

# First Hit

Supervised Learning || Classification || K-NN ||  
Training & Testing || Learning method || Non-  
Parametric Statistical Learning || Voronoi Diagram ||  
Data Visualization

---

Ahmed Kowsher

AI Scientist at Hishab Ltd, Bangladesh

Research Scientist at Voicebridge, Germany

Formar AI Engineer, NKSoft, Dalas, USA

# Outline & Content

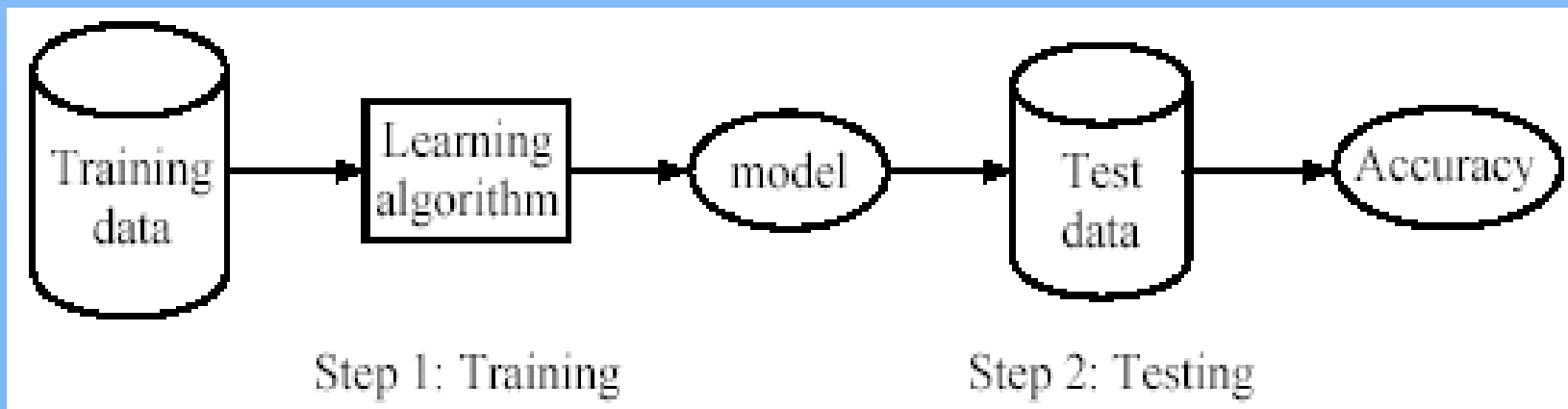
---

- ▣ Supervised Learning
- ▣ Pre-processing ( training and testing)
- ▣ How to learn a Machine
- ▣ Classifications
- ▣ K-NN
- ▣ Theory of K-NN
- ▣ Graphical view of K-NN
- ▣ Code performance
- ▣ Assignment
- ▣ Next Class work

# Supervised learning process: two steps

- **Learning (training)**: Learn a model using the training data
- **Testing**: Test the model using **unseen** test data to assess the model accuracy

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$



# What do we mean by learning?

- **Given**

- a data set  $D$ ,
- a task  $T$ , and
- a performance measure  $M$ ,

a computer system is said to **learn** from  $D$  to perform the task  $T$  if after learning the system's performance on  $T$  improves as measured by  $M$ .

- In other words, the learned model helps the system to perform  $T$  better as compared to no learning.

# An example

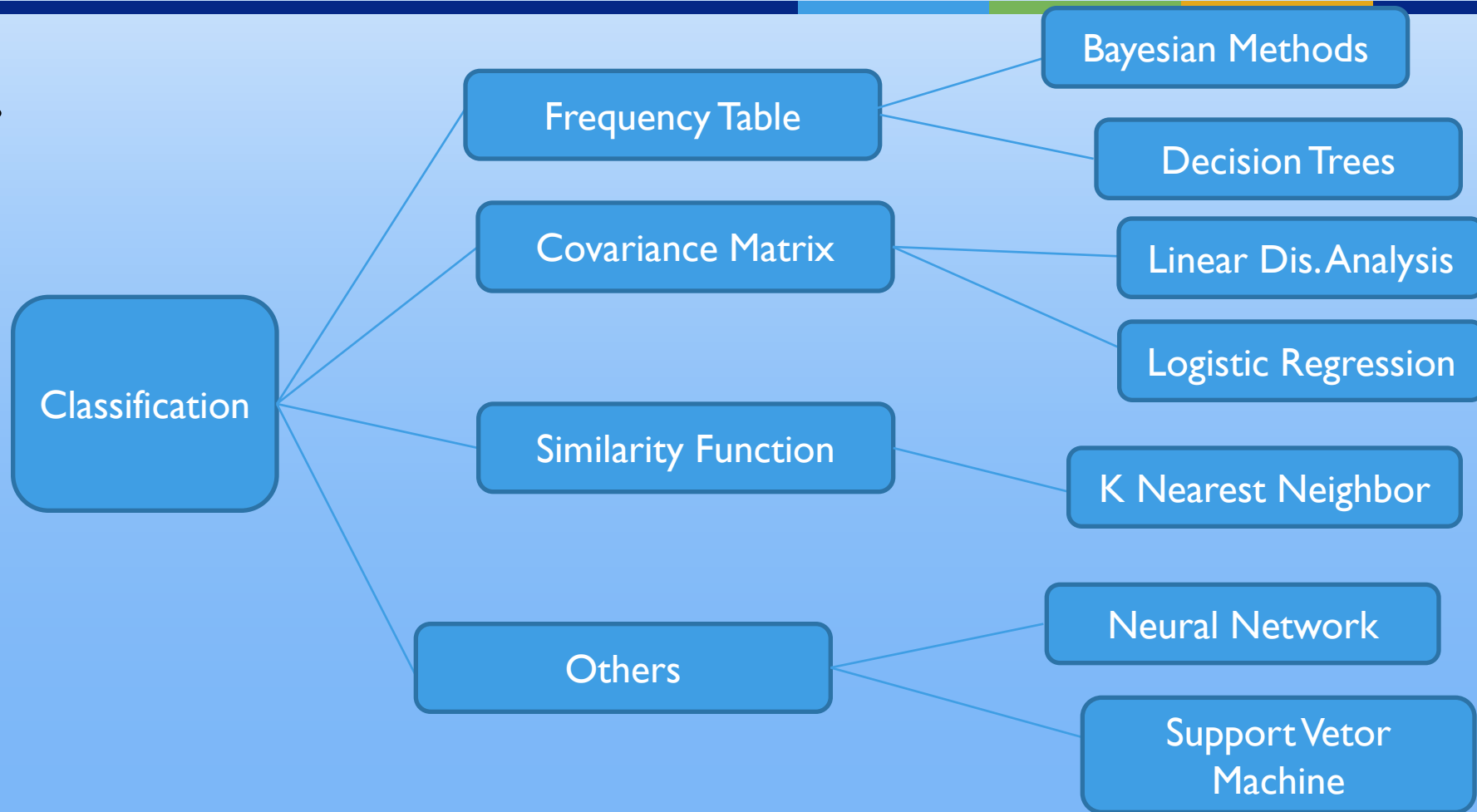
- ▣ **Data**: Loan application data
- ▣ **Task**: Predict whether a loan should be approved or not.
- ▣ **Performance measure**: accuracy.

**No learning**: classify all future applications (test data) to the majority class (i.e., **Yