

FLP - Rozhodovacie stromy

Autor: Martin Takács

Login: xtakac07

Rok: 2025

Čo som stihol?

Stihol som naimplementovať všetko, čo bolo v zadaní požadované. Aj načítanie z textových súborov, aj klasifikáciu, aj trénovanie.

Čo som nestihol?

Z povinných častí nič.

Na testovacích dátach nemám všade 100-percentnú úspešnosť, iba na niektorých. Avšak na všetkých testovacích datasetoch mám úspoešnosť cez 90 percent, s čím som osobne spokojný.

Moja pôvodná implementácia trénovania zahrnovala maximálnu hĺbku a to tým spôsobom, že som pri každom uzle odstránil príznak, podľa ktorého sa rozhodovalo. Po konzultácii s pánom Ing. Poliakovom som túto funkcionality jednoducho odkomentoval. Z časového hľadiska som to viac neoptimalizoval. Štyri zakomentované riadky v zdrojovom súbore *Decisiontree.hs* som ponechal pre vysvetlenie, ako to bolo predtým, keby náhodou niektorá časť kódu nebola zrozumiteľná, že prečo som to tak spravil.

Použité knižnice navyše

V kóde používam okrem Prelude aj balíky ***System.Environment*** (*getArgs*), ***Data.Ord*** (*comparing*), ***Data.List*** (*sort,nub,elemIndex,sortBy,minimumBy,maximumBy,group,partition*).

Ako to funguje?

Spustenie

Príkaz **make** vygeneruje spustiteľný program *flp-fun* a pomocné súbory

Príkaz **make clean** pomocné súbory vymaže. *flp-fun* ponechá.

Zdojové kódy

Hlavným vstupom programu je ***Main.hs***, kde sa nerobí nič iné iba načítanie argumentov, na základe ktorých sa spustí buď klasifikácia alebo trénovanie.

Vlastný dátový typ stromu **Tree** a typové aliasy sú definované v module ***Datatypes***.

Modul ***Classification*** načíta najskôr dáta zo zadanej *data_path* a spustí na nich preprocessing funkciou *convertData*. Následne načíta strom z *tree_path* a pomocou funkcie *loadTree* ho prevedie do dátového typu **Tree**. Nakoniec sa vypočítajú predikcie a vypíšu sa na std výstup.

Modul ***Training*** načíta trénovací dataset, preprocesne dáta a spustí trénovanie (vytváranie) nového stromu. Na rozhodovanie pri uzloch sa využíva Gini index.

Všetka funkcionality stromu je definovaná v module ***Decisiontree***. Obsahuje funkcie na prechádzanie stromom, klasifikáciu a trénovanie nového stromu.

Pomocné funkcie sú definované v module ***Utils***.