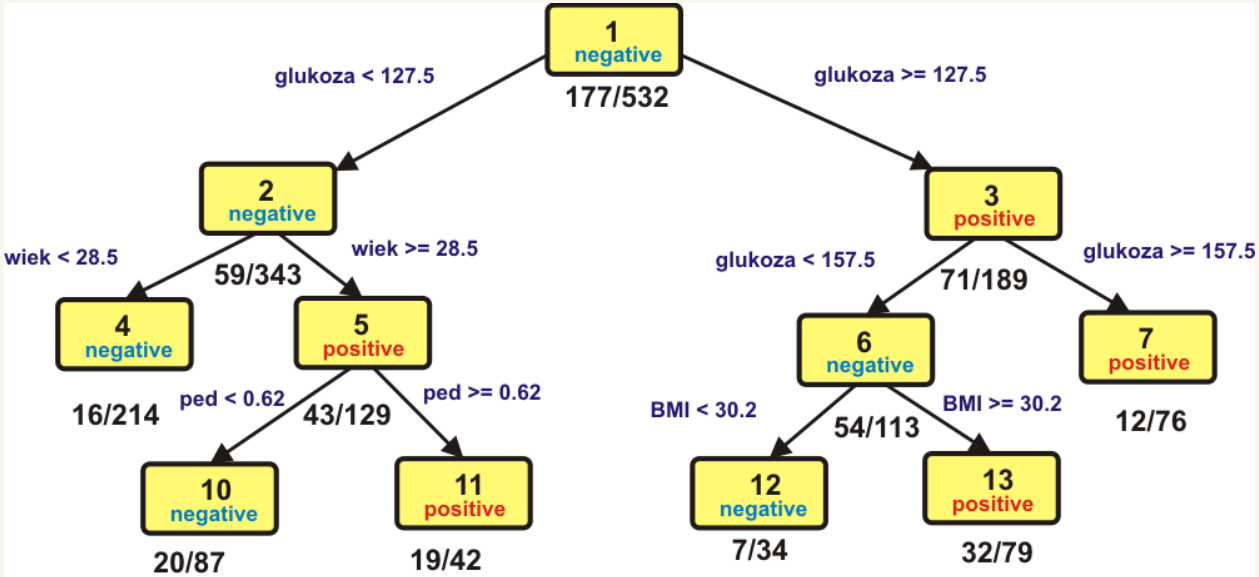


[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

- obok gałęzi $\rightarrow$ warunek podziału (np. liść nr. 4: poziom glukozy $<$ 127.5 i wiek $<$ 28.5 to osoba zdrowa
- w węzłach podana jest klasa większościowa oraz numer węzła
- ułamek błędów podany jest poniżej węzła
- numeracja odpowiada nieskończonemu drzewu binarnemu (1-2-4-8-16...).

Cel drzewa

Przykładowe drzewo klasyfikacyjne

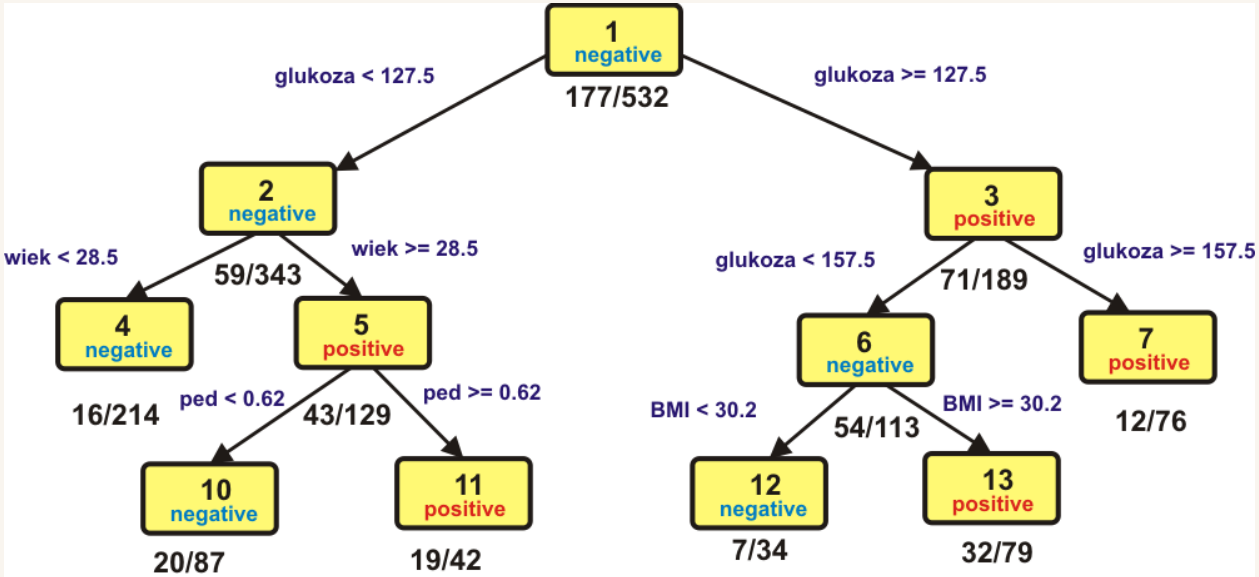


[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

Cel drzewa klasyfikującego

Umożliwienie klasyfikacji obserwacji, o których nie wiemy, do jakich klas należą.

Przykładowe drzewo klasyfikacyjne



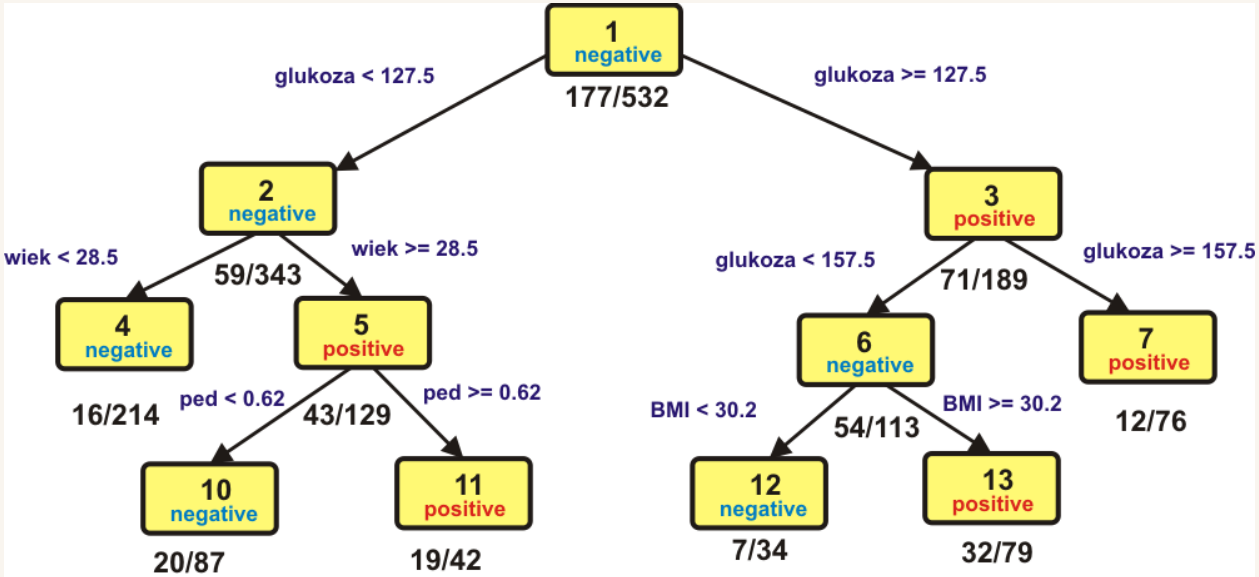
[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

Cel drzewa klasyfikującego

Umożliwienie klasyfikacji obserwacji, o których nie wiemy, do jakich klas należą.

