

NPFL142.C4DHI – domácí úkol #4

Příklad 4.1

Testovací množina obsahuje 1 000 příkladů. Binární klasifikátor z nich 400 klasifikoval do pozitivní třídy a 600 do negativní s úspěšností (accuracy) 0,7 a specificitou (specificity) 0,8. Ve vašem sdíleném dokumentu

- napište kompletní matici konfuze
- vypočtěte přesnost (precision) klasifikátoru
- vypočtěte úplnost (recall) klasifikátoru

Pozn. specificita = $TN / (TN + FP)$

Příklad 4.2

Máme kolekci 200 studentů a o každém studentovi víme jeho pohlaví a výsledky testů ve čtení, psaní a matematice. Úkolem je předpovědět, zda student je ve třídě s vyznamenáním, nebo ne. Jedná se o binární klasifikační úlohu s cílovým atributem $hon = \{0, 1\}$ a k predikci použijte příznaky `female`, `read`, `write`, `math`.

Na clusteru v adresáři 4/ vytvořte podadresář `students/` a uložte do něho tento kód <https://ufal.mff.cuni.cz/~hladka/2024/docs/students.h.R>. Kód načítá data a rozděljuje je do trénovacích a testovacích podmnožin. Vaším úkolem je pokračovat v kódu analogicky k řešení úlohy v [tutoriálu](#), a sice

- pomocí algoritmu Naive Bayes natrénovat model na trénovacích datech `train` a otestovat ho na testovacích datech `test`. Matici konfuze a hodnoty Accuracy, Precision, Recall a F1 vložte do vašeho sdíleného dokumentu.
- pomocí algoritmu Decision Trees natrénovat model na trénovacích datech `train` a otestovat ho na testovacích datech `test`. Matici konfuze a hodnoty Accuracy, Precision, Recall a F1 vložte do vašeho sdíleného dokumentu.

Příklad 4.3

Zopakujte si obsah přednášek 1-4, protože v pátek 3.5. budete psát test. V testu budou otázky na vysvětlení pojmů, případně nějaké drobné počítání. Otázky na programování v R v něm nebudou.

Termín odevzdání: 2.5.2024 15:00