

SYSTEMY UCZĄCE SIĘ

WYKŁAD 4. DRZEWA REGRESYJNE, INDUKCJA REGUŁ

Dr hab. inż. Grzegorz Dudek
Wydział Elektryczny
Politechnika Częstochowska

Częstochowa 2014

Sposób konstrukcji i przycinania drzewa regresyjnego jest podobny do opisanego w poprzednim wykładzie.

W kryterium wyboru testu stosuje **zanieczyszczenie** węzła (*impurity*). Węzeł jest "czysty", jeśli wszystkie przykłady docierające do niego mają tę samą wartość pożądaną y . Im bardziej ta wartość jest zróżnicowana (większa wariancja) tym węzeł jest bardziej "zanieczyszczony".

Dąży się do tego, aby węzły potomne (tworzone w wyniku podziału) miały mniejsze zanieczyszczenie od węzła rodzicielskiego. Jako miary zanieczyszczenia używa się wariancji:

$$\sigma^2(P) = \frac{1}{|P|} \sum_{i \in \Pi} (y_i - \bar{y}(P))^2$$

gdzie: P , Π – zbiór przykładów i zbiór numerów przykładów przekazanych do węzła; y_i – wartość pożądana dla i -tego przykładu; $\bar{y}(P)$ – średnia wartość pożądana przykładów ze zbioru P .

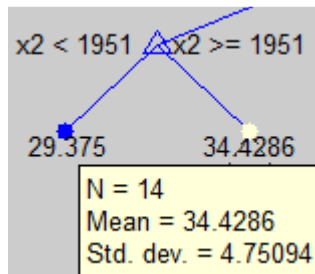
Kryterium wyboru testu – wybierz test maksymalizujący funkcję:

$$g_t(P) = \sigma^2(P) - \sum_{r \in R_t} \frac{|P_{tr}|}{|P|} \sigma^2(P_{tr})$$

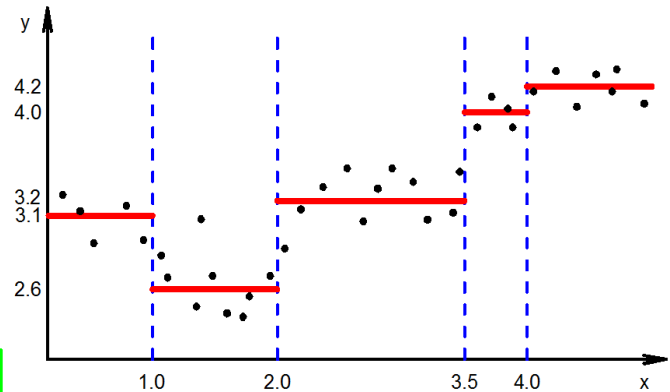
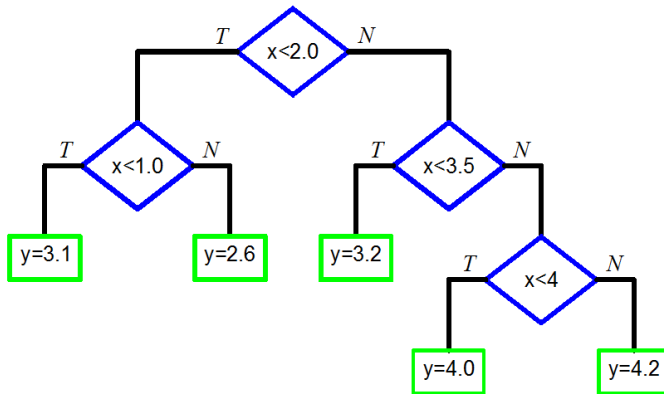
gdzie P_{tr} jest zbiorem przykładów, które w wyniku wykonania testu t uzyskały wynik r .