Drzewo jest uczone (trenowane) na podstawie **próby uczącej**:

Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

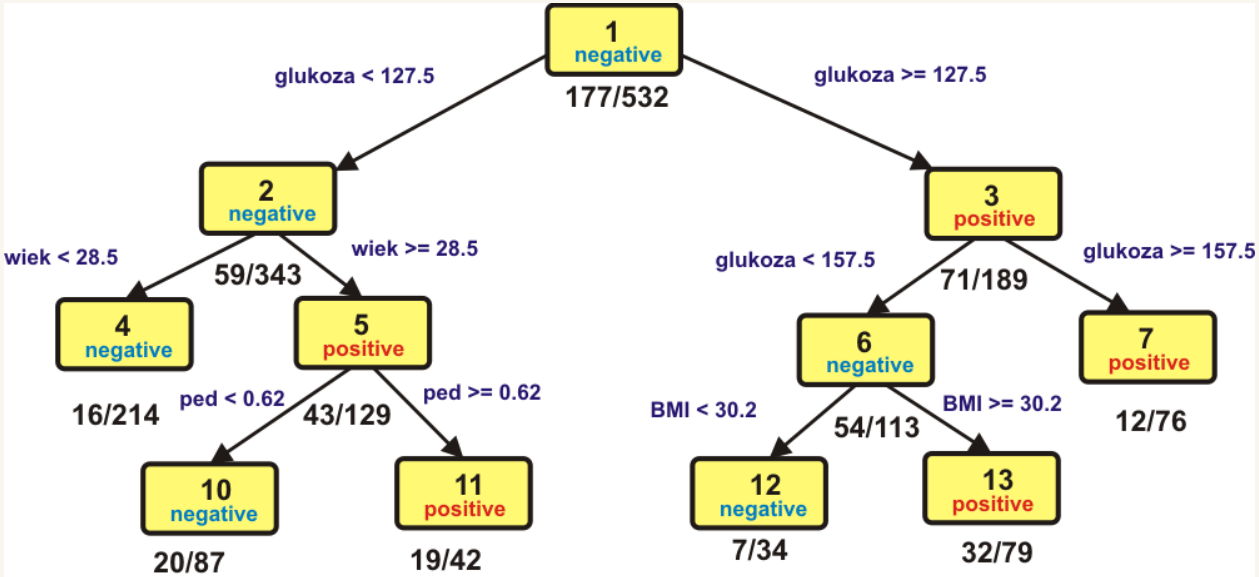
Cel drzewa klasyfikującego

Umożliwienie klasyfikacji obserwacji, o których nie wiemy, do jakich klas należą.
Drzewo jest uczone (trenowane) na podstawie **próby uczącej**:

- od niej zależy postać **warunków podziału**

Cel drzewa

Przykładowe drzewo klasyfikacyjne



[Drzewo zaczerpnięte z Koronacki, Ćwik, *Statystyczne systemy uczące się*, wyd. drugie s. 133]

Cel drzewa klasyfikującego

Umożliwienie klasyfikacji obserwacji, o których nie wiemy, do jakich klas należą.

Drzewo jest uczone (trenowane) na podstawie **próby uczącej**:

- od niej zależy postać **warunków podziału**
- ona determinuje, które węzły określa się jako liście.

Reguły podziału - ogólny zarys

Podział

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

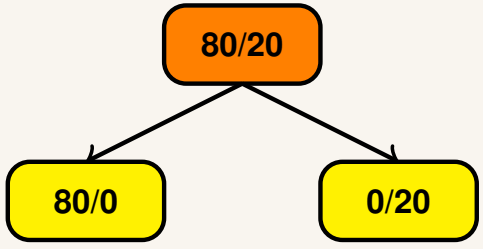
Różnorodność otrzymywanych części jest możliwie największa

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

Różnorodność otrzymywanych części jest możliwie największa

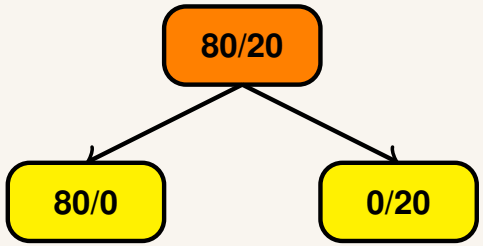


Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

Różnorodność otrzymywanych części jest możliwie największa



Potrzebne jest:

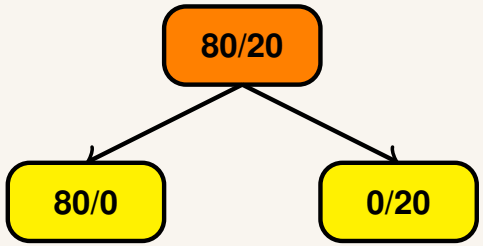
- podanie stosownej miary różnorodności klas

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

Różnorodność otrzymywanych części jest możliwie największa



Potrzebne jest:

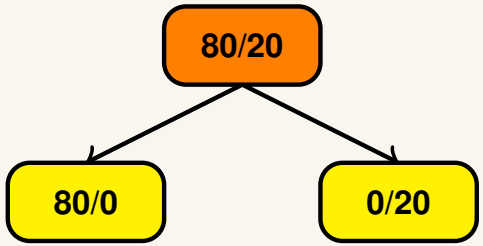
- podanie stosownej miary różnorodności klas
- podanie miary różnicy pomiędzy różnorodnością klas w węźle-rodzicu i węzłach-dzieciach

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

Różnorodność otrzymywanych części jest możliwie największa



Potrzebne jest:

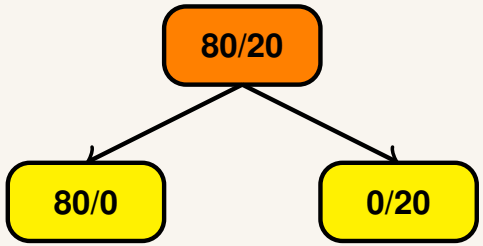
- podanie stosownej miary różnorodności klas
- podanie miary różnicy pomiędzy różnorodnością klas w węźle-rodzicu i węzłach-dzieciach
- podanie algorytmu maksymalizacji różnorodności.

Podział

- dokonywany jest tylko na podstawie tych elementów PU, które znalazły się w danym węźle
- polega na **najlepszym rozdzieleniu** podpróby na 2 części przechodzące do węzłów dzieci

Najlepsze rozdzielenie

Różnorodność otrzymywanych części jest możliwie największa



Potrzebne jest:

- podanie stosownej miary różnorodności klas
- podanie miary różnicy pomiędzy różnorodnością klas w węźle-rodzicu i węzłach-dzieciach
- podanie algorytmu maksymalizacji różnorodności.

UWAGA! jest to podział *lokalnie* najlepszy, ale prowadzi do dobrych *globalnych* wyników

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g
- próba ucząca to pary $(\mathbf{x}_i, y_i)$, $i = 1, ..., n$

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g
- próba ucząca to pary $(\mathbf{x}_i, y_i)$, $i = 1, ..., n$
- rozważamy dowolny, ustalony węzeł m,

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g
- próba ucząca to pary $(\mathbf{x}_i, y_i)$, $i = 1, ..., n$
- rozważamy dowolny, ustalony węzeł m,
- liczność próby, która trafiła do węzła m to n_m, liczba obserwacji z klasy k w węźle m to n_{mk}

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g
- próba ucząca to pary $(\mathbf{x}_i, y_i)$, $i = 1, ..., n$
- rozważamy dowolny, ustalony węzeł m,
- liczność próby, która trafiła do węzła m to n_m, liczba obserwacji z klasy k w węźle m to n_{mk}
- ułamek obserwacji z klasy k w węźle m, $\hat{p}_{mk} = \frac{1}{n_m} \sum_{\mathbf{x}_i} \delta_{y_i,k} = \frac{n_{mk}}{n_m}$ ($\delta_{a,b}$ - delta Kroneckera)

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g
- próba ucząca to pary $(\mathbf{x}_i, y_i)$, $i = 1, ..., n$
- rozważamy dowolny, ustalony węzeł m,
- liczność próby, która trafiła do węzła m to n_m, liczba obserwacji z klasy k w węźle m to n_{mk}
- ułamek obserwacji z klasy k w węźle m, $\hat{p}_{mk} = \frac{1}{n_m} \sum_{\mathbf{x}_i} \delta_{y_i,k} = \frac{n_{mk}}{n_m}$ ($\delta_{a,b}$ - delta Kroneckera)
- obserwacje w węźle m klasyfikowane do klasy k: $k(m) = \arg\max \hat{p}_{mk}$
  - jeżeli m jest liściem $\rightarrow$ ostateczny wyniki klasyfikacji $\mathbf{x}$ przez drzewo,

