

Zaawansowane uczenie maszynowe: *ćwiczenia do wykładu 9*

Paweł Cichosz

x	a_1	a_2	a_3	c	f
1	1	1	2	0	2
2	1	2	3	0	5
3	1	3	4	0	8
4	2	1	1	0	3
5	2	2	2	0	4
6	1	1	1	1	1
7	1	2	1	1	3
8	1	3	4	1	6
9	2	2	3	1	7
10	2	3	2	1	5

1. Dla drzewa regresji h_1 do predykcji f na podstawie a_1, a_2, a_3 , w którego korzeniu znajduje się podział $a_1 = 1$, w obu węzłach potomnych podziały $a_3 \leq 2$, a ich potomkami są liście, potraktowano różnice $f(x) - h_1(x)$ jako nowe wartości docelowe dla budowy kolejnego drzewa h_2 , w którym stosowano podziały wybierane na podstawie wariancji wartości docelowych, zaś liście tworzone w przypadku braku możliwości podziału lub po przekroczeniu poziomu 2 (tzn. poniżej węzłów potomnych korzenia). Wyznaczyć spadek średniego błędu bezwzględnego wynikający z wykorzystania przy predykcji dodatkowo drugiego drzewa.
2. Wyznaczyć \mathcal{G}_x i \mathcal{H}_x dla algorytmu *xgboost* z wykładniczą funkcją straty:

$$\mathcal{L}(f(x), h(x)) = e^{-f(x)h(x)}$$

3. Przyjmując, że algorytm *xgboost* jest stosowany ze stratą logarytmiczną do tworzenia modelu klasyfikacji przewidującego prawdopodobieństwo $c = 1$ na podstawie a_1, a_2, a_3 , h_1 jest drzewem złożonym z jednego liścia z wartością 0, a liczba poziomów węzłów nie może przekraczać 2, wyznaczyć drzewo h_2 dla $\lambda = 1$ i $\gamma = 0.1$.
4. Przyjmując, że algorytm *xgboost* jest stosowany ze stratą kwadratową do tworzenia modelu regresji przewidującego f na podstawie a_1, a_2, a_3 , h_1 jest drzewem złożonym

z jednego liścia z wartością 0, a liczba poziomów węzłów nie może przekraczać 2, wyznaczyć drzewo h_2 dla $\lambda = 1$ i $\gamma = 0.1$.