Kryterium stopu (utwórz liść zamiast węzła) – wariancja jest mniejsza od założonego progu V lub liczba przykładów jest mniejsza od wartości progowej U : $\sigma^2(P) < V \vee |P| < U$

Wpis do etykiety liścia – średnia wartość pożądana przykładów docierających do liścia $\bar{y}(P)$



Aproksymacja odcinkami równoległymi do osi OX



- Reprezentacja drzewiasta lub zbiór reguł decyzyjnych "jeśli – to"
- Działanie na atrybutach nominalnych, porządkowych i ciągłych
- Wbudowany mechanizm selekcji atrybutów
- Podział przestrzeni cech na hiperprostopadłościany
- Lokalna aproksymacja funkcji stałą wewnątrz hiperprostopadłościanu (aproksymacja dyskretna)
- Zależnie od funkcji docelowej drzewo decyzyjne może pełnić rolę klasyfikatora lub modelu regresyjnego

Hipotezy reprezentowane są za pomocą zbiorów reguł postaci:

JEŚLI warunki TO decyzja

Część warunkowa reguły (przesłanka) reprezentowana jest tu przez tzw. **kompleksy**, a decyzja (konkluzja) określa etykietę klasy przykładów pokrywanych przez te kompleksy.

Reguły pozwalają zapisać wiedzę w sposób czytelny dla ludzi.

Jeśli przez $a_1, a_2, \dots, a_n$ oznaczymy atrybuty przykładów x , to reguła może być zapisana:

$$(\forall x \in X) (\alpha[a_1(x), a_2(x), \dots, a_n(x)] \rightarrow h(x) = d)$$

gdzie $\alpha(.)$ jest pewną formułą logiczną działającą na atrybutach przykładów.

Reguła ta wyraża zdanie: każdemu przykładowi z dziedziny spełniającemu warunki α hipoteza h przypisuje klasę d .

Kompleks opisuje warunki, jakie muszą spełniać wartości atrybutów przykładu, aby został on przez ten kompleks **pokryty**, a tym samym sklasyfikowany jako pozytywny przez odpowiednią hipotezę.

Kompleks jest wektorem interpretowanym jako koniunkcja (AND) warunków (tzw. **selektorów**) nakładanych na pojedyncze atrybuty. Selektor umieszczony na i -tej pozycji w kompleksie reprezentuje warunek nakładany na wartość i -tego atrybutu.

Selektory dzielimy na cztery grupy:

- selektor pojedynczy – spełniony dla przykładu x , jeśli atrybut a_i tego przykładu ma pewną wskazaną wartość: $a_i(x) = v$
- selektor dysjunkcyjny – spełniony dla przykładu x , jeśli atrybut a_i tego przykładu ma jedną z wymienionych wartości: $a_i(x) = v_1 \vee a_i(x) = v_2 \vee \dots \vee a_i(x) = v_m$
- selektor uniwersalny (oznaczany $?$) – spełniony dla przykładu x przez dowolną wartość atrybutu a_i z zakresu jego wartości A_i : $a_i(x) \in A_i$
- selektor pusty (oznaczany $\emptyset$) – nie spełniony dla przykładu x przez żadną wartość atrybutu a_i z jego dziedziny.

Kompleks $\mathbf{k} = \langle s_1, s_2, \dots, s_n \rangle$ pokrywa przykład $\mathbf{x}$, jeżeli każdy selektor s_i dla $i = 1, 2, \dots, n$ pokrywa przykład $\mathbf{x}$, tzn. wartości wszystkich atrybutów przykładu $\mathbf{x}$ spełniają warunki ujęte w odpowiadających im selektorach. Piszemy wówczas $\mathbf{k} \triangleright \mathbf{x}$.

Np. kompleks $\langle 7, ?, \text{mały} \vee \text{średni}, 1 \vee 2 \vee 5 \rangle$ pokrywa przykłady, których pierwszy atrybut ma wartość 7, drugi atrybut ma dowolną wartość (?), trzeci atrybut ma wartość *mały* lub *średni*, a czwarty atrybut ma wartość, 1, 2 lub 5.

Ten kompleks można zapisać w postaci formuły logicznej:

$$a_1(x) = 7 \wedge (a_3(x) = \text{mały} \vee a_3(x) = \text{średni}) \wedge (a_4(x) = 1 \vee a_4(x) = 2 \vee a_4(x) = 5)$$

Mając zbiór reguł $r \in R$ znajdujemy reguły pokrywające klasyfikowany przykład $\mathbf{x}$ (tzn. reguły których kompleksy pokrywają $\mathbf{x}$). Mogą zaistnieć trzy przypadki:

- tylko jedna reguła pokrywa $\mathbf{x}$.

Wtedy przykład klasyfikuje się do klasy zgodnej z decyzją tej reguły.

- kilka reguł pokrywa $\mathbf{x}$.