Sformułowanie problemu

- mamy problem dyskryminacji o g klasach 1,2,...g
- próba ucząca to pary $(\mathbf{x}_i, y_i)$, $i = 1, ..., n$
- rozważamy dowolny, ustalony węzeł m,
- liczność próby, która trafiła do węzła m to n_m, liczba obserwacji z klasy k w węźle m to n_{mk}
- ułamek obserwacji z klasy k w węźle m, $\hat{p}_{mk} = \frac{1}{n_m} \sum_{\mathbf{x}_i} \delta_{y_i,k} = \frac{n_{mk}}{n_m}$ ($\delta_{a,b}$ - delta Kroneckera)
- obserwacje w węźle m klasyfikowane do klasy k: $k(m) = \arg\max \hat{p}_{mk}$
  - jeżeli m jest liściem $\rightarrow$ ostateczny wyniki klasyfikacji $\mathbf{x}$ przez drzewo,
  - w przeciwnym razie $\rightarrow$ $k(m)$ to informacja, która klasa jest najliczniej reprezentowana.

Miara różnorodności

Miara różnorodności

Sensowna miara różnorodnosci to taka, która przyjmuje

Miara różnorodności

Sensowna miara różnorodnosci to taka, która przyjmuje

- wartośc 0, gdy wszystkie obserwacje należą do tej samej klasy,

Miara różnorodności

Sensowna miara różnorodnosci to taka, która przyjmuje

- wartośc 0, gdy wszystkie obserwacje należą do tej samej klasy,
- wartość maksymalną, gdy mamy do czynienia z rozkładem jednostajnym $\hat{p}_{m1} = ... = \hat{p}_{mg} = 1/g$

Przykłady miar różnorodności $Q_n(T)$

Miara różnorodności

Sensowna miara różnorodnosci to taka, która przyjmuje

- wartośc 0, gdy wszystkie obserwacje należą do tej samej klasy,
- wartość maksymalną, gdy mamy do czynienia z rozkładem jednostajnym $\hat{p}_{m1} = ... = \hat{p}_{mg} = 1/g$

Przykłady miar różnorodności $Q_n(T)$

1. ułamek błednych klasyfikacji:
   $\frac{1}{n_m} \sum_i (1 - \delta_{y_i, k(m)}) = 1 - \hat{p}_{mk(m)}$

Miara różnorodności

Sensowna miara różnorodnosci to taka, która przyjmuje

- wartośc 0, gdy wszystkie obserwacje należą do tej samej klasy,
- wartość maksymalną, gdy mamy do czynienia z rozkładem jednostajnym $\hat{p}_{m1} = ... = \hat{p}_{mg} = 1/g$

Przykłady miar różnorodności $Q_n(T)$

1. ułamek błednych klasyfikacji:
   $\frac{1}{n_m} \sum_i (1 - \delta_{y_i, k(m)}) = 1 - \hat{p}_{mk(m)}$
2. wskaźnik (indeks) Giniego (oszacowanie ułamka błednych klasyfikacji, gdy obserwacje są klasyfikowane do klasy k z prawdpodobieństwem $\hat{p}_{mk}$):
   $\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1}^{g} \hat{p}_{mk}(1 - \hat{p}_{mk})$

Miara różnorodności

Sensowna miara różnorodnosci to taka, która przyjmuje

- wartośc 0, gdy wszystkie obserwacje należą do tej samej klasy,
- wartość maksymalną, gdy mamy do czynienia z rozkładem jednostajnym $\hat{p}_{m1} = ... = \hat{p}_{mg} = 1/g$

Przykłady miar różnorodności $Q_n(T)$

1. ułamek błednych klasyfikacji:
   $\frac{1}{n_m} \sum_i (1 - \delta_{y_i, k(m)}) = 1 - \hat{p}_{mk(m)}$
2. wskaźnik (indeks) Giniego (oszacowanie ułamka błednych klasyfikacji, gdy obserwacje są klasyfikowane do klasy k z prawdpodobieństwem $\hat{p}_{mk}$):
   $\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1}^{g} \hat{p}_{mk}(1 - \hat{p}_{mk})$
3. entropia:
   $-\sum_{k=1}^{g} \hat{p}_{mk} \log \hat{p}_{mk}$

Przypadek $g = 2$

Przypadek $g = 2$

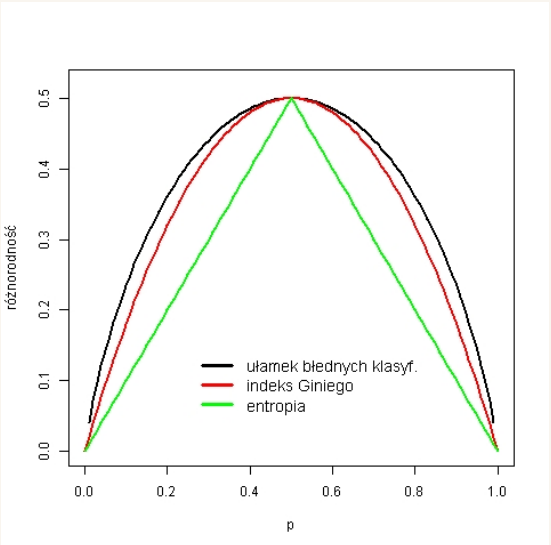
$$Q_m(T) = \begin{cases} 1 - \max(1, 1-p) \\ 2p(1-p) \\ p \log p - (1-p)\log(1-p) \end{cases}$$

p - ułamek (prawdopodobieństwo) przynależności do klasy 2

Przypadek $g = 2$

$$Q_m(T) = \begin{cases} 1 - \max(1, 1-p) \\ 2p(1-p) \\ p \log p - (1-p)\log(1-p) \end{cases}$$

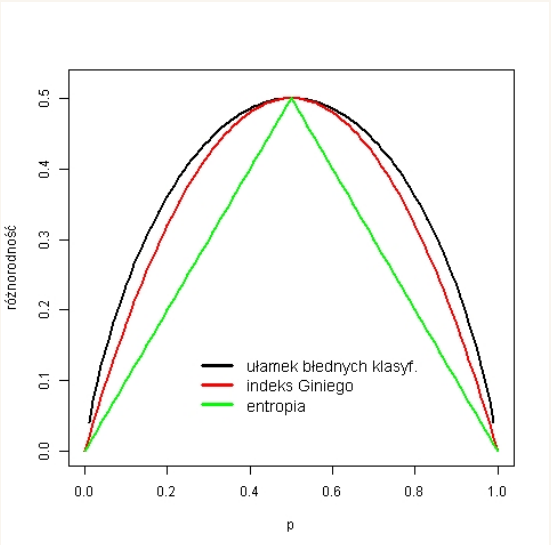
p - ułamek (prawdopodobieństwo) przynależności do klasy 2

