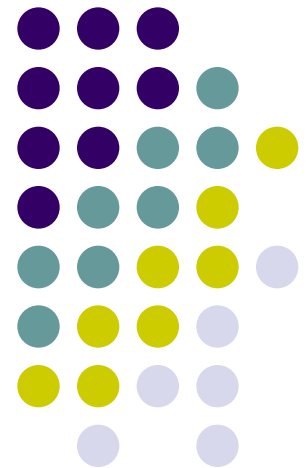
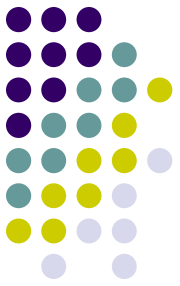


# Основи географічного моделювання (4 курс, 2 семестр)

---





# План курсу

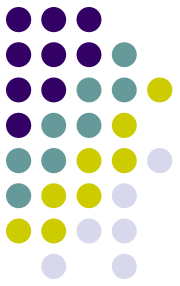
- Продовження вивчення таксономічного аналізу
- Концепція поля в сучасній географії
- Теорія графів та її застосування в географічних дослідженнях



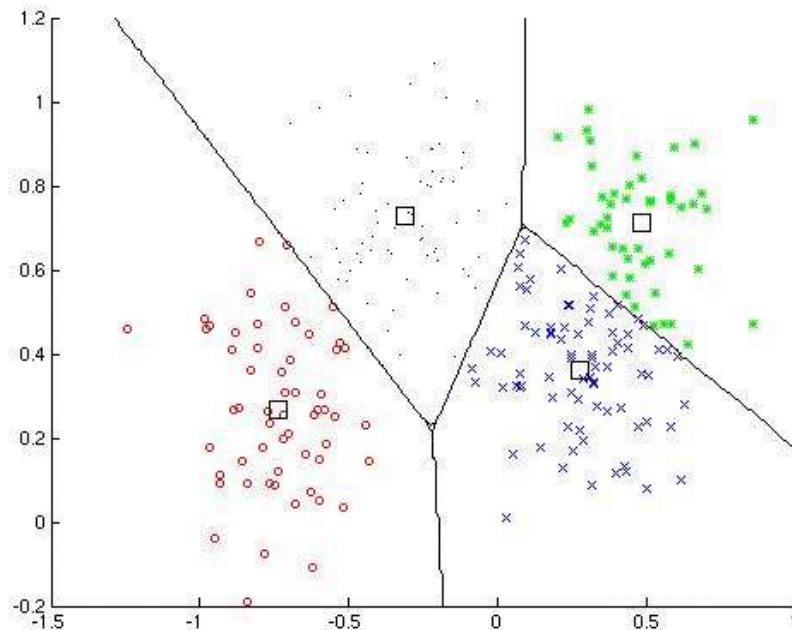
Основи таксономічного аналізу

# МЕТОД К-СЕРЕДНІХ, МЕТОД ISODATA

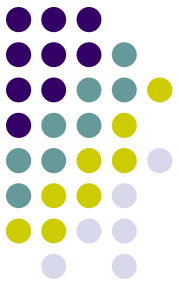
# Сутність методу к-середніх



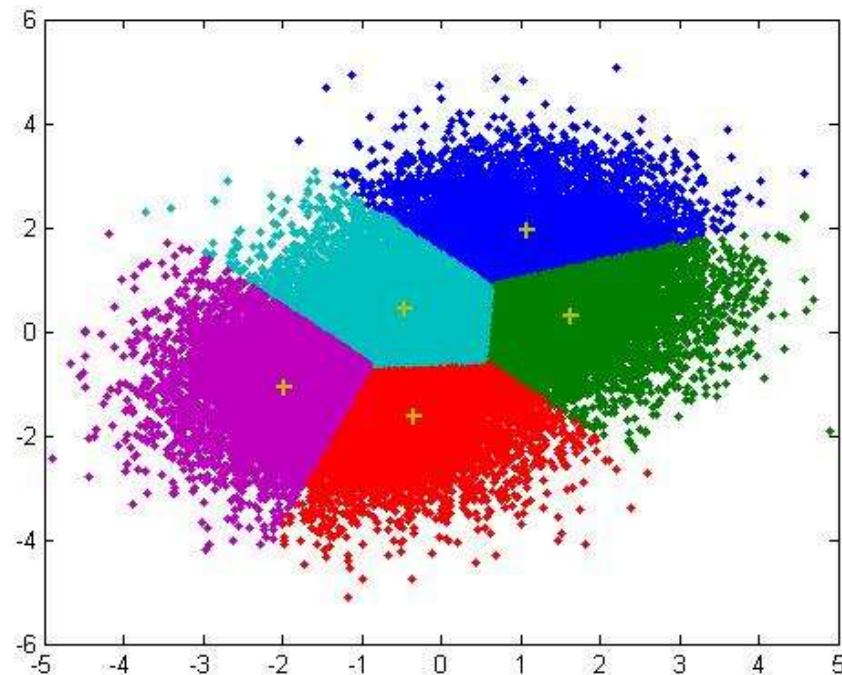
- **Метод К-середніх** — метод визначення приналежності елементів кластерам (таксонам) за допомогою мінімізації різниці між елементами кластера і максимізації відстані між кластерами.