Wtedy wyznacza się liczby przykładów trenujących pokrywanych przez te reguły $|T_r|$. Klasa reprezentowana przez reguły pokrywające największą liczbę przykładów trenujących i przykład $\mathbf{x}$ przypisywana jest do przykładu $\mathbf{x}$:

$$h(\mathbf{x}) = \arg \max_{d \in C} \sum_{r \in R_x^d} |T_r|$$

gdzie: $R_x^d = \{r \in R \mid r \triangleright \mathbf{x} \wedge d_r = d\}$.

Jeśli reguły pokrywają jednocześnie przykłady z różnych klas, przewidywaną klasę przykładu $\mathbf{x}$ można wyznaczyć ze wzoru:

$$h(\mathbf{x}) = \arg \max_{d \in C} \sum_{\mathbf{r} \in R_{\mathbf{x}}} |T_{\mathbf{r}}^d|$$

gdzie: $R_{\mathbf{x}} = \{\mathbf{r} \in R \mid \mathbf{r} \triangleright \mathbf{x}\}$, $|T_{\mathbf{r}}^d|$ – liczba przykładów trenujących z klasy d pokrywanych przez regułę $\mathbf{r}$.

- żadna reguła nie pokrywa $\mathbf{x}$.

Przypisanie do $\mathbf{x}$ klasy większościowej w zbiorze trenującym lub określenie klasy wykorzystując miarę częściowego pokrywania przykładów przez reguły: $\mu(\mathbf{k}_r, \mathbf{x}) \in [0, 1]$. $\mu(\mathbf{k}_r, \mathbf{x}) = 1$, gdy kompleks reguły $\mathbf{r}$ pokrywa $\mathbf{x}$, i $\mu(\mathbf{k}_r, \mathbf{x}) < 1$, gdy pokrycie jest częściowe. Miara częściowego pokrywania:

$$\mu(\mathbf{k}, \mathbf{x}) = \prod_{i=1}^n \mu(s_i^k, \mathbf{x})$$

gdzie s_i^k jest selektorem znajdującym się na pozycji i w kompleksie $\mathbf{k}$, a $\mu(s_i^k, \mathbf{x})$ jest miarą pokrywania przykładu $\mathbf{x}$ przez ten selektor, np.:

$$\mu(s_i^k, \mathbf{x}) = \begin{cases} 1, & \text{jeżeli } a(\mathbf{x}) \in V_s \\ \frac{|V_s|}{|A|}, & \text{jeżeli } a(\mathbf{x}) \notin V_s \end{cases}$$

gdzie V_s jest zbiorem wartości selektora s , a A jest zbiorem wartości atrybutu $a(\mathbf{x})$.

Przewidywaną klasę przykładu $\mathbf{x}$ można wyznaczyć ze wzoru:

$$h(\mathbf{x}) = \arg \max_{d \in C} \sum_{\mathbf{r} \in R} \mu(\mathbf{k}_r, \mathbf{x}) |T_r^d|$$

Dany jest zbiór reguł:

$$r_1 = \langle \text{słoneczna} \vee \text{deszczowa}, \text{zimna} \vee \text{ciepła}, \text{duża}, ? \rangle \rightarrow 0,$$

$$r_2 = \langle \text{pochmurna} \vee \text{deszczowa}, ?, ?, \text{słaby} \rangle \rightarrow 1,$$

$$r_3 = \langle \text{deszczowa}, \text{zimna} \vee \text{umiarkowana}, ?, \text{silny} \rangle \rightarrow 0,$$

$$r_4 = \langle \text{słoneczna} \vee \text{pochmurna}, \text{zimna} \vee \text{umiarkowana}, ?, \text{silny} \rangle \rightarrow 1,$$

$$r_5 = \langle \text{słoneczna} \vee \text{pochmurna}, \text{zimna} \vee \text{umiarkowana}, \text{duża}, \text{słaby} \rangle \rightarrow 0,$$

$$r_6 = \langle ?, ?, \text{normalna}, \text{słaby} \rangle \rightarrow 1.$$

Liczba przykładów trenujących pokrywanych przez reguły:

$$|T_{r_1}| = |T_{r_1}^0| = |\{1, 2\}| = 2,$$

$$|T_{r_2}| = |T_{r_2}^1| = |\{3, 4, 5, 10, 13\}| = 5,$$

$$|T_{r_3}| = |T_{r_3}^0| = |\{6, 14\}| = 2,$$

$$|T_{r_4}| = |T_{r_4}^1| = |\{7, 11, 12\}| = 3,$$

$$|T_{r_5}| = |T_{r_5}^0| = |\{8\}| = 1,$$

$$|T_{r_6}| = |T_{r_6}^1| = |\{5, 9, 10, 13\}| = 4.$$

Wyznacz klasy dwóch przykładów:

