

V mailu jsou potřebná data k řešení semestrální práce. Jedná se o množinu dvojdimenzionálních vektorů. Vaším úkolem je

- 1) automaticky určit počet tříd
  - a) metodou **shlukové hladiny** (automaticky nalezněte hladinu  $h$ )
  - b) metodou **řetězové mapy** (zkuste několik různých počátků)
  - c) metodou **MAXIMIN**a metody vzájemně porovnat
- 2) metodou k-means rozdělit všechna data do zjištěného počtu tříd
  - porovnat nerovnoměrné binární dělení s přímým dělením do cílového počtu tříd
- 3) na výsledné rozdělení dat do jednotlivých tříd z bodu 2 vyzkoušet iterativní optimalizaci
- 4) na základě informací od učitele (informace o zařazení do jednotlivých tříd  $\omega_i$  z bodu 2 popřípadě informace z bodu 3) natrénovat
  - a) **Bayesův klasifikátor** - tady nepředpokládám explicitně řešení té hranice (kuželosečky) stačí odhadnout parametry jednosložkového normálního rozložení a nějakým dostatečně jemným rastrem ohodnotit body v prostoru (kde se vyskytují trénovací data) kam který bod má největší pravděpodobnost.
  - b) **klasifikátor podle minimální vzdálenosti** - jako v předchozím bodě zakreslete pomocí rastru body v prostoru (trénovacích dat) které klasifikujeme do jednotlivých tříd
  - c) **klasifikátor podle nejbližšího souseda** - vyzkoušejte klasifikaci podle jednoho a podle dvou nejbližších sousedů a podobně jako v předchozím bodě zakreslete pomocí rastru body v prostoru (trénovacích dat) které klasifikujeme do jednotlivých tříd
  - d) **klasifikátor s lineárními diskriminačními funkcemi** - porovnejte potřebný počet iterací při použití Rosenblattova alg., metody konstantních přírůstků a upravené metody konstantních přírůstků pro několik zvolených konstant učení. Podobně jako v předchozím bodě zakreslete pomocí rastru body v prostoru (trénovacích dat) které klasifikujeme do jednotlivých tříd a které případně klasifikovat nemůžeme (diskutujte proč)

Úlohu můžete řešit v libovolném jazyce. Ale pokud zvolíte nestandardní řešení (mimo C++, MATLAB, Python) tak mě musíte před odevzdáním informovat, abych se na vás připravil.

Výsledky a postup prezentujte krátkou zprávou - záměrně jsem zvolil dimenzi 2, aby se řešení úloh dalo pěkně zobrazit (očekávám spoustu obrázků).