# Сутність методу



- Кожен кластер ідентифікується за допомогою **центроїду**, який обчислюється як осереднений вектор від усіх його елементів.

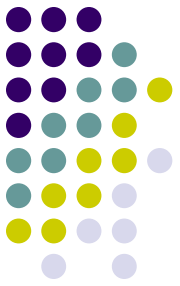




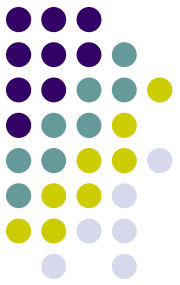
# Сутність методу

- « $K$ » означає довільну апріорно задану кількість точок, що використовуються для формування початкових значень процесу кластеризації. Алгоритм  $K$ -середніх обчислює квадрати євклідових відстаней між записами даних в кластері і вектором середніх значень даного кластера. Метод сходиться, видаючи остаточний набір з  $K$  кластерів, коли згадана сума мінімізована.

# Сутність методу



- Алгоритм  $K$ -середніх діє ітераціями. Спочатку кластери формуються довільним чином за допомогою перших  $K$  спостережень, а кожне з подальших спостережень приписується до найближчого з цих  $K$  кластерів. Після цього переобчислюються центроїди всіх кластерів.



# Сутність методу

- Можливі умови зупинки циклу:
  - Кількість ітерацій
  - Склад груп не змінюється (або не перевищує поріг збіжності)
  - Місцеположення центроїдів не змінюється





