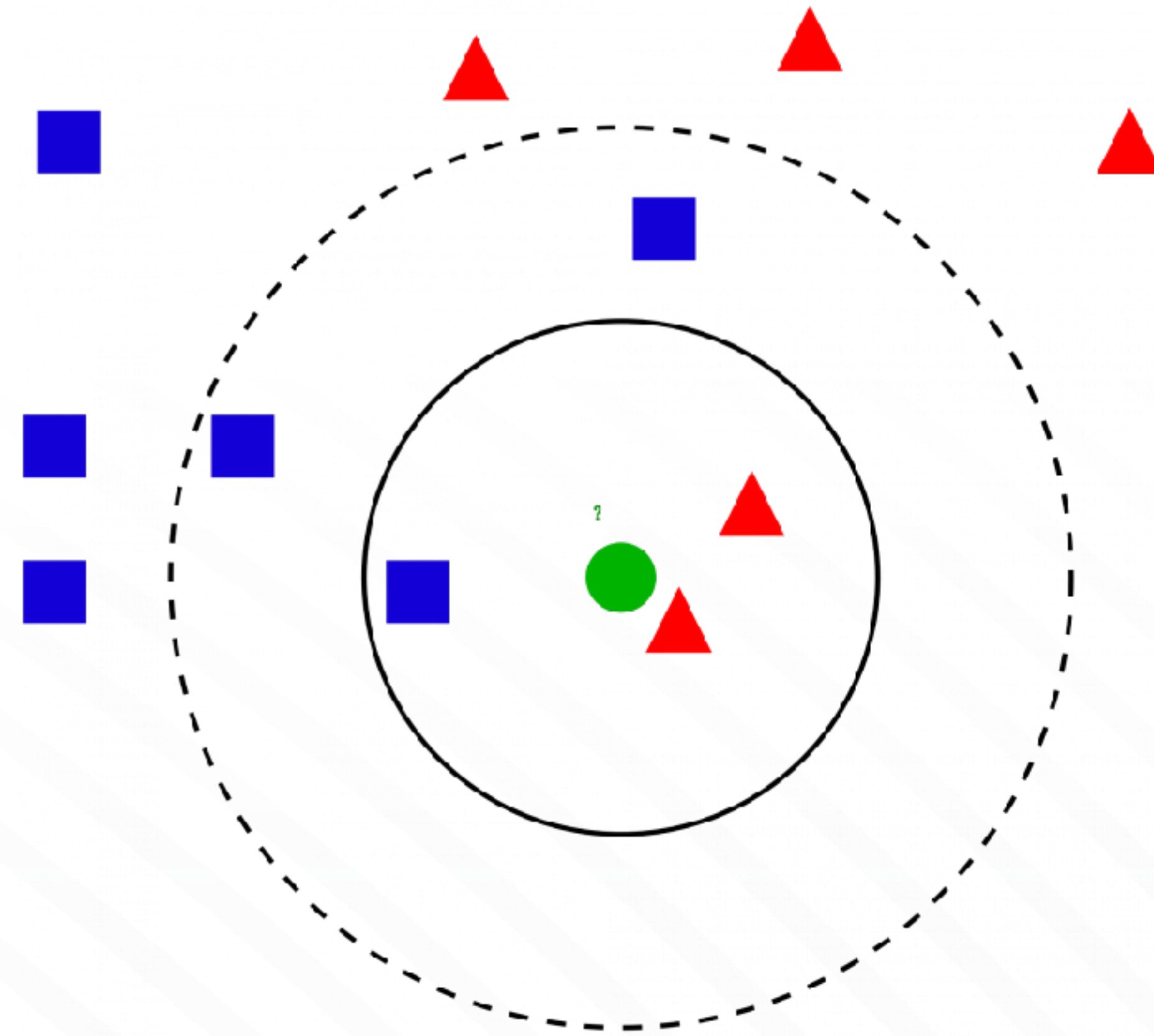




مقدمة في تعلم الآلة

خوارزميات التصنيف (k-nearest neighbors algorithm (KNN))