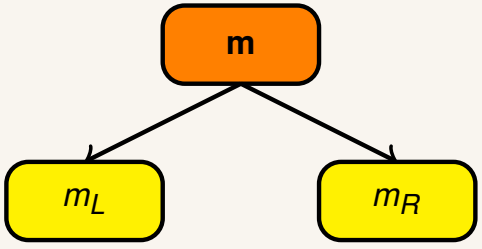


Przypadek $g = 2$

$$Q_m(T) = \begin{cases} 1 - \max(1, 1-p) \\ 2p(1-p) \\ p \log p - (1-p)\log(1-p) \end{cases}$$

p - ułamek (prawdopodobieństwo) przynależności do klasy 2

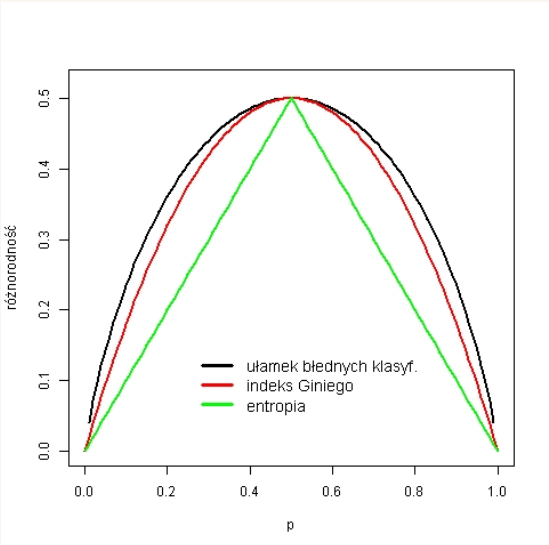


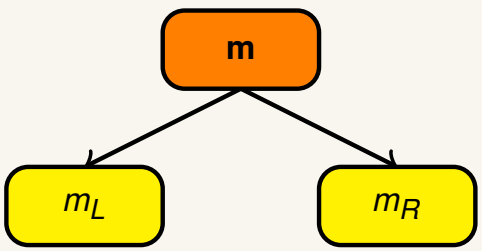


Przypadek $g = 2$

$$Q_m(T) = \begin{cases} 1 - \max(1, 1-p) \\ 2p(1-p) \\ p \log p - (1-p)\log(1-p) \end{cases}$$

p - ułamek (prawdopodobieństwo) przynależności do klasy 2



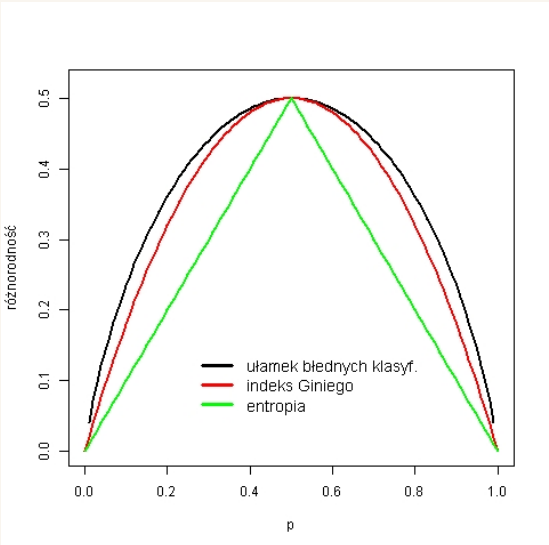


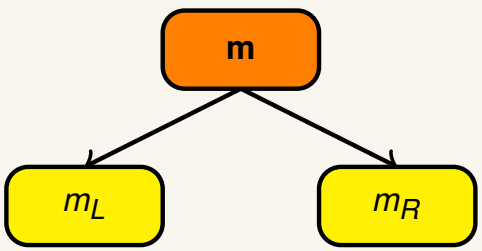
$\hat{p}_L$ - ułamek elementów PU, które z węzła m przeszły do m_L, $\hat{p}_L = \frac{n_{mL}}{n_m}$
$\hat{p}_R = 1 - \hat{p}_L$ - ułamek elementów PU, które z węzła m przeszły do m_R

Przypadek $g = 2$

$$Q_m(T) = \begin{cases} 1 - \max(1, 1-p) \\ 2p(1-p) \\ p \log p - (1-p)\log(1-p) \end{cases}$$

p - ułamek (prawdopodobieństwo) przynależności do klasy 2





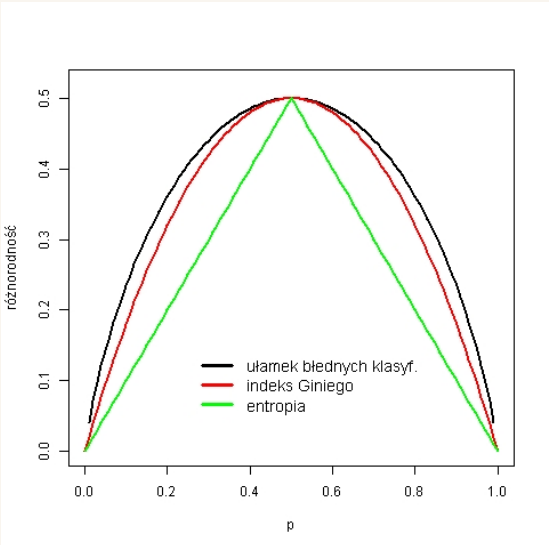
$\hat{p}_L$ - ułamek elementów PU, które z węzła m przeszły do m_L, $\hat{p}_L = \frac{n_{mL}}{n_m}$
$\hat{p}_R = 1 - \hat{p}_L$ - ułamek elementów PU, które z węzła m przeszły do m_R

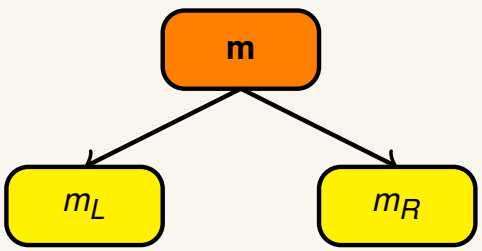
$Q_{m_L,m_R} = \hat{p}_L Q_{m_L} + \hat{p}_R Q_{m_R}$ - łączna miara różnorodności w dzieciach

Przypadek $g = 2$

$$Q_m(T) = \begin{cases} 1 - \max(1, 1-p) \\ 2p(1-p) \\ p \log p - (1-p)\log(1-p) \end{cases}$$

p - ułamek (prawdopodobieństwo) przynależności do klasy 2

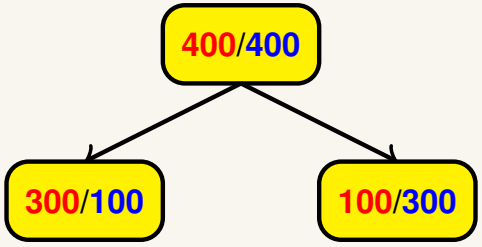




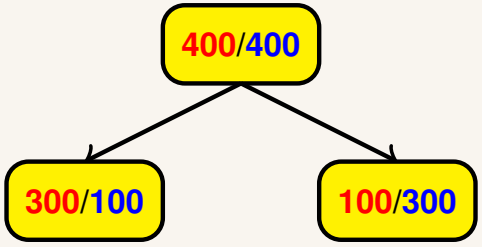
$\hat{p}_L$ - ułamek elementów PU, które z węzła m przeszły do m_L, $\hat{p}_L = \frac{n_{mL}}{n_m}$
$\hat{p}_R = 1 - \hat{p}_L$ - ułamek elementów PU, które z węzła m przeszły do m_R

$Q_{m_L,m_R} = \hat{p}_L Q_{m_L} + \hat{p}_R Q_{m_R}$ - łączna miara różnorodności w dzieciach
$\Delta Q_{m,m_L,m_R} = Q_m - Q_{m_L,m_R}$ - różnica między różnorodności klas

