Zaawansowane uczenie maszynowe: *ćwiczenia do wykładu 5*

Paweł Cichosz

x	a_1	a_2	a_3	a_4	a_5	c
1	1	1	2	1	1	0
2	1	2	3	2	6	0
3	1	2	3	2	9	0
4	2	1	1	3	3	0
5	2	2	2	3	5	0
6	1	1	1	3	0	1
7	1	2	1	1	2	1
8	1	2	3	2	6	1
9	2	2	3	3	8	1
10	2	2	2	3	7	1

1. Dla każdego węzła drzewa decyzyjnego do predykcji c, które w korzeniu zawiera podział z użyciem warunku $a_2 = 1$, a w obu węzłach potomnych – podziały z użyciem warunku $a_1 = 1$, sprawdzić, przy jakiej wartości m byłby przycięty z wykorzystaniem kryterium MEP.

2. Dla każdego węzła drzewa decyzyjnego do predykcji c, które w korzeniu zawiera podział z użyciem warunku $a_2 = 1$, a w obu węzłach potomnych – podziały z użyciem warunku $a_1 = 1$, sprawdzić, przy jakiej wartości α byłby przycięty z wykorzystaniem kryterium CCP w postaci nierównościowej:

$$e_{T_\mathbf{n}}(\mathbf{l}) \leq e_{T_\mathbf{n}}(\mathbf{n}) + \alpha \mathcal{C}(\mathbf{n})$$

3. Dla każdego węzła drzewa decyzyjnego do predykcji c, które w korzeniu zawiera podział z użyciem warunku $a_2 = 1$, a w obu węzłach potomnych – podziały z użyciem warunku $a_1 = 1$, sprawdzić, przy jakiej wartości α byłby przycięty z wykorzystaniem metody CCP polegającej na wyborze przyciętego wariantu drzewa, który minimalizuje:

$$e_T(\mathbf{n}_1) + \alpha \mathcal{C}(\mathbf{n}_1)$$

gdzie $\mathbf{n}_1$ oznacza węzeł-korzeń drzewa.

$$e_{T_\mathbf{n}}(\mathbf{l}) \leq e_{T_\mathbf{n}}(\mathbf{n}) + \alpha \mathcal{C}(\mathbf{n})$$

4. Wykorzystując technikę przykładów ułamkowych, dokonać predykcji prawdopodobieństwa klasy 1 dla następujących przykładów z podanego wyżej zbioru:
   - 1 po usunięciu wartości atrybutu a_1,
   - 7 po usunięciu wartości atrybutu a_2,
   - 10 po usunięciu wartości atrybutów a_1 i a_2,

   za pomocą drzewa decyzyjnego, które w korzeniu zawiera podział z użyciem atrybutu a_2, a w obu węzłach potomnych – podziały z użyciem atrybutu a_1.

5. Traktując atrybuty jako numeryczne, zbudować drzewo regresji do predykcji f na podstawie a_1, a_2, a_3, w którym:
   - stosowane są binarne podziały nierównościowe wybierane według odchylenia standardowego,
   - liście tworzone są po osiągnięciu odchylenia standardowego poniżej 1, mniej niż 3 przykładów lub wyczerpaniu możliwości podziału,

   oraz wyznaczyć jego błąd średniokwadratowy oraz współczynnik determinacji na zbiorze trenującym.