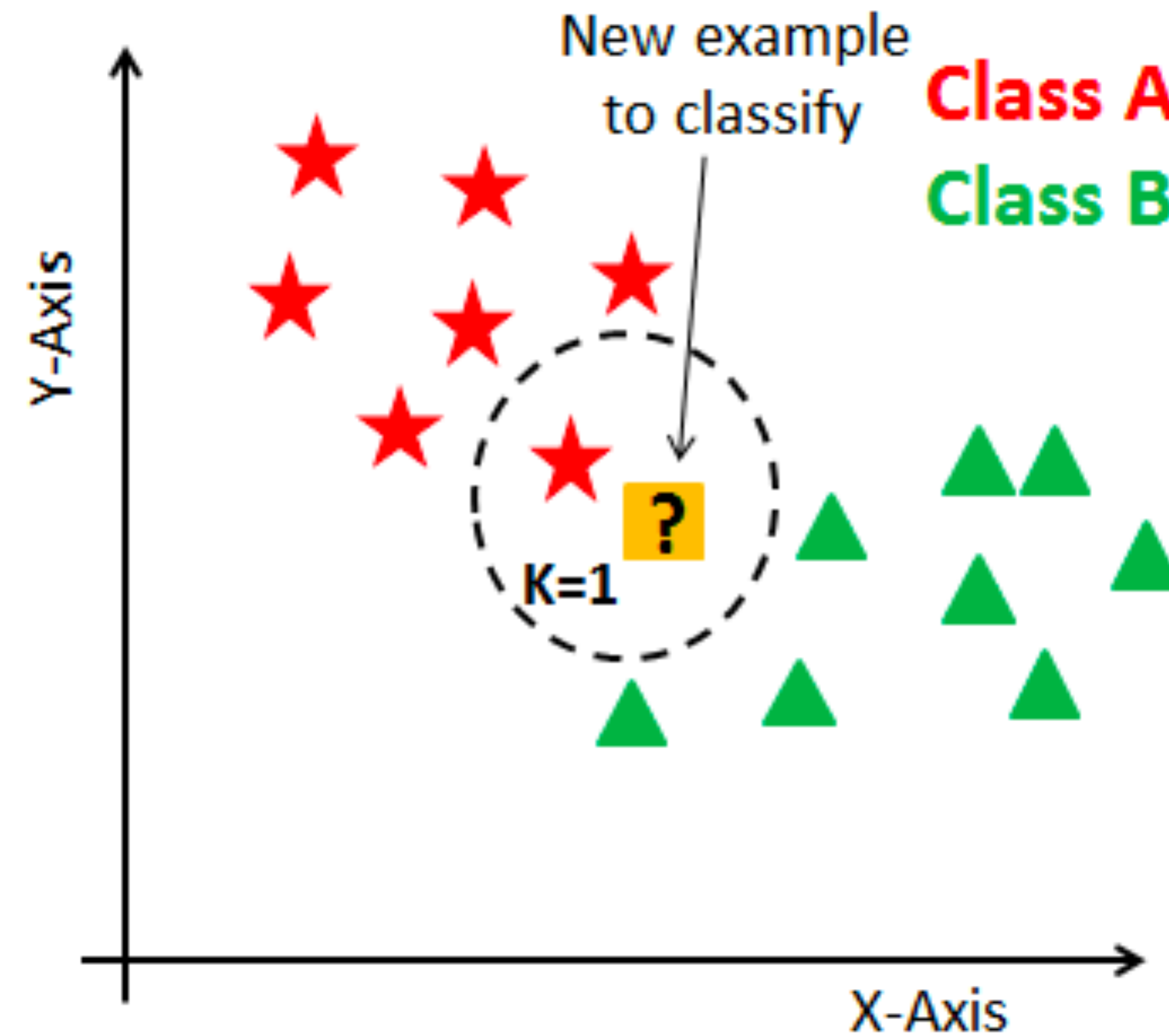
خوارزميات التصنيف (k-nearest neighbors algorithm (KNN))

- ظهرت في بداية 1950s ولكن كان الاستخدام الفعلي لها 1960s عندما تطورت قدرة أجهزة الكمبيوتر على المعالجة.
- أحد أشكال الخوارزميات المستخدمة في Supervised Learning، وتهدف هذه الخوارزمية إلى التنبؤ بأي قيمة جديدة عن طريق البحث عن أقرب نقطة في مجموعة بيانات التدريب.
- تستخدم مفهوم Closeness والمقصود به distance metric.