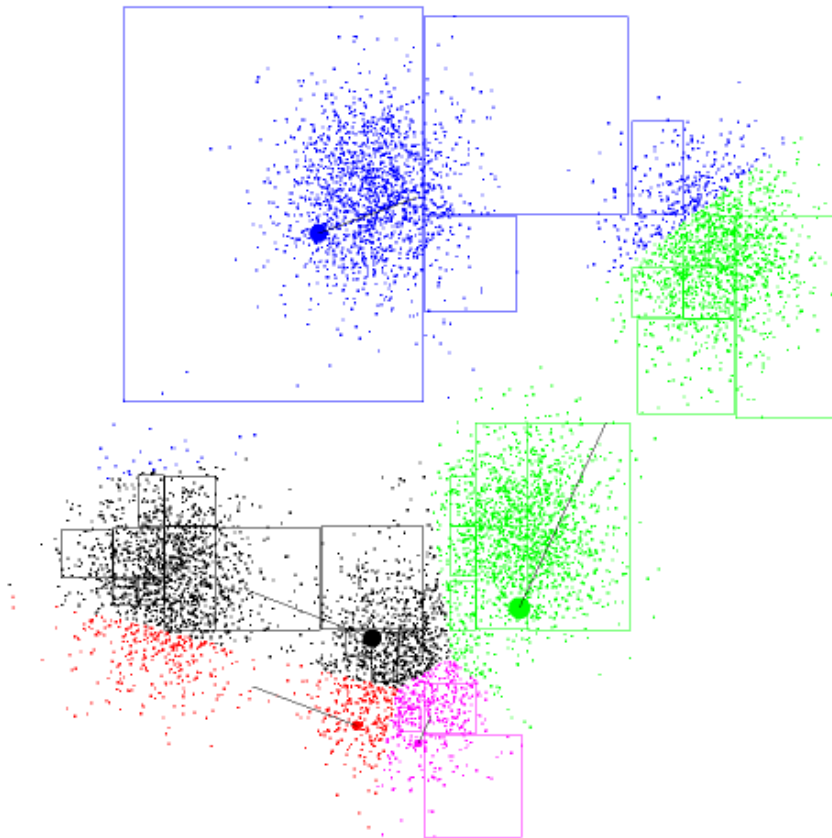
# Сутність методу

- Недоліки:
  - Результат кластеризації залежить від вибору стартових елементів та кількості ітерацій
  - Значення  $K$  обирається поза алгоритмом
  - Кластери не перетинаються (відсутня можливість формування “перехідних зон” між кластерами – елементи виборки жорстко відносяться до одного з кластерів)
  - Відсутність можливості прослідкувати етапи об’єднання елементів у кластери
  - Невідома кількість часу для повного перерозподілу елементів між кластерами

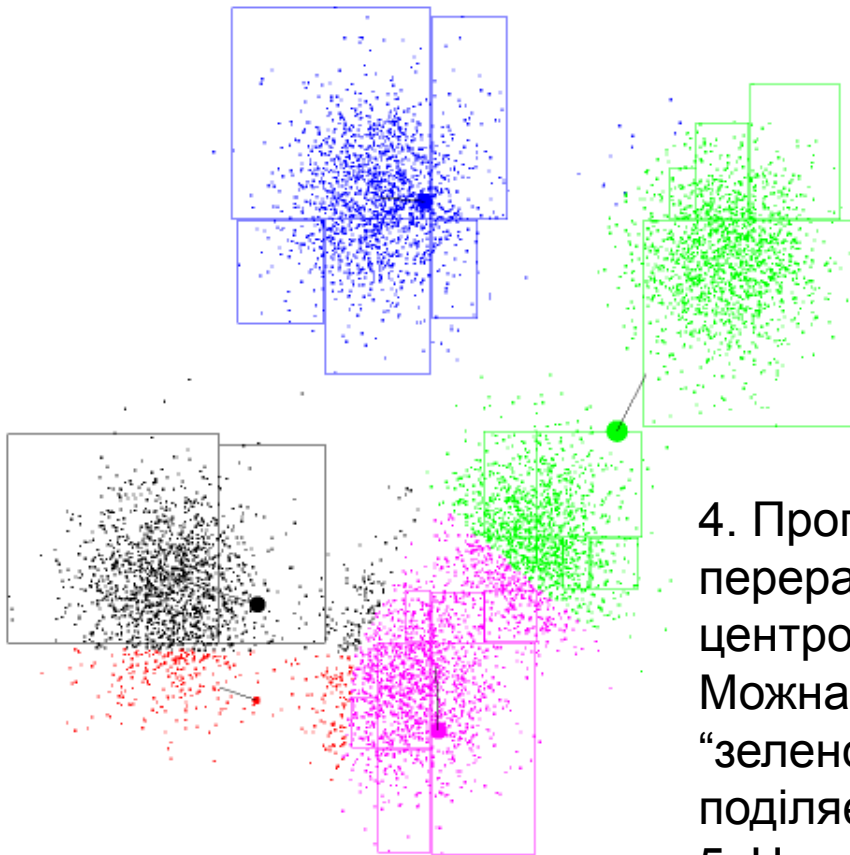
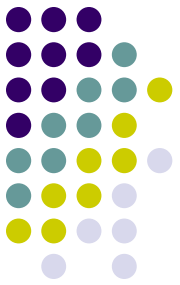


# Приклад дії алгоритму

- Прослідкуємо алгоритм роботи методу на прикладі 8000 точок у двовимірному просторі.
  1. На початку **K-means** вибрав 5 випадкових точок-центроїдів, що показані великими синіми, зеленими, червоними, чорними, і бузковими крапками.



