

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií

# Umelá Inteligencia

Zadanie 4a – klasifikácia

ZS 2021/2022

Samuel Kováč

Študijný program: B-INFO4  
Vyučujúci: Ing. Ivan Kapustík  
Cvičenia: Streda 13:00  
December 2021

## Úloha :

Máme 2D priestor, ktorý má rozmery X a Y, v intervaloch od -5000 do +5000. V tomto priestore sa môžu nachádzať body, pričom každý bod má určenú polohu pomocou súradníc X a Y. Každý bod má unikátne súradnice (t.j. nemalo by byť viac bodov na presne tom istom mieste). Každý bod patrí do jednej zo 4 tried, pričom tieto triedy sú: red (R), green (G), blue (B) a purple (P). Na začiatku sa v priestore nachádza 5 bodov pre každú triedu (dokopy teda 20 bodov). Súradnice počiatočných bodov sú:

R: [-4500, -4400], [-4100, -3000], [-1800, -2400], [-2500, -3400] a [-2000, -1400]

G: [+4500, -4400], [+4100, -3000], [+1800, -2400], [+2500, -3400] a [+2000, -1400]

B: [-4500, +4400], [-4100, +3000], [-1800, +2400], [-2500, +3400] a [-2000, +1400]

P: [+4500, +4400], [+4100, +3000], [+1800, +2400], [+2500, +3400] a [+2000, +1400]

Vašou úlohou je naprogramovať klasifikátor pre nové body – v podobe funkcie `classify(int X, int Y, int k)`, ktorá klasifikuje nový bod so súradnicami X a Y, pridá tento bod do nášho 2D priestoru a vráti triedu, ktorú pridelila pre tento bod. Na klasifikáciu použijete k-NN algoritmus, pričom k môže byť 1, 3, 7 alebo 15.

Na demonštráciu Vášho klasifikátora vytvorte testovacie prostredie, v rámci ktorého budete postupne generovať nové body a klasifikovať ich (volaním funkcie `classify`). Celkovo vygenerujte 20000 nových bodov (5000 z každej triedy). Súradnice nových bodov generujte náhodne, pričom nový bod by mal mať zakaždým inú triedu (dva body vygenerované po sebe by nemali byť rovnakej triedy):

- R body by mali byť generované s 99% pravdepodobnosťou s  $X < +500$  a  $Y < +500$
- G body by mali byť generované s 99% pravdepodobnosťou s  $X > -500$  a  $Y < +500$
- B body by mali byť generované s 99% pravdepodobnosťou s  $X < +500$  a  $Y > -500$
- P body by mali byť generované s 99% pravdepodobnosťou s  $X > -500$  a  $Y > -500$

(Zvyšné jedno percento bodov je generované v celom priestore.)

Návratovú hodnotu funkcie `classify` porovnávajte s triedou vygenerovaného bodu. **Na základe týchto porovnaní vyhodnoťte úspešnosť** Vášho klasifikátora pre daný experiment.

Experiment vykonajte 4-krát, pričom zakaždým Váš klasifikátor použije iný parameter k (pre k = 1, 3, 7 alebo 15) a vygenerované body budú pre každý experiment rovnaké.

## Riešenie :

Na začiatku programu vygenerujem 20 000 bodov - pre každú farbu 5000, pričom súradnice týchto bodov vyberám podľa pravdepodobností uvedených v zadaní. Tieto body rozdelím podľa farby do štyroch *zásobníkov* z ktorých ich budem neskôr postupne vyberať. Tiež si vytvorím 20 bodov určených zo zadania. Po vygenerovaní všetkých bodov prejdem do hlavnej slučky programu.

Na začiatku slučky náhodne vyberiem farbu pričom nesmie byť rovnaká ako farba posledného klasifikovaného bodu. Po vybraní farby vytiahnem bod zo zásobníka rovnakej farby a posuniem ho do funkcie *classify()*. Klasifikačná funkcia bodu priradí farbu podľa ostatných doteraz priradených bodov. Funkcia vypočíta vzdialenosť bodu od všetkých ostatných priradených bodov a vyberie *k* najbližších. Bodu bude pridelená farba ktorá je najfrekventovanejšia medzi najbližšími bodmi. Následne porovnáme túto farbu z pôvodnou farbou bodu zo zásobníka. Ak sa farby nezhodujú, pripočítame chybu. Túto slučku opakujeme až dokým nebudú prázdne všetky zásobníky.

Tento proces sa zopakuje pre všetky hodnoty *k* (1,3,7,15) s rovnakými bodmi v zásobníkoch. Na konci zostrojíme grafy podľa vygenerovaných polí prostredníctvom knižnice *matplotlib*.

## Experiment

```
Pri klasifikácii (k = 1) bolo nájdených 4279 chýb, klasifikácia trvala 143 s  
Pri klasifikácii (k = 3) bolo nájdených 4455 chýb, klasifikácia trvala 138 s  
Pri klasifikácii (k = 7) bolo nájdených 4696 chýb, klasifikácia trvala 143 s  
Pri klasifikácii (k = 15) bolo nájdených 8005 chýb, klasifikácia trvala 124 s
```