Przykład 1

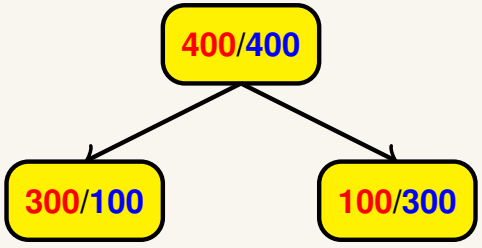


Przykład 1



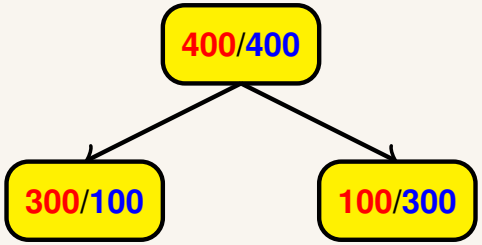
- Ułamek błędnych klasyfikacji

Przykład 1



- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$

Przykład 1

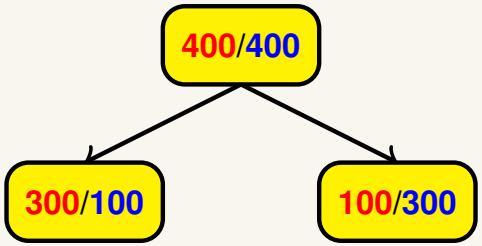


- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$

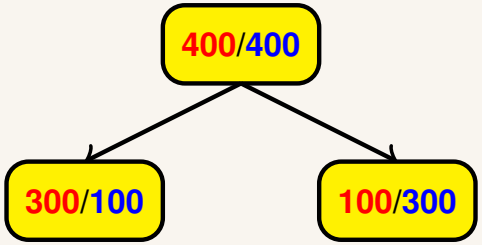
Przykład 1



- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$

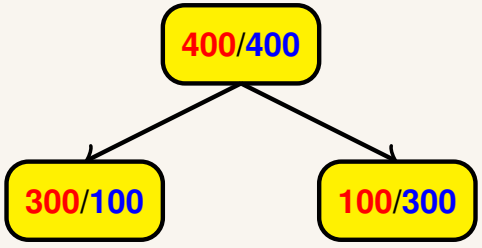
Przykład 1



- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

Przykład 1

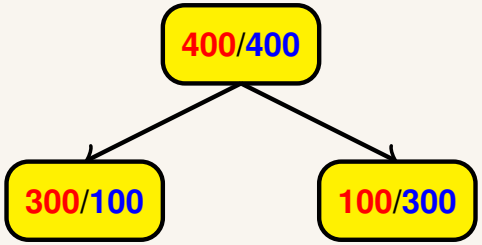


- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

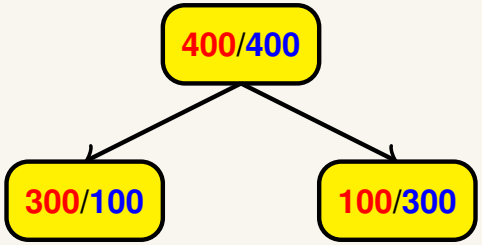
- Indeks Giniego

Przykład 1



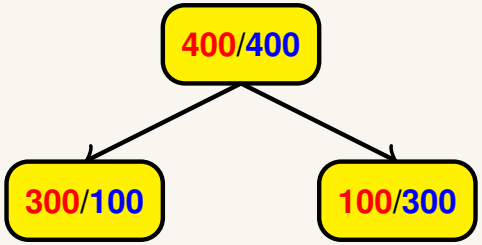
- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$
  $Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$

Przykład 1



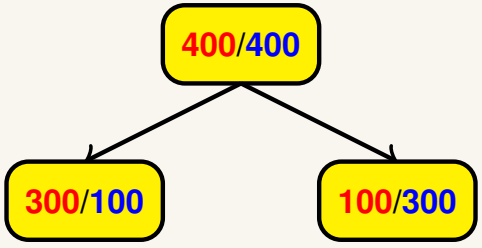
- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$
  $Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$

Przykład 1



- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$
  $Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$
  $Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$

Przykład 1



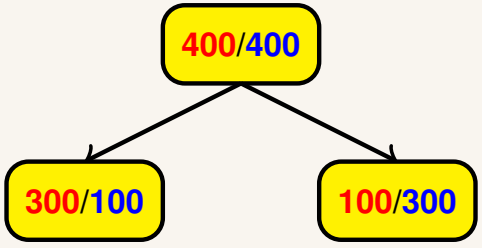
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = 1 - \tfrac{1}{2} = \tfrac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\tfrac{1}{4}, \tfrac{3}{4}\right) = 1 - \tfrac{3}{4} = \tfrac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\tfrac{3}{4}, \tfrac{1}{4}\right) = 1 - \tfrac{3}{2} = \tfrac{1}{4}$$
$$\Delta Q^f = \tfrac{1}{2} - \left(\tfrac{1}{2} \cdot \tfrac{1}{4} + \tfrac{1}{2} \cdot \tfrac{1}{4}\right) = \tfrac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\tfrac{1}{2}\tfrac{1}{2} = \tfrac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \tfrac{1}{4} \cdot \tfrac{3}{4} = \tfrac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \tfrac{3}{4} \cdot \tfrac{1}{4} = \tfrac{3}{8}$$
$$\Delta Q^g = \tfrac{1}{2} - \left(\tfrac{1}{2} \cdot \tfrac{3}{8} + \tfrac{1}{2} \cdot \tfrac{3}{8}\right) = \tfrac{1}{2} - \tfrac{3}{8} = \tfrac{1}{8}$$

Przykład 1



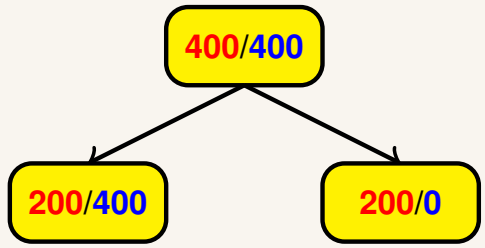
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

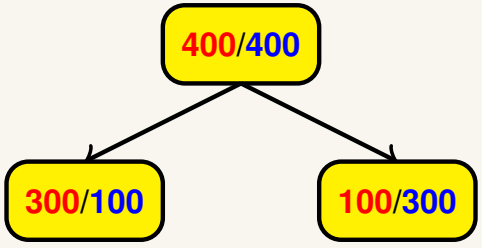
- Indeks Giniego

$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

Przykład 2



Przykład 1



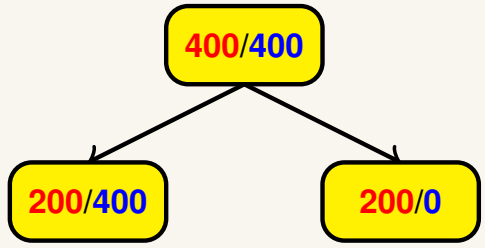
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

- Indeks Giniego

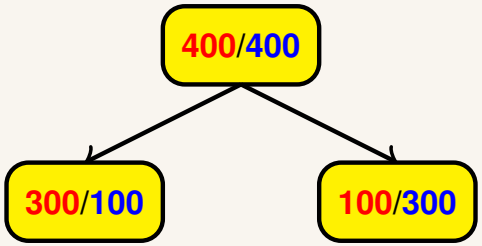
$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

Przykład 2



- Ułamek błędnych klasyfikacji

Przykład 1



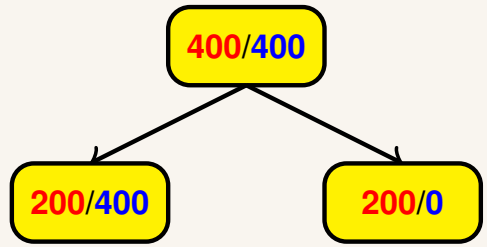
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

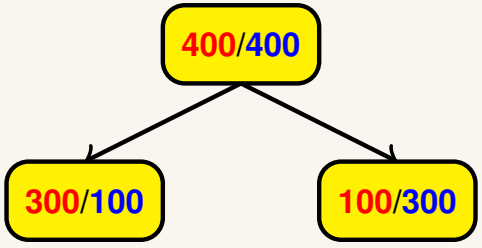
Przykład 2



- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$

Przykład 1



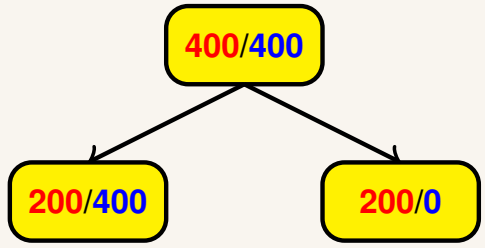
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = 1 - \tfrac{1}{2} = \tfrac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\tfrac{1}{4}, \tfrac{3}{4}\right) = 1 - \tfrac{3}{4} = \tfrac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\tfrac{3}{4}, \tfrac{1}{4}\right) = 1 - \tfrac{3}{2} = \tfrac{1}{4}$$
$$\Delta Q^f = \tfrac{1}{2} - \left(\tfrac{1}{2} \cdot \tfrac{1}{4} + \tfrac{1}{2} \cdot \tfrac{1}{4}\right) = \tfrac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\tfrac{1}{2}\tfrac{1}{2} = \tfrac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \tfrac{1}{4} \cdot \tfrac{3}{4} = \tfrac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \tfrac{3}{4} \cdot \tfrac{1}{4} = \tfrac{3}{8}$$
$$\Delta Q^g = \tfrac{1}{2} - \left(\tfrac{1}{2} \cdot \tfrac{3}{8} + \tfrac{1}{2} \cdot \tfrac{3}{8}\right) = \tfrac{1}{2} - \tfrac{3}{8} = \tfrac{1}{8}$$