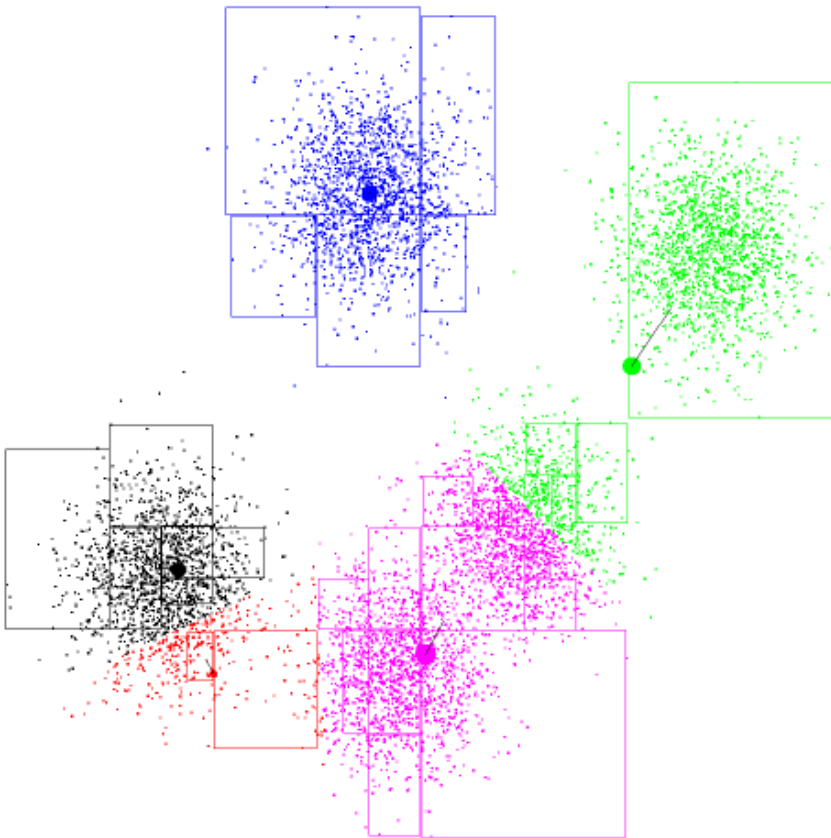
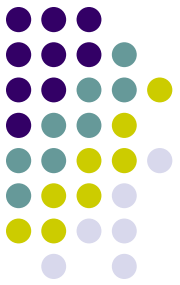
2. На даному етапі визначається належність решти точок до того, чи іншого кластеру (точки фарбуються відповідними кольорами)
3. Після визначення належності перевизначаються центроїди кожного з кластерів (показано лініями)



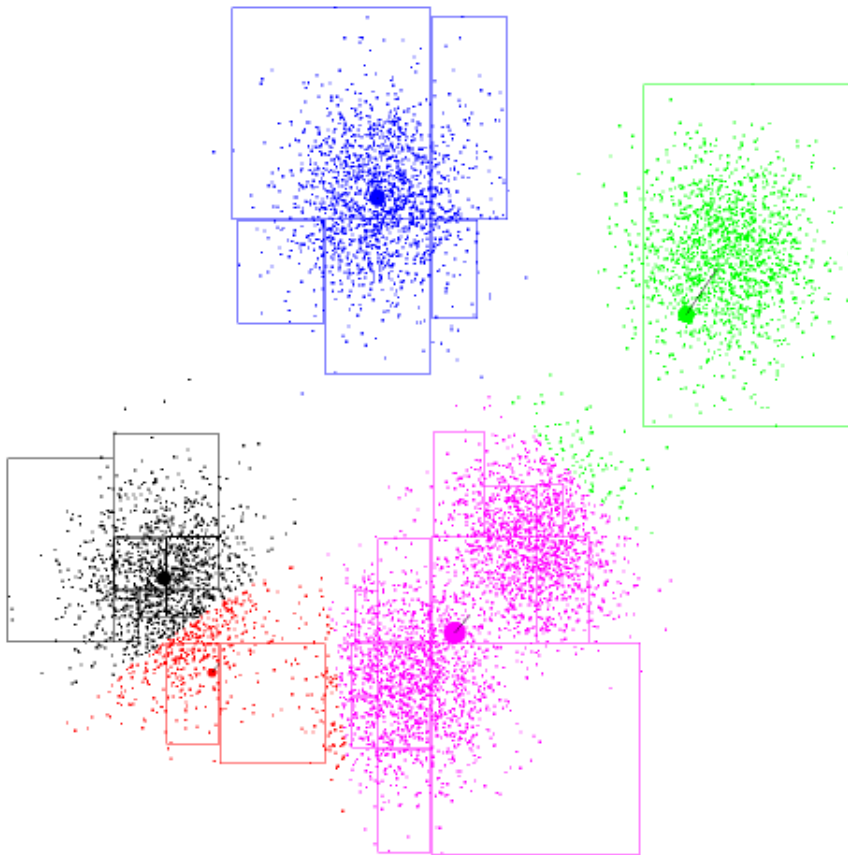
4. Програма перемістила центри, і повторно перерахувала відстань від крапок до центроїдів.