ZASADY KLASYFIKACJI – PRZYKŁAD

x	<i>aura</i>	<i>temperatura</i>	<i>wilgotność</i>	<i>wiatr</i>
15	<i>słoneczna</i>	<i>zimna</i>	<i>duża</i>	<i>silny</i>
16	<i>deszczowa</i>	<i>ciepła</i>	<i>normalna</i>	<i>silny</i>

Przykład $\mathbf{x}_{15}$ pokrywany jest przez $\mathbf{r}_1$ i $\mathbf{r}_4$, więc $R_{15} = \{\mathbf{r}_1, \mathbf{r}_4\}$.

$$\sum_{\mathbf{r} \in R_{15}^0} |T_{\mathbf{r}}| = |T_{\mathbf{r}_1}| = 2, \quad \sum_{\mathbf{r} \in R_{15}^1} |T_{\mathbf{r}}| = |T_{\mathbf{r}_4}| = 3, \quad \text{stąd } h(\mathbf{x}_{15}) = 1.$$

Przykład $\mathbf{x}_{16}$ nie jest pokrywany przez żadną regułę, wykorzystujemy częściowe pokrywanie:

$$\begin{aligned} \mu(\mathbf{r}_1, 16) &= 1 \cdot 1 \cdot \frac{1}{2} \cdot 1 = 0,5, & \mu(\mathbf{r}_2, 16) &= 1 \cdot 1 \cdot 1 \cdot \frac{1}{2} = 0,5, & \mu(\mathbf{r}_3, 16) &= 1 \cdot \frac{2}{3} \cdot 1 \cdot 1 = 0,667, \\ \mu(\mathbf{r}_4, 16) &= \frac{2}{3} \cdot \frac{2}{3} \cdot 1 \cdot 1 = 0,444, & \mu(\mathbf{r}_5, 16) &= \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = 0,111, & \mu(\mathbf{r}_6, 16) &= 1 \cdot 1 \cdot 1 \cdot \frac{1}{2} = 0,5. \end{aligned}$$

Wobec tego:

$$\begin{aligned} \sum_{\mathbf{r} \in R} \mu(\mathbf{r}, 16) |T_{\mathbf{r}}^1| &= 0,5 \cdot 0 + 0,5 \cdot 5 + 0,667 \cdot 0 + 0,444 \cdot 3 + 0,111 \cdot 0 + 0,5 \cdot 4 = 5,832, \\ \sum_{\mathbf{r} \in R} \mu(\mathbf{r}, 16) |T_{\mathbf{r}}^0| &= 0,5 \cdot 2 + 0,5 \cdot 0 + 0,667 \cdot 2 + 0,444 \cdot 0 + 0,111 \cdot 1 + 0,1 \cdot 0 = 2,445. \end{aligned}$$

Stąd $h(\mathbf{x}_{16}) = 1$.

INDUKCJA REGUŁ — POKRYWANIE SEKWENCYJNE

- Generujemy pojedynczo kolejne reguły, z których każda pokrywa część przykładów klasy c zbioru trenującego.
- Proces jest kontynuowany aż do uzyskania pokrycia wszystkich przykładów trenujących klasy c .
- Każda reguła pokrywa przykłady tej samej klasy lub przynajmniej z wyraźnie dominującą klasą większościową.
- Klasa ta jest umieszczana w części decyzyjnej reguły.

funkcja *sekwencyjne-pokrywanie*(T)

argumenty wejściowe: T — zbiór trenujący dla pojęcia c ;

zwraca: zbiór reguł reprezentujący hipotezę przybliżającą c ;

- 1: $\mathbb{R} := \emptyset$; $P := T$;
- 2: **jak długo** $P \neq \emptyset$ **wykonaj**
- 3: $\mathbf{k} := \text{znajdź-kompleks}(T, P)$;
- 4: $d := \text{kategoria}(\mathbf{k}, T, P)$;
- 5: $\mathbb{R} := \mathbb{R} \cup \{\mathbf{k} \rightarrow d\}$;
- 6: $P := P - P_{\mathbf{k}}$;
- 7: **koniec jak długo**
- 8: **zwróć** $\mathbb{R}$.