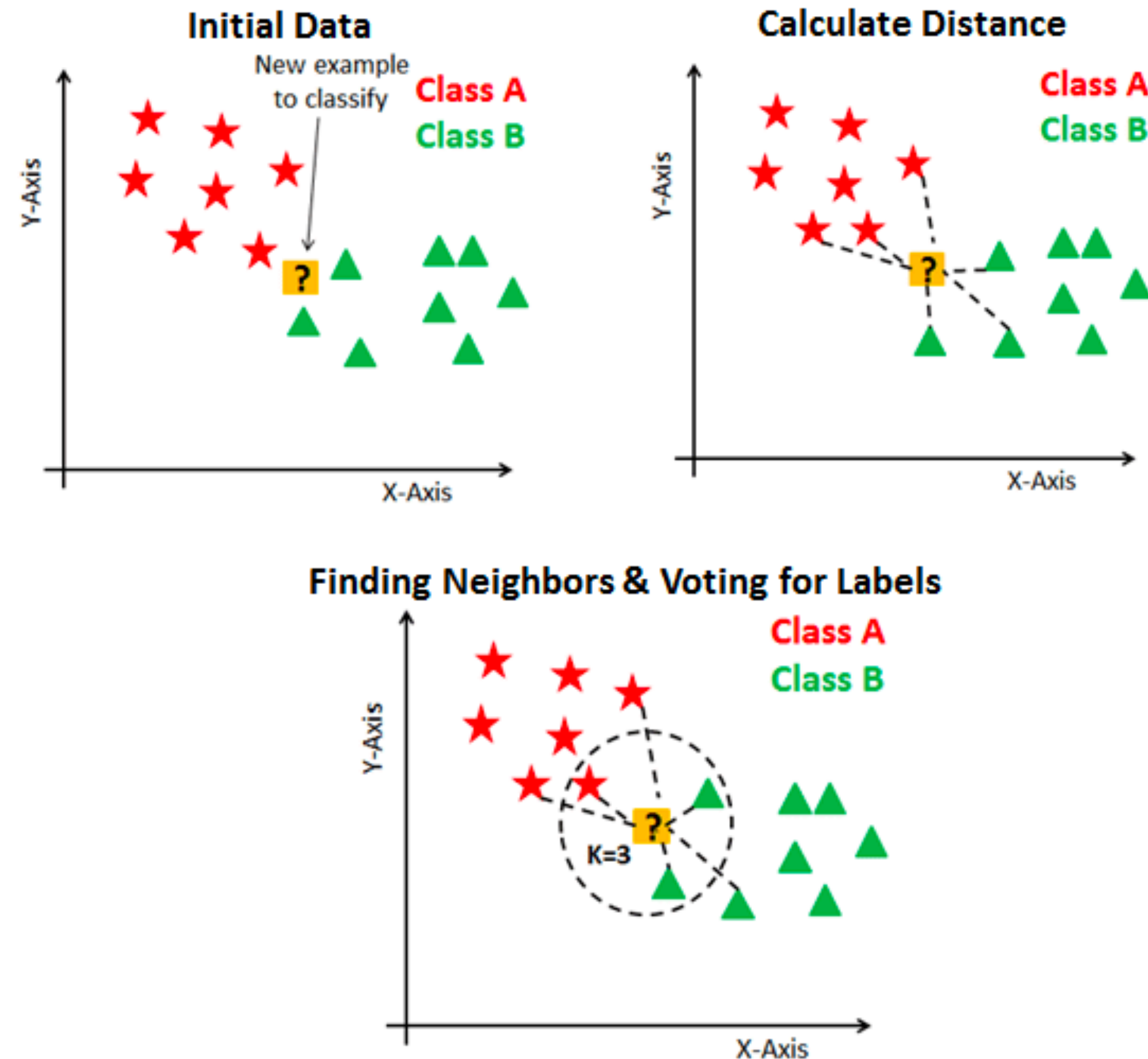
أمثلة لاستخدامات الخوارزمية:

- اكتشاف الأنماط (pattern recognition) خصوصا في مجال retail.
- أنظمة التوصية (Recommended Systems).

خوارزميات التصنيف (k-nearest neighbors algorithm (KNN))



خوارزميات التصنيف (k-nearest neighbors algorithm (KNN))



الخطوات الأساسية في KNN:

- حساب المسافة
- البحث عن أقرب neighbors
- التصويت على labels

خوارزميات التصنيف (k-nearest neighbors algorithm (KNN))

Distance functions

Euclidean

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

تعتبر خوارزمية KNN من نوع distance-based وتقوم بإعطاء equal weight لجميع المتغيرات وبالتالي تعاني من ضعف الدقة.

Manhattan

$$\sum_{i=1}^k |x_i - y_i|$$

لذلك يتم حل هذه المشكلة عن طريق:

- تعيين attribute weighting.

- حذف القيم الغير مفيدة pruning noisy data.

Minkowski

$$\left(\sum_{i=1}^k (|x_i - y_i|)^q \right)^{1/q}$$



خوارزميات التصنيف (k-nearest neighbors algorithm (KNN))

- كيف يتم تحديد حساب المسافة للبيانات الغير رقمية (nominal or categorical) مثل اللون؟
- كيف يمكن تحديد قيمة K بشكل جيد (number of neighbors)؟

Resources

- Data Science: The Big Picture [<https://app.pluralsight.com/library/courses/data-science-big-picture/table-of-contents>].
- Introduction to Data Science [<https://link.springer.com/book/10.1007/978-3-319-50017-1>].
- Data Mining: Concepts and Techniques [<https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>].