Можна побачити, що , після переміщення “зеленого” центроїду, його межа з синім вже не поділяє “природні” угруповання точок.

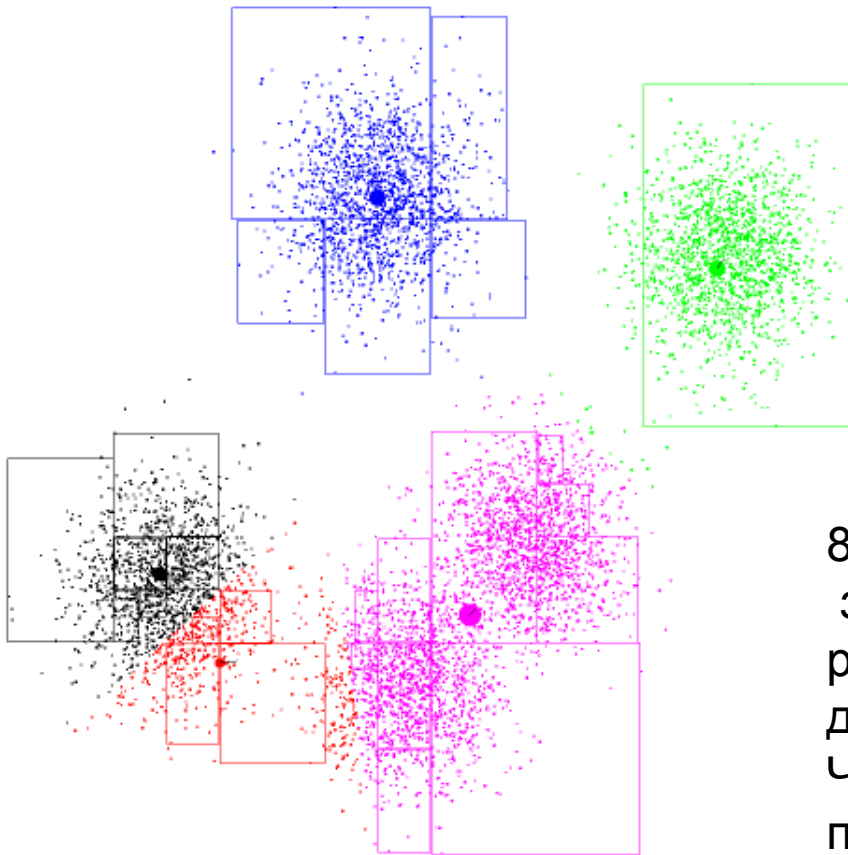
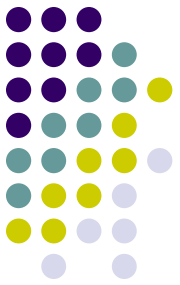
5. Чорний та червоний центроїди продовжують Зміщуватись вліво, поділяючи одне й теж “природне” угруповання. Це також зумовлює наявність дуже великого “зеленого” кластеру



