

Úkol C – Klasifikace srdečních arytmií – Arrhythmia data

Cíl

Cílem třetího úkolu je seznámit se s praktickým použitím klasifikačních metod na příkladě klasifikace pacientů se srdečními arytmiemi.

Data

Arrhythmia data byla vytvořena pro účely klasifikace normálního průběhu EKG a 15 arytmií. Informace o pacientech spolu s průběhy 12-ti svodového EKG byly popsány pomocí 279 příznaků. Mezi nimi je prvních 15, které představují obecné patientské charakteristiky, a jejich popis je tabulce níže. Zbytek jsou parametry z jednotlivých EKG svodů popsané pomocí délek segmentů a jejich amplitud, a pak specifické jevy typické pro arytmiie.

Klasifikace pacientů podle arytmií

Arrhythmia data obsahují informace o pacientech s různými typy arytmií. Pro účely třetího úkolu uvažujte pouze rozdělení pacientů na pacienty s arytmií a pacienty bez arytmií. Pokuste se vytvořit takový klasifikátor, který by predikoval, zda pacient trpí srdeční arytmií nebo ne; tedy jakýsi screeningový klasifikátor pro srdeční arytmiie – pacientovi je nahráno EKG, extrahují se z něj příznaky a váš klasifikátor má rozhodnout, zda pacient trpí arytmií a má být důkladněji vyšetřen, nebo arytmií netrpí a není potřeba ho dále vyšetřovat. Klasifikátor zhodnoťte z různých praktických i teoretických hledisek.

Požadované kroky analýzy

- Vyberte alespoň dva klasifikátory vhodné pro tento typ dat. Podle čeho budete vybírat? [2 b]
- Na vhodné podmnožině dat vybrané klasifikátory natrénujte. Jaké úspěšnosti dosahují? [2 b]
- Lze pro učení klasifikátorů nějak využít znalosti, které jste získali o datech v předchozích úkolech (shluková analýza)? Pokud ano, jakým způsobem? [2 b]
- Podle jakých příznaků se klasifikátor rozhoduje? Dává to smysl? Lze na základě vaší analýzy omezit počet měřených příznaků při zachování stejné úspěšnosti klasifikace? [4 b]
- Jakou úspěšnost klasifikace očekáváte v hypotetickém reálném nasazení vašeho klasifikátoru, tj. v případě nově příchozího pacienta? [2 b]
- Jaká je pravděpodobnost na základě dat, že nově příchozí pacient bude nemocný? [2 b]
- Jaká je pravděpodobnost, že nově příchozí pacient bude klasifikovaný jako nemocný? Jaká bude naproti tomu pravděpodobnost, že nově příchozí nemocný pacient bude klasifikován jako zdravý. Výsledky diskutujte. [4 b]
- Má ve vašem případě na přesnost klasifikace vliv to, zda je trénovací (testovací) množina vyvážená? Pokud ano, jaký? Vyvažovali jste trénovací (testovací) množinu? Pokud ano, proč a jak? Pokud ne, proč ne? [2 b]

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**. Maximální délka zprávy je 3 stránky. Zprávu ve formátu pdf odevzdejte pomocí UploadSystemu.