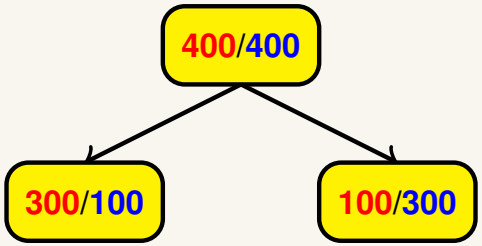
Przykład 2



- Ułamek błędnych klasyfikacji

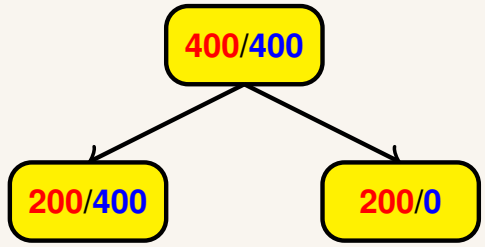
$$Q_m^f = 1 - \max\left(\tfrac{1}{2}, \tfrac{1}{2}\right) = 1 - \tfrac{1}{2} = \tfrac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\tfrac{1}{3}, \tfrac{2}{3}\right) = 1 - \tfrac{2}{3} = \tfrac{1}{3}$$

Przykład 1



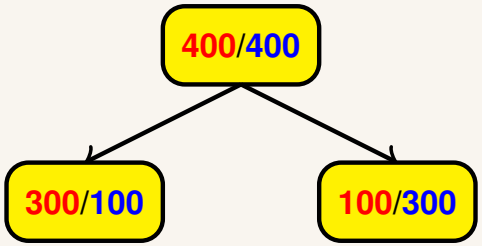
- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$
  $Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$
  $Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$
  $\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$

Przykład 2



- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$
  $Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$

Przykład 1



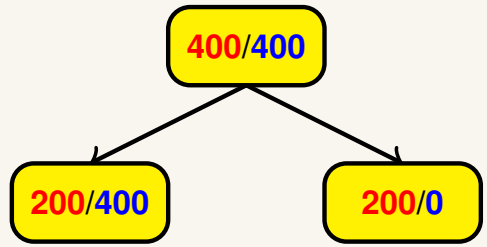
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

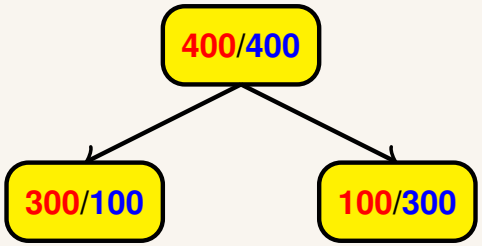
Przykład 2



- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$$
$$Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot 0\right) = \frac{1}{4}$$

Przykład 1



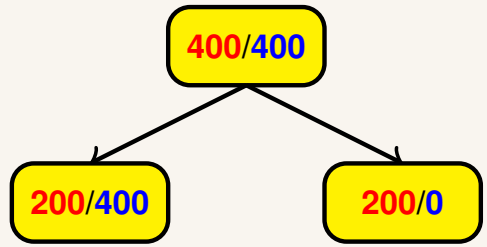
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

Przykład 2

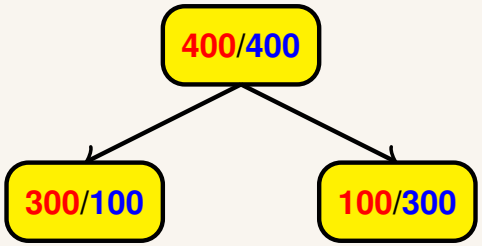


- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$$
$$Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot 0\right) = \frac{1}{4}$$

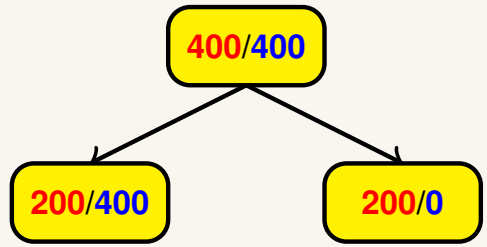
- Indeks Giniego

Przykład 1



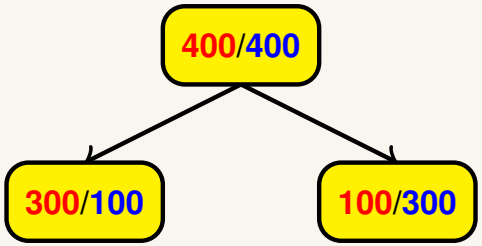
- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$
  $Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$
  $Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$
  $\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$

Przykład 2



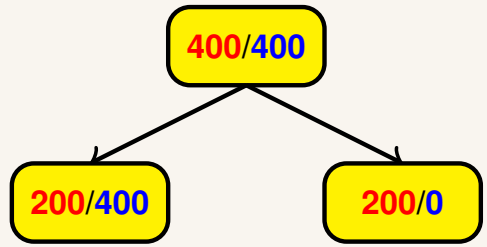
- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$
  $Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot 0\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$

Przykład 1



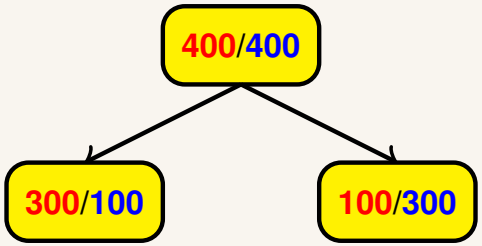
- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$
  $Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$
  $Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$
  $\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$

Przykład 2



- Ułamek błędnych klasyfikacji
  $Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$
  $Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$
  $\Delta Q^f = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot 0\right) = \frac{1}{4}$
- Indeks Giniego
  $Q_m^g = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$
  $Q_{mL}^g = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$

Przykład 1



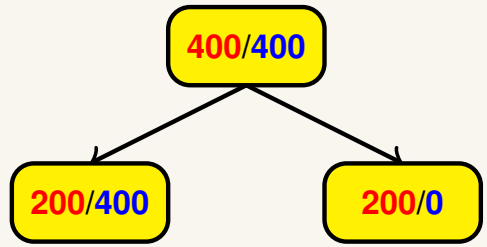
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

Przykład 2



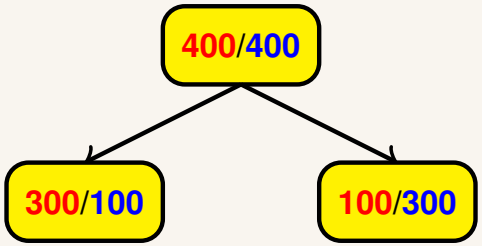
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$$
$$Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot 0\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$$
$$Q_{mR}^g = 2 \cdot 0 \cdot 1 = 0$$

Przykład 1



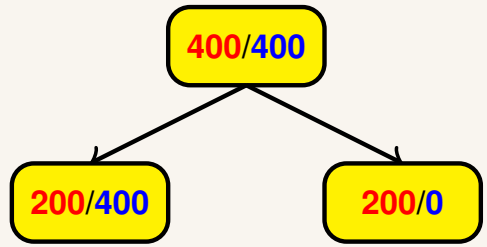
- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{4}, \frac{3}{4}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$
$$Q_{mR}^f = 1 - \max\left(\frac{3}{4}, \frac{1}{4}\right) = 1 - \frac{3}{2} = \frac{1}{4}$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{4}\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2\frac{1}{2}\frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{8}$$
$$Q_{mR}^g = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{8}$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{1}{2} \cdot \frac{3}{8} + \frac{1}{2} \cdot \frac{3}{8}\right) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8}$$

Przykład 2



- Ułamek błędnych klasyfikacji

$$Q_m^f = 1 - \max\left(\frac{1}{2}, \frac{1}{2}\right) = 1 - \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^f = 1 - \max\left(\frac{1}{3}, \frac{2}{3}\right) = 1 - \frac{2}{3} = \frac{1}{3}$$
$$Q_{mR}^f = 1 - \max(0, 1) = 1 - 1 = 0$$
$$\Delta Q^f = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot 0\right) = \frac{1}{4}$$

- Indeks Giniego

$$Q_m^g = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$
$$Q_{mL}^g = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9}$$
$$Q_{mR}^g = 2 \cdot 0 \cdot 1 = 0$$
$$\Delta Q^g = \frac{1}{2} - \left(\frac{3}{4} \cdot \frac{4}{9} + \frac{1}{4} \cdot 0\right) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

Generalnie: wskaźnik Giniego i entropia są **bardziej czułe** na zmiany klas rozkładów!