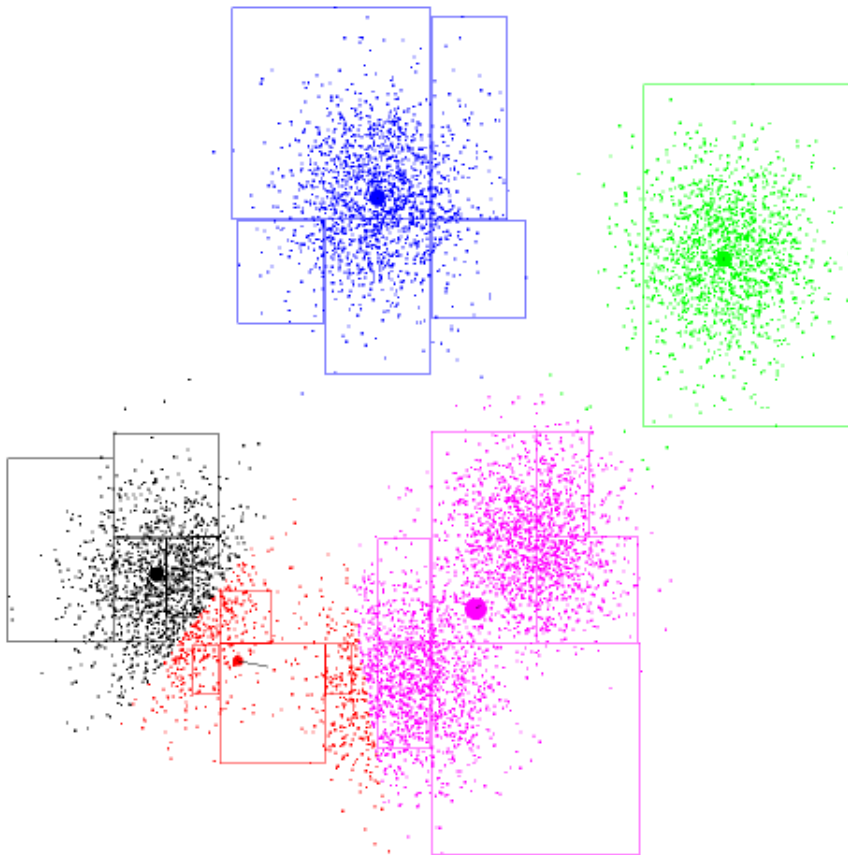
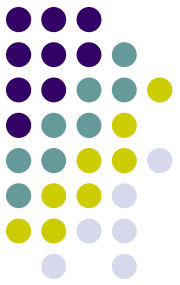
6. Спостерігається зміщення межі між зеленим та бузковим кластером вгору та праворуч. Також спостерігається зміщення межі між чорним та червоним кластером вниз та праворуч



7. Ще одна ітерація з зазначеними тенденціями

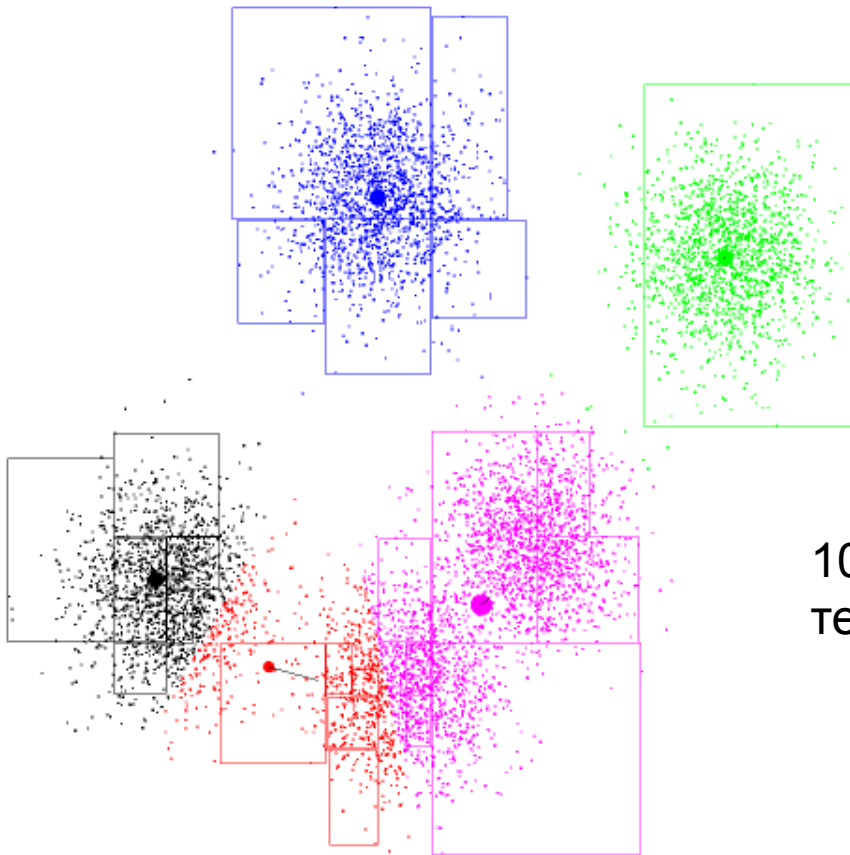


8. Зараз межі синій-зелений та  
зелений-бузковий кластери є  
репрезентативними (знаходяться там,  
де їм необхідно бути)  
Червоний кластер змістився ще більше  
праворуч

