

Kolokwium ML grupa 2

1. Drzewa decyzyjne:

- a. Oblicz entropię początkową, entropię warunkową, information gain, intrinsic information, gain ratio dla podziału według atrybutu „gorączka” (pierwszy podział w drzewie) (5 pkt).

Gorączka	Osłabienie	Ból gardła	Klasa
Duża	Tak	Tak	Chory
Średnia	Tak	Tak	Chory
Duża	Nie	Tak	Chory
Mała	Nie	Tak	Zdrowy
Duża	Nie	Tak	Chory
Mała	Nie	Nie	Zdrowy
Średnia	Tak	Nie	Zdrowy
Srednia	Tak	Tak	Chory
Mała	Nie	Nie	Zdrowy

- b. Opisz sposób radzenia sobie z wartościami ciągłymi w atrybutach (2 pkt).
- c. Na czym polega pre- i post-pruning w drzewach decyzyjnych (3 pkt)?

2. K-NN:

- a. Opisz szczegółowo działanie algorytmu klasyfikacyjnego k-nn (tylko dla atrybutów liczbowych) (3 pkt).
- b. Wymień 3 dowolne wykorzystywane miary dystansu w tym algorytmie (1 pkt).
- c. Opisz jak można sobie poradzić z wartościami nominalnymi w atrybutach (1 pkt).

3. K-MEANS:

- a. Opisz czym jest przekleństwo wielowymiarowości (1 pkt).
- b. Opisz szczegółowo działanie algorytmu k-means (tylko dla atrybutów liczbowych) (3 pkt).
- c. Wymień dwie dowolne metody doboru początkowych centroid (1 pkt).

4. Regresja:

- a. Na czym polega regresja logistyczna i do czego służy (3 pkt).
- b. Wymień dwie metody dopasowywania regresora liniowego (1 pkt).
- c. Czym różni się regresja wieloraka od prostej (multiple vs simple) (1 pkt)?

5. AHC (wykład):

- a. Opisz działanie algorytmu grupowania aglomeracyjnego hierarchicznego (5 pkt).