Algortym maksymalizacji ΔQ

Algortym maksymalizacji ΔQ

- jest to algorytm poszukiwania najlepszego podziału,

Algortym maksymalizacji ΔQ

- jest to algorytm poszukiwania najlepszego podziału,
- dopuszcza wszystkie podziały na wszystkich atrybutach

Algortym maksymalizacji ΔQ

- jest to algorytm poszukiwania najlepszego podziału,
- dopuszcza wszystkie podziały na wszystkich atrybutach
  Dla ustalonego atrybutu (L wartości) mamy $\frac{1}{2}2^L - 1 = 2^{L-1} - 1$ podziałów (ogólnie możliwe jest dokładnie 2^L podziałów, ale nie interesuje nas zbiór pusty oraz ważny jest sam podział)

Algortym maksymalizacji ΔQ

- jest to algorytm poszukiwania najlepszego podziału,
- dopuszcza wszystkie podziały na wszystkich atrybutach
  Dla ustalonego atrybutu (L wartości) mamy $\frac{1}{2}2^L - 1 = 2^{L-1} - 1$ podziałów (ogólnie możliwe jest dokładnie 2^L podziałów, ale nie interesuje nas zbiór pusty oraz ważny jest sam podział)
- czyli trzeba dokonać $2^{L-1} - 1$ podziałów i wybrać taki, który maksymalizuje kryterium,

Algortym maksymalizacji ΔQ

- jest to algorytm poszukiwania najlepszego podziału,
- dopuszcza wszystkie podziały na wszystkich atrybutach Dla ustalonego atrybutu (L wartości) mamy $\frac{1}{2}2^L - 1 = 2^{L-1} - 1$ podziałów (ogólnie możliwe jest dokładnie 2^L podziałów, ale nie interesuje nas zbiór pusty oraz ważny jest sam podział)
- czyli trzeba dokonać $2^{L-1} - 1$ podziałów i wybrać taki, który maksymalizuje kryterium,
- wybór najlepszego podziału danego atrybutu należy powtórzyć dla wszystkich atrybutów i wybrać najlepszy podział na najlepszym atrybucie
- mamy więc kłopot obliczeniowy (np. dla $L = 50$ trzeba sprawdzić $\mathbf{2^{49} - 1}$ **podziałów**!

Składowe o atrybutach liczbowych i porządkowych

Składowe o atrybutach liczbowych i porządkowych

- jednak w przypadku argumentu liczbowego (nawet ciągłego — ma on zawsze skończoną liczbę wartości) oraz porządkowego ograniczamy się do **podziałów monotonicznych**: $x^{(l)} \leq c$ (albo $x^{(l)} < c$), gdzie c to jakaś zaobserwowana w danych wartość atrybutu,

Składowe o atrybutach liczbowych i porządkowych

- jednak w przypadku argumentu liczbowego (nawet ciągłego — ma on zawsze skończoną liczbę wartości) oraz porządkowego ograniczamy się do **podziałów monotonicznych**: $x^{(l)} \leq c$ (albo $x^{(l)} < c$), gdzie c to jakaś zaobserwowana w danych wartość atrybutu,
- dzięki temu liczba możliwych podziałów spada do $L - 1$.

Składowe o atrybutach liczbowych i porządkowych

- jednak w przypadku argumentu liczbowego (nawet ciągłego — ma on zawsze skończoną liczbę wartości) oraz porządkowego ograniczamy się do **podziałów monotonicznych**: $x^{(l)} \leq c$ (albo $x^{(l)} < c$), gdzie c to jakaś zaobserwowana w danych wartość atrybutu,
- dzięki temu liczba możliwych podziałów spada do $L - 1$.

A co ze składowymi o atrybutach nominalnych?

Składowe o atrybutach liczbowych i porządkowych

- jednak w przypadku argumentu liczbowego (nawet ciągłego — ma on zawsze skończoną liczbę wartości) oraz porządkowego ograniczamy się do **podziałów monotonicznych**: $x^{(l)} \leq c$ (albo $x^{(l)} < c$), gdzie c to jakaś zaobserwowana w danych wartość atrybutu,
- dzięki temu liczba możliwych podziałów spada do $L - 1$.

A co ze składowymi o atrybutach nominalnych?

Niech $g = 2$, a atrybut ma L poziomów. Załóżmy, że poziomy składowej $x^{(i)}$ zostały ułożone wg. rosnących wartości prawdopodobieństw $p(1|x^{(i)})$

$$p(1|x^{(1)}) \leq p(1|x^{(2)}) \leq ... \leq p(1|x^{(L)})$$

Wówczas jeden z $L - 1$ podziałów typu

$$\{x^{(1)}, ..., x^{(l)}\}, \{x^{(1+1)}, ..., x^{(L)}\}$$

maksymalizuje ΔQ^g i ΔQ^e.

Jak długo należy budować drzewo?

Jak długo należy budować drzewo?

- reguły **tworzenia** drzewa mamy, ale skąd wiemy, kiedy należy **zakończyć** konstrukcję?

Jak długo należy budować drzewo?

- reguły **tworzenia** drzewa mamy, ale skąd wiemy, kiedy należy **zakończyć** konstrukcję?
- czy budowę kontynuujemy tak długo, jak to jest możliwe, czyli do otrzymania liści, w których będą obserwacje tylko z jednej klasy?

Jak długo należy budować drzewo?

- reguły **tworzenia** drzewa mamy, ale skąd wiemy, kiedy należy **zakończyć** konstrukcję?
- czy budowę kontynuujemy tak długo, jak to jest możliwe, czyli do otrzymania liści, w których będą obserwacje tylko z jednej klasy?
- **raczej nie** $\rightarrow$ takie drzewo będzie nadmiernie dopasowane do próby uczącej, czyli **przeuczone** (**przetrenowane**)

Jak długo należy budować drzewo?

- reguły **tworzenia** drzewa mamy, ale skąd wiemy, kiedy należy **zakończyć** konstrukcję?
- czy budowę kontynuujemy tak długo, jak to jest możliwe, czyli do otrzymania liści, w których będą obserwacje tylko z jednej klasy?
- **raczej nie** $\rightarrow$ takie drzewo będzie nadmiernie dopasowane do próby uczącej, czyli **przeuczone** (**przetrenowane**)
- taka konstrukcja ma sens tylko i wyłącznie w przypadku zupełnie determinstycznym (brak losowości w danych)

Etapowa konstrukcja drzewa

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.
  - w liściach są elementy tylko jednej klasy,

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.
  - w liściach są elementy tylko jednej klasy,
  - w liściach są wektory obserwacji o tej samej wartości choć różnej przynależności do klas,

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.
  - w liściach są elementy tylko jednej klasy,
  - w liściach są wektory obserwacji o tej samej wartości choć różnej przynależności do klas,
  - uznanie z góry za liść węzła, do którego dotrało nie więcej niż 5 elemntów PU,

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.
  - w liściach są elementy tylko jednej klasy,
  - w liściach są wektory obserwacji o tej samej wartości choć różnej przynależności do klas,
  - uznanie z góry za liść węzła, do którego dotrało nie więcej niż 5 elemntów PU,
  - można też ograniczyć maksymalną długość drogi od korzenia do liścia

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.
  - w liściach są elementy tylko jednej klasy,
  - w liściach są wektory obserwacji o tej samej wartości choć różnej przynależności do klas,
  - uznanie z góry za liść węzła, do którego dotrało nie więcej niż 5 elemntów PU,
  - można też ograniczyć maksymalną długość drogi od korzenia do liścia
- sprawdzamy zdolność klasyfikacyjną drzewa na próbie walidacyjnej lub testowej

