**FIIT STU**

**Umelá Inteligencia**

**Zadanie 3a**

**Klasifikácia pomocou KNN algoritmu**

Meno: Jakub Gašparín

Cvičenie: Utorok 14:15:40

Cvičiaci: Ing. Martin Komák, PhD.

Obsah

[1.0 Zadanie 2](#_Toc121395159)

[2.0 Riešenie 3](#_Toc121395160)

[2.1 Úvodné zadanie prvých 20 bodov 3](#_Toc121395161)

[2.2 Generovanie správnych bodov 5](#_Toc121395162)

[2.3 Generovanie bodov mimo daného rozsahu 6](#_Toc121395163)

[2.4 Riešenie KNN 7](#_Toc121395164)

[3.0 Výsledky 8](#_Toc121395165)

[4.0 Záver 10](#_Toc121395166)

# 1.0 Zadanie

Máme zadané 2d priestor s rozmermi -5000 až +5000. V tomto priestore máme dopredu vygenerované body:

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]

Ďalej budeme generovať body podľa nasledujúcich pravidiel:

* 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.)

Úlohou je navrhnúť program s použitím KNN algoritmu na klasifikovanie všetkých bodov v priestore tak, že rovnako zafarbené body budú spolu pri sebe.

# 2.0 Riešenie

## 2.1 Úvodné zadanie prvých 20 bodov

Najprv si musíme vytvoriť triedu kde budeme si pamätať všetky možné body. Obrázok, na ktorom je text

Automaticky generovaný popis

V tejto triede si budem pamätať ich x-ovú, y-ovú súradnicu a ich farbu.

Keďže máme prvých 20 bodov pevne daných, musím ich aj vypísať ručne:

Obrázok, na ktorom je text

Automaticky generovaný popis

## 2.2 Generovanie správnych bodov

Generovanie robím cez jeden for cyklus. Obrázok, na ktorom je text

Automaticky generovaný popis

Jednotlivé funkcie mi vrátia body ktoré sú na 100% v danom rozsahu pre jednotlivé farby. Obrázok, na ktorom je text

Automaticky generovaný popis

## 2.3 Generovanie bodov mimo daného rozsahu

Tieto body som generoval rovnako, iba som tam pridal podmienku že musia byť mimo rozsahu, inak nevráti žiadnu hodnotu

Obrázok, na ktorom je text

Automaticky generovaný popis

Obrázok, na ktorom je text

Automaticky generovaný popis

## 2.4 Riešenie KNN

Samotné riešenie sa deje v jednom for cykle ktorý zbehne iba 4-krát, raz pre každú jednu farbu. Farby riešim postupne, teda idem red>green>blue>purple>red>green... atď.

Obrázok, na ktorom je text

Automaticky generovaný popis

Keď zabezpečím, že ide po porade tak prejdem na samotné riešenie algoritmu.

Obrázok, na ktorom je text

Automaticky generovaný popis

Keď si zistím, o akú farbu ide, tak idem riešiť samotný KNN.