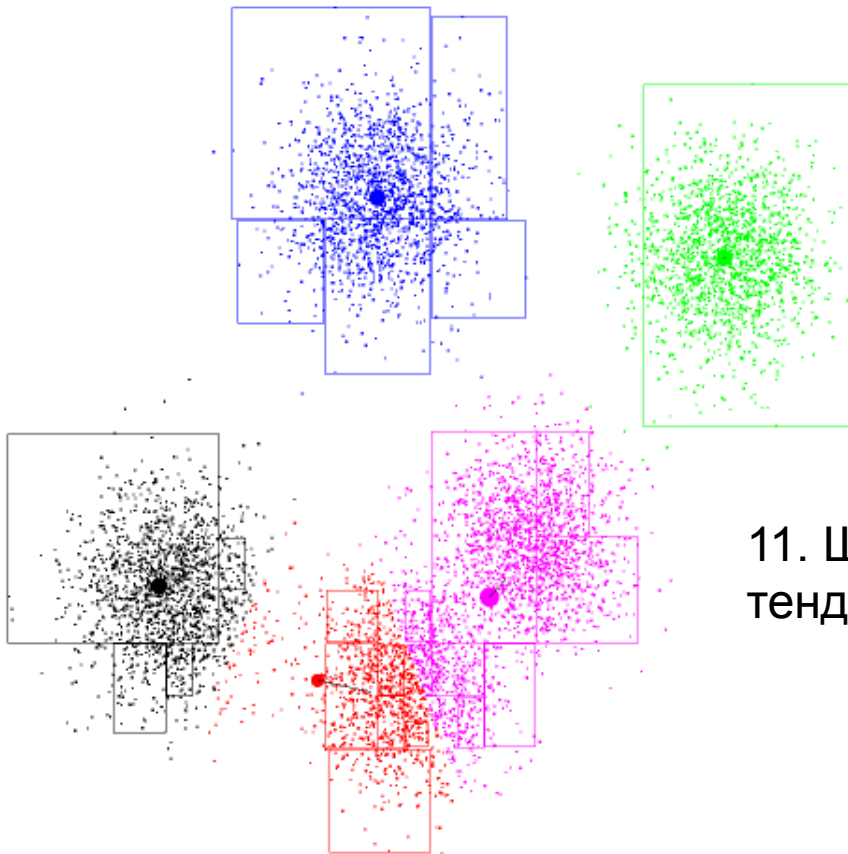


9. Червоний кластер стало зміщується праворуч та захоплює точки бузкового кластеру

