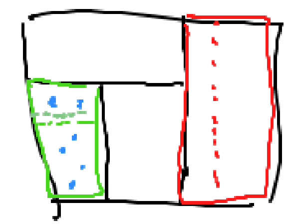
Tworzenie drzewa – algorytm zachłanny.

· dzielimy obszar $X = \{(x_1, x_2, \dots, x_p)\}$ na prostokąty parami rozłączne R_j; na początku: jeden prostokąt X

· jeśli mamy pewien podział X na $R_1, R_2, \dots, R_J$, to rozważamy wszystkie możliwe podziały dowolnego z prostokątów R_ℓ otrzymane przez hiperpłaszczyzny $\{x_j < s\}$;

$$R^{(1)}(j,s) = R_\ell \cap \{(x_1, \dots, x_p) : x_j < s\}$$
$$R^{(2)}(j,s) = R_\ell \cap \{(x_1, \dots, x_p) : x_j \geq s\}$$



: wybieramy najlepszy podział – tzn. taki, który minimalizuje:

np. dla ~~d~~ przypadku regresji

$$\sum_{i: x^{(i)} \in R^{(1)}(j,s)} (y_i - \hat{y}_{R^{(1)}})^2 + \sum_{i: x^{(i)} \in R^{(2)}(j,s)} (y_i - \hat{y}_{R^{(2)}})^2$$

(+ suma dla pozostałych prostokątów)

dla klasyfikacji:
$$\hat{p}_{mk} = \frac{\#\text{ próbek w } R_m\text{, które są klasy } k}{\#\text{ próbek w } R_m}$$

$$\hat{p}_{m1} + \hat{p}_{m2} + \dots + \hat{p}_{mK} = 1$$

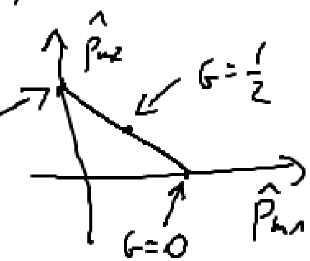
minimalizujemy uśredniamy ważone po wszystkich m z wagami odp. liczbie próbek

$$E = 1 - \max_k \hat{p}_{mk}$$

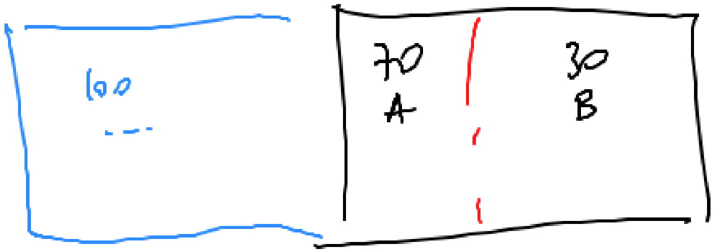
$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) = 1 - \sum_{k=1}^{K} (\hat{p}_{mk})^2$$

wskaźnik Giniego

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk}$$ – entropia



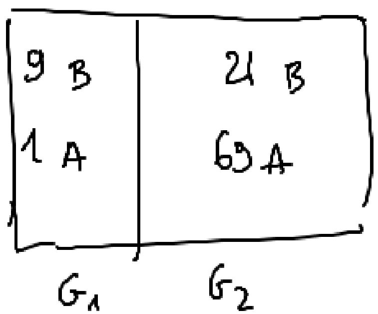
$$-\log(\underbrace{1-x}_{y}) \approx \overbrace{+x}^{(1-y)}$$
$x \approx 0$, $y \approx 1$



$$G = 1 - \left(\left(\frac{30}{100}\right)^2 + \left(\frac{70}{100}\right)^2 \right) = 1 - 0.09 - 0.49 = 0.42$$

średnia: $\frac{100}{100} \cdot G = 0.42$

podział:



0.5

$$G_1 = 1 - \left(\left(\frac{9}{10}\right)^2 + \left(\frac{1}{10}\right)^2 \right) = 1 - 0.81 - 0.01 = 0.18$$

$$G_2 = 1 - \left(\left(\frac{21}{90}\right)^2 + \left(\frac{69}{90}\right)^2 \right) \approx 0.36$$

~~Suma $G_1 + G_2$~~ =

średnia: $0.9 \cdot G_2 + 0.1 \cdot G_1 \approx 0.32 + 0.02 = 0.34$

0.45 0.05

Tworzymy drzewo za pomocą algorytmu zachłannego tak długo, jak się da, lub zatrzymujemy się po osiągnięciu pewnej wielkości drzewa (np. liczba liści, głębokość, liczba próbek w liściu). Następnie przycinamy.

Przycinanie

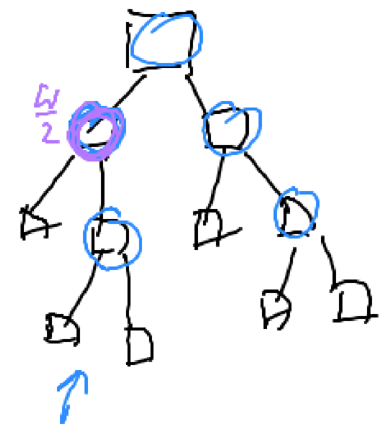
„cost complexity pruning" (przycinanie wg kosztu złożoności (?))

$\alpha \geq 0$

dla regresji:

minimalizujemy

$$(*) \quad \underbrace{\sum_{m=1}^{|T|} \sum_{i: x_i \in R_m} (y_i - \hat{y}_{R_m})^2}_{S} + \alpha \underbrace{|T|}_{\text{liczba liści w drzewie}}$$



dla każdego węzła możemy łatwo znaleźć α, przy której zaczyna się opłacać przyciąć drzewo w tym węźle

po wszystkich możliwych poddrzewach T naszego drzewa.

- Zaczynamy od wyjściowego drzewa T_1
- rozważamy najmniejsze α, dla którego (*) jest minimalna dla pewnego właściwego poddrzewa T_{k+1} $\quad$ drzewo $T_k \longrightarrow T_{k+1}$
- powtarzamy dla drzewa T_{k+1} w miejsce T_k

Np. dla węzła O:

suma S się zwiększy po jego obcięciu o pewną liczbę W

liczba liści zmniejszy się o 2

(*) zwiększy się o $\longrightarrow (W - 2\alpha) \leq 0 \quad \alpha = \frac{W}{2}$

Jaką α wybrać?

$\alpha = 0$ drzewo oryginalne

α duże sam korzeń

1) Trenujemy drzewa używając części danych (treningowych) i szukamy α sprawdzając drzewo na pozostałych danych (testowych)

2) Używamy walidacji krzyżowej