

## Úkol C – Predikce komplikací – Point data

### Cíl

Cílem třetího úkolu je seznámit se s praktickým použitím klasifikačních metod na příkladě predikce komplikací spojených s chirurgickým zákrokem.

### Data

Point data byla sesbírána v průběhu několika let na více než 2000 pacientech podstupujících náročné lékařské zákroky. Cílem studie bylo posoudit výhody a nevýhody přesné kontroly glykémie pacienta (koncentrace glukózy v krvi) během operace. Velmi nízká koncentrace glukózy (hypoglykémie) stejně jako velmi vysoká koncentrace glukózy (hyperglykémie) mohou být pro pacienty velmi nebezpečné. V rámci studie byly použity dva protokoly při provádění operace, v prvním případě byla hladina glukózy velmi přesně regulována, zatímco v druhém případě nebyla hladina glukózy kontrolována, ale pouze měřena. Z důvodu posouzení výhod byly pacienti sledováni při pobytu na anesteziologicko - resuscitačním oddělení a veškeré pooperační komplikace byly zaznamenány.

### Predikce komplikací

Point data obsahují informace o operacích a s nimi spojenými komplikacemi. Pro účely třetího úkolu uvažujte pouze rozdělení pacientů na pacienty s komplikacemi během a po operaci a pacienty bez komplikací. Pokuste se vytvořit takový klasifikátor, který by predikoval, zda pacient bude mít zvýšené riziko komplikací během a po operaci; tedy jakýsi screeningový klasifikátor na základě předoperačního vyšetření. 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 má zvýšené riziko komplikací během a po operaci? [2 b]
- Jaká je pravděpodobnost, že nově příchozí pacient bude klasifikovaný jako rizikový? Jaká bude naproti tomu pravděpodobnost, že nově příchozí rizikový pacient bude klasifikován jako bez rizika komplikací. 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.