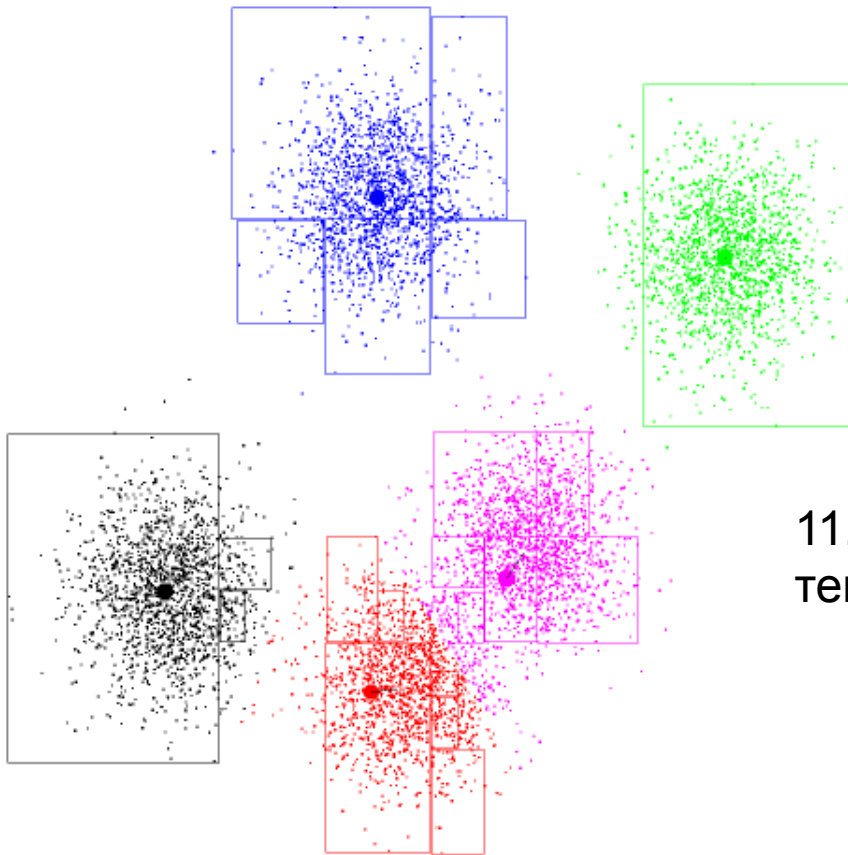




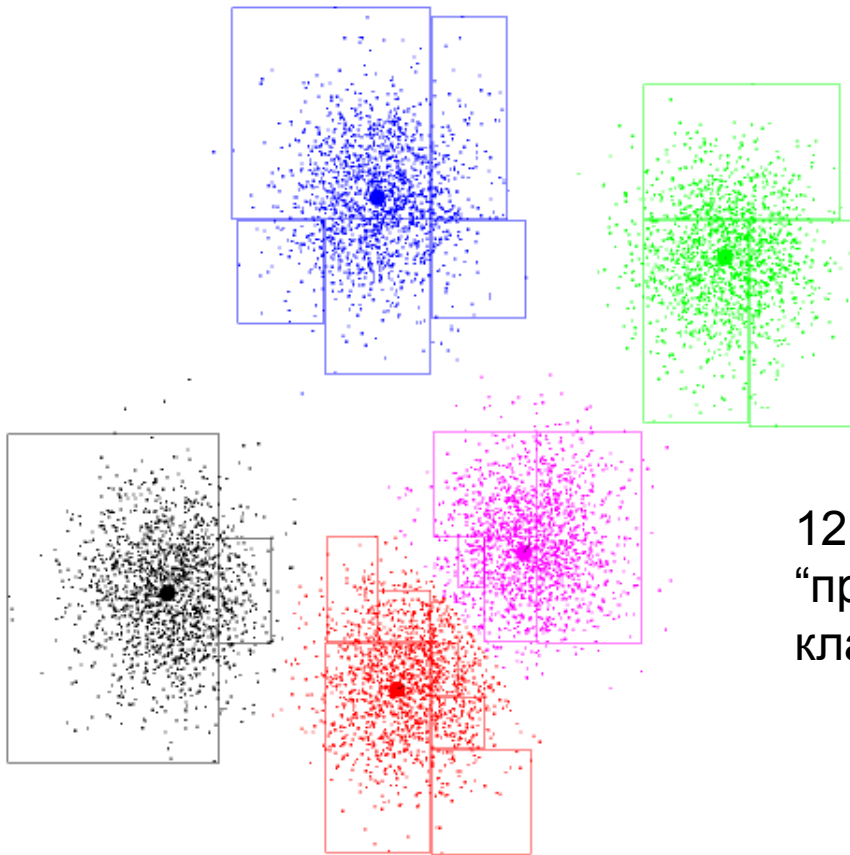
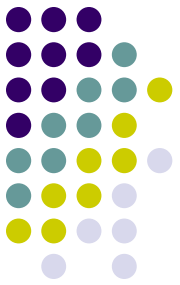
10. Ще одна ітерація з зазначеними тенденціями



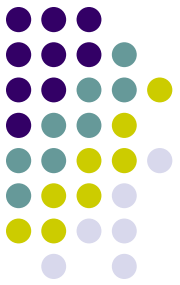
11. Ще одна ітерація з зазначеними тенденціями



11. Ще одна ітерація з зазначеними тенденціями



12. Червоний повністю захопив  
“природне” угруповання. Виділення  
кластерів закінчено



# Метод ISODATA

- Iterative Self-Organising Data Analysis Technique .... А – ітеративний спосіб аналізу даних, що здатний самоорганізовуватися
- Застосовується, в першу чергу, в аналізі даних дистанційного зондування Землі.



# Метод ISODATA

- Метод є схожим з “к-середніх”.
- Для визначення перших центрів кластерів досліджується статистичний розподіл в кожному спектральному діапазоні (макс, мін, середнє, стандартне відхилення).
- Потім весь простір спектральний ознак поділяється на необхідну кількість діапазонів, в кожному з яких визначаються центроїди. Далі – аналогічно “к-середнім”