Obrázok, na ktorom je text

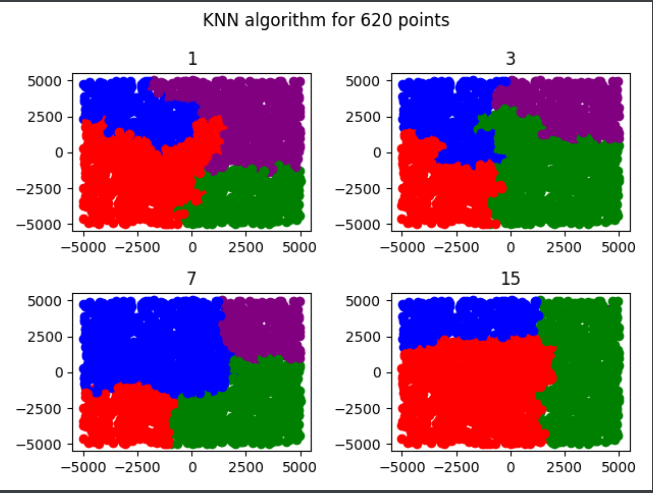
Automaticky generovaný popis

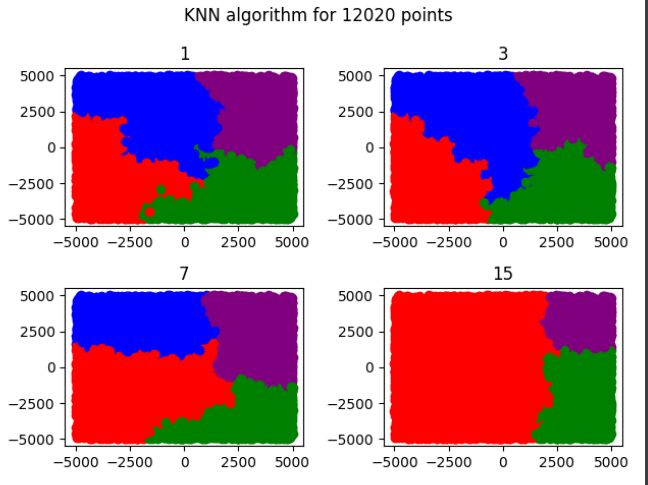
V týchto dátach si pamätám všetky body s ktorými chcem porovnávať ich k najbližších susedov. Vypočítam si ich euklidovskú vzdialenosť, zoradím ich a zoberiem k najbližších. Tie potom si zapamätám. Tento postup opakujem až dovtedy, dokiaľ neprejdem všetky body.

# 3.0 Výsledky

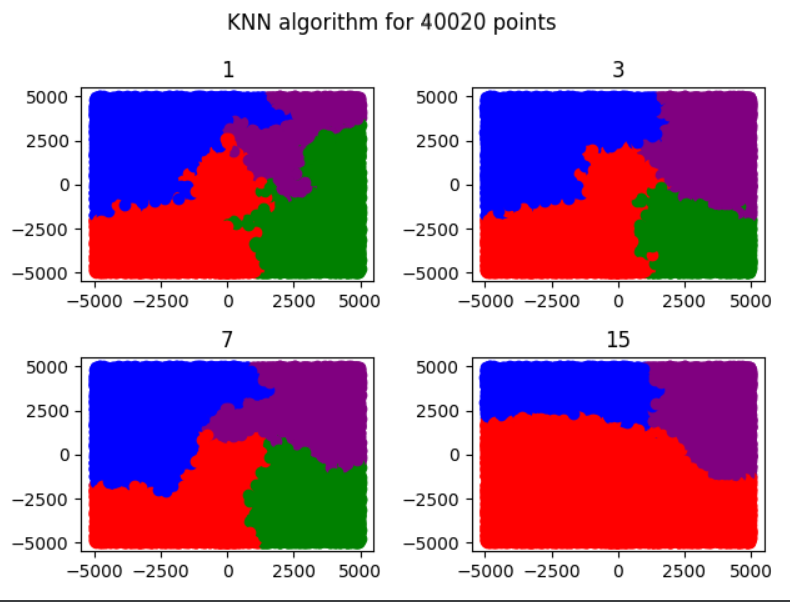
Program som spustil na troch vstupoch: na malom, stredne veľkom a so v zadaní požadovaných 40 000 bodmi.

Na malom vstupe som použil 600 bodov + 20 počiatočních. Celý program zbehol do jednej minúty



Na stredne veľkom vstupe som testoval s 12000 + 20 počiatočných bodov. 

Na najväčšom vstupe som testoval so 40000 + 20 počiatočnými bodmi. Celkovo program trval okolo hodiny a pol.



Program bežal približne 75 minút.

# 4.0 Záver

Najlepšie výsledky som mal keď som testoval s KNN = 7. V tomto grafe dáta vyzerali najlepšie rozdelené. V KNN 1 boli dáta občas veľmi neusporiadané, v KNN 5 to vyzeralo oveľa lepšie ale občas sa našli nejaké výchylky a KNN 15 bol príliš veľmi agresívny: „zhltol“ jednu alebo viaceré farby.