

Uczenie maszynowe: ćwiczenia do wykładu 3

Paweł Cichosz

1. Dla dziedziny X , opisywanej przez atrybuty nominalne $a_1 : X \mapsto \{1, 2, 3\}$, $a_2 : X \mapsto \{1, 2, 3\}$ i $a_3 : X \mapsto \{1, 2, 3\}$, zbioru klas $C = \{0, 1\}$ oraz klasy pojęć i przestrzeni modeli reprezentowanych przez drzewa decyzyjne zawierające dokładnie dwa poziomy węzłów z równościowymi warunkami podziału ($a = v$), oszacować liczbę przykładów trenujących wystarczającą, aby (dla ustalonych δ i ϵ) z prawdopodobieństwem co najmniej $1 - \delta$ uzyskać model o błędzie rzeczywistym nieprzekraczającym ϵ przy użyciu spójnego algorytmu uczenia się.
2. Dla dziedziny X , opisywanej przez atrybuty nominalne $a_1 : X \mapsto \{1, 2, 3\}$, $a_2 : X \mapsto \{1, 2, 3\}$ i $a_3 : X \mapsto \{1, 2, 3\}$, zbioru klas $C = \{0, 1\}$, klasy pojęć reprezentowanych przez dowolne funkcje $\{1, 2, 3\}^3 \rightarrow \{0, 1\}$ oraz przestrzeni modeli reprezentowanych przez koniunkcje warunków przynależności do klasy 1, oszacować liczbę przykładów trenujących wystarczającą, aby (dla ustalonych δ i ϵ) z prawdopodobieństwem co najmniej $1 - \delta$ uzyskać model o błędzie rzeczywistym nieprzekraczającym błędem na zbiorze trenującym o więcej niż ϵ przy użyciu spójnego algorytmu uczenia się.
3. Dla dziedziny X , opisywanej przez atrybuty nominalne $a_1 : X \mapsto \{1, 2, 3\}$, $a_2 : X \mapsto \{1, 2, 3\}$ i $a_3 : X \mapsto \{1, 2, 3\}$, zbioru klas $C = \{0, 1\}$, klasy pojęć reprezentowanych przez dowolne funkcje $\{1, 2, 3\}^3 \rightarrow \{0, 1\}$ oraz przestrzeni modeli reprezentowanych przez drzewa decyzyjne zawierające dokładnie dwa poziomy węzłów z równościowymi warunkami podziału ($a = v$), oszacować liczbę przykładów trenujących wystarczającą, aby (dla ustalonych δ i ϵ) z prawdopodobieństwem co najmniej $1 - \delta$ uzyskać model o błędzie rzeczywistym nieprzekraczającym błędem na zbiorze trenującym ϵ przy użyciu spójnego algorytmu uczenia się.
4. Wyznaczyć wymiar VC dla przestrzeni modeli reprezentowanych przez binarne drzewa decyzyjne o dokładnie dwóch poziomach podziałów.
5. Dla dziedziny $X = \{0, 1\}^3$ z atrybutami binarnymi a_1, a_2, a_3 , zbioru klas $C = \{0, 1\}$ oraz przestrzeni modeli reprezentowanych przez dowolne funkcje boolowskie, porównać oszacowania liczby przykładów trenujących wystarczającej, aby (dla ustalonych δ i ϵ) z prawdopodobieństwem co najmniej $1 - \delta$ uzyskać model o błędzie

rzeczywistym nieprzekraczającym jego błędu na zbiorze trenującym o więcej niż ϵ , uzyskane na podstawie rozmiaru przestrzeni modeli i jej wymiaru VC.

6. Dla dziedziny $X = \{0, 1\}^3$ z atrybutami binarnymi a_1, a_2, a_3 , zbioru klas $C = \{0, 1\}$ oraz przestrzeni modeli reprezentowanych przez koniunkcje funkcje boolowskie, porównać oszacowania liczbę przykładów trenujących wystarczającej, aby (dla ustalonych δ i ϵ) z prawdopodobieństwem co najmniej $1 - \delta$ uzyskać model o błędzie rzeczywistym nieprzekraczającym jego błędu na zbiorze trenującym o więcej niż ϵ , uzyskane na podstawie rozmiaru przestrzeni modeli i jej wymiaru VC.
7. Dla dziedziny $X = \mathcal{R}^2$, opisywanej przez atrybuty ciągłe a_1, a_2 , zbioru klas $C = \{0, 1\}$ oraz przestrzeni modeli reprezentowanych przez prostokąty o bokach równoległych do osi układu współrzędnych oszacować liczbę przykładów trenujących wystarczającą, aby (dla ustalonych δ i ϵ) z prawdopodobieństwem co najmniej $1 - \delta$ uzyskać model o błędzie rzeczywistym nieprzekraczającym ϵ , wykorzystując wymiar VC przestrzeni modeli.