Etapowa konstrukcja drzewa

- najpierw rozbudowujemy drzewo tak długo, jak to jest możliwe albo do do spełnienia naturalnej reguły zatrzymania budowy np.
  - w liściach są elementy tylko jednej klasy,
  - w liściach są wektory obserwacji o tej samej wartości choć różnej przynależności do klas,
  - uznanie z góry za liść węzła, do którego dotrało nie więcej niż 5 elemntów PU,
  - można też ograniczyć maksymalną długość drogi od korzenia do liścia
- sprawdzamy zdolność klasyfikacyjną drzewa na próbie walidacyjnej lub testowej
- przeprowadzamy systematyczne *przycinanie* drzewa (odcinanie końcowych gałęzi, potem rodziców) tak, aby otrzymać największą możliwą zdolność klasyfikacyjną

Algorytm przycinania

Algorytm przycinania

- oznaczmy pełne drzewo jako T_0, a jego rozmiar (liczbę liści) jako $|T_0|$,

Algorytm przycinania

- oznaczmy pełne drzewo jako T_0, a jego rozmiar (liczbę liści) jako $|T_0|$,
- zdolność klasyfikacyjną mierzymy jako ułamek błednych klasyfikacji R (na próbie walidacyjnej lub testowej) i obliczamy ją dla pełnego drzewa $R(T_0)$,

Algorytm przycinania

- oznaczmy pełne drzewo jako T_0, a jego rozmiar (liczbę liści) jako $|T_0|$,
- zdolność klasyfikacyjną mierzymy jako ułamek błednych klasyfikacji R (na próbie walidacyjnej lub testowej) i obliczamy ją dla pełnego drzewa $R(T_0)$,
- następnie znajdujemy poddrzewo zakorzenione T_1 drzewa T_0 o $|T_0| - 1$ liściach i minimalnej wartości błednych klasyfikacji $R(T_1)$ (wśród wszystkich drzew o $|T_0| - 1$ lisciach),

Algorytm przycinania

- oznaczmy pełne drzewo jako T_0, a jego rozmiar (liczbę liści) jako $|T_0|$,
- zdolność klasyfikacyjną mierzymy jako ułamek błednych klasyfikacji R (na próbie walidacyjnej lub testowej) i obliczamy ją dla pełnego drzewa $R(T_0)$,
- następnie znajdujemy poddrzewo zakorzenione T_1 drzewa T_0 o $|T_0| - 1$ liściach i minimalnej wartości błednych klasyfikacji $R(T_1)$ (wśród wszystkich drzew o $|T_0| - 1$ lisciach),
- potem szukamy tego samego dla drzewa T_2 o $|T_0| - 2$ liściach itd, aż do uzyskania minimum globalnego wśród drzew wszystkich możliwych rozmiarów,

Algorytm przycinania

- oznaczmy pełne drzewo jako T_0, a jego rozmiar (liczbę liści) jako $|T_0|$,
- zdolność klasyfikacyjną mierzymy jako ułamek błednych klasyfikacji R (na próbie walidacyjnej lub testowej) i obliczamy ją dla pełnego drzewa $R(T_0)$,
- następnie znajdujemy poddrzewo zakorzenione T_1 drzewa T_0 o $|T_0| - 1$ liściach i minimalnej wartości błednych klasyfikacji $R(T_1)$ (wśród wszystkich drzew o $|T_0| - 1$ lisciach),
- potem szukamy tego samego dla drzewa T_2 o $|T_0| - 2$ liściach itd, aż do uzyskania minimum globalnego wśród drzew wszystkich możliwych rozmiarów,
- wadą algorytmu jest to, że kolejne podrzewa **nie muszą** tworzyć rodziny zagnieżdżonej

Algorytm kosztu-złożoności

Algorytm kosztu-złożoności

- kompromis pomiędzy kosztem dokonania błędnej klasyfikacji (R) i kosztem wynikającym z konieczności zbudowania drzewa (proporcjonalnym do liczy liści),

Algorytm kosztu-złożoności

- kompromis pomiędzy kosztem dokonania błędnej klasyfikacji (R) i kosztem wynikającym z konieczności zbudowania drzewa (proporcjonalnym do liczy liści),
- zadanie polega na wybraniu drzewa zakorzenionego T drzewa pełnego T_0, dla którego minimum osiąga kryterium postaci

$$R_\alpha(T) = R(T) + \alpha|T|,$$

gdzie $\alpha \geq 0$ jest **współczynnikiem złożoności**

Algorytm kosztu-złożoności

- kompromis pomiędzy kosztem dokonania błędnej klasyfikacji (R) i kosztem wynikającym z konieczności zbudowania drzewa (proporcjonalnym do liczy liści),
- zadanie polega na wybraniu drzewa zakorzenionego T drzewa pełnego T_0, dla którego minimum osiąga kryterium postaci

$$R_\alpha(T) = R(T) + \alpha|T|,$$

  gdzie $\alpha \geq 0$ jest **współczynnikiem złożoności**
- dla T_0 mamy $R_0(T_0) = 0$ (bo $\alpha = 0$ i powtórne podstawienie daje brak błędu), ale wraz ze wzrostem α następuje moment, kiedy T_0 przestaje być optymalne,

Algorytm kosztu-złożoności

- kompromis pomiędzy kosztem dokonania błędnej klasyfikacji (R) i kosztem wynikającym z konieczności zbudowania drzewa (proporcjonalnym do liczy liści),
- zadanie polega na wybraniu drzewa zakorzenionego T drzewa pełnego T_0, dla którego minimum osiąga kryterium postaci

$$R_\alpha(T) = R(T) + \alpha|T|,$$

  gdzie $\alpha \geq 0$ jest **współczynnikiem złożoności**
- dla T_0 mamy $R_0(T_0) = 0$ (bo $\alpha = 0$ i powtórne podstawienie daje brak błędu), ale wraz ze wzrostem α następuje moment, kiedy T_0 przestaje być optymalne,
- można pokazać, że istnieje taka rodzina zagnieżdżonych poddrzew T_j, że każde z nich jest optymalne dla wszystkich α z pewnego przedziału $\alpha \in [\alpha_j, \alpha_{j+1})$, $\alpha_1 < \alpha_2 < ...\infty$

Algorytm kosztu-złożoności (cd)

Algorytm kosztu-złożoności (cd)

- czyli najpierw budujemy ciąg optymalnych poddrzew T_j w kolejnych przedziałach $[\alpha_j, \alpha_{j+1})$

Algorytm kosztu-złożoności (cd)

- czyli najpierw budujemy ciąg optymalnych poddrzew T_j w kolejnych przedziałach $[\alpha_j, \alpha_{j+1})$
- kolejne j-te drzewo buduje się dla ustalonej wartości współczynnika α np. $\alpha'_j = \sqrt{\alpha_j \alpha_{j+1}}$ (środek geometryczny odcinka),

Algorytm kosztu-złożoności (cd)

- czyli najpierw budujemy ciąg optymalnych poddrzew T_j w kolejnych przedziałach $[\alpha_j, \alpha_{j+1})$
- kolejne j-te drzewo buduje się dla ustalonej wartości współczynnika α np. $\alpha'_j = \sqrt{\alpha_j \alpha_{j+1}}$ (środek geometryczny odcinka),
- drzewo optymalne nie musi być jednoznaczne, algorytm pozwala wybrać najmniesze drzewo zagnieżdżone,

Algorytm kosztu-złożoności (cd)

- czyli najpierw budujemy ciąg optymalnych poddrzew T_j w kolejnych przedziałach $[\alpha_j, \alpha_{j+1})$
- kolejne j-te drzewo buduje się dla ustalonej wartości współczynnika α np. $\alpha'_j = \sqrt{\alpha_j \alpha_{j+1}}$ (środek geometryczny odcinka),
- drzewo optymalne nie musi być jednoznaczne, algorytm pozwala wybrać najmniesze drzewo zagnieżdżone,
- dla wszystkich drzew T_j obliczamy ułamek błednych klasyfikacji i wybieramy to drzewo, dla którego ten ułamek jest najmiejszy.