INDUKCJA REGUŁ – ALGORYTM AQ

Algorytm AQ realizuje schemat pokrywania sekwencyjnego. Kolejne kroki algorytmu AQ zostaną opisane w sposób uproszczony.

Niech P_s oznacza zbiór przykładów pozytywnych $\mathbf{x}_s$ (klasy i), a P_n – zbiór przykładów negatywnych $\mathbf{x}_n$ (o innych klasach niż i). Tworzymy regułę klasyfikującą przykłady do klasy i .

1. Wybierz przykład $\mathbf{x}_s$ (tzw. **ziarno**) spośród przykładów niepokrytych.
2. Dla każdego przykładu $\mathbf{x}_n$ ze zbioru P_n zbuduj tzw. **różnicę** złożoną z wartości atrybutów $\mathbf{x}_s$, które są różne od wartości tych atrybutów w $\mathbf{x}_n$. Atrybut wraz z jego wartością będziemy traktować jako selektor.
3. Wybierz po jednym selektorze z każdej różnicy i połącz je każdy z każdym poprzez koniunkcję ($\wedge$). W ten sposób otrzymujemy zbiór kompleksów K . Zauważ, że żaden z tych kompleksów nie pokrywa przykładu negatywnego.
4. Dokonaj uogólnienia zbioru K poprzez odrzucenie powtarzających się selektorów w tym samym kompleksie i odrzucenie powtarzających się kompleksów.

5. Ze zbioru K wybierz najlepszy kompleks k^* według przyjętego kryterium, np. kompleks pokrywający największą liczbę przykładów pozytywnych lub kompleks z najmniejszą liczbą atrybutów.
6. Ze zbioru P_s usuń przykłady pokryte przez kompleks k^* .
7. Dołącz kompleks k^* do warunku reguły.
8. Jeśli P_s jest zbiorem pustym, przejdź do p. 9, w przeciwnym przypadku przejdź do p. 1.
9. Przedstaw uzyskane wyniki.

Przykład. Dane są obiekty z klasy 1: $\mathbf{x}_1 = [0, 2, 5]$ i $\mathbf{x}_2 = [7, 2, 8]$

oraz z klasy 2: $\mathbf{x}_3 = [0, 3, 4]$ i $\mathbf{x}_4 = [1, 3, 8]$.

Wygeneruj kompleksy i utwórz regułę decyzyjną do rozpoznawania klasy 1.

1. Wybieramy przykład $\mathbf{x}_1$.
2. Tworzymy różnice: $\mathbf{d}_{1,1} = [-, 2, 5]$, $\mathbf{d}_{1,2} = [0, 2, 5]$
3. Tworzymy kompleksy: $\mathbf{k}_{2-1} = \langle 0, 2, ? \rangle$, $\mathbf{k}_{2-2} = \langle ?, 2, ? \rangle$, $\mathbf{k}_{2-3} = \langle ?, 2, 5 \rangle$, $\mathbf{k}_{3-1} = \langle 0, ?, 5 \rangle$,
 $\mathbf{k}_{3-2} = \langle ?, 2, 5 \rangle$, $\mathbf{k}_{3-3} = \langle ?, ?, 5 \rangle$
4. Odrzucamy kompleks $\mathbf{k}_{3-2}$ identyczny z $\mathbf{k}_{2-3}$.
5. Wybieramy kompleks pokrywający największą liczbę przykładów z klasy 1: $\mathbf{k}^* = \mathbf{k}_{2-2}$.
6. Usuwamy przykłady $\mathbf{x}_1$ i $\mathbf{x}_2$ pokryte przez $\mathbf{k}^*$ ze zbioru P_s .
7. Dołączamy kompleks $\mathbf{k}^*$ do warunku reguły: JEŚLI $\langle ?, 2, ? \rangle$, TO klasa = 1
8. Zbiór $P_s = \emptyset$.

Utworzona reguła ma postać: JEŚLI $a_2(\mathbf{x}) = 2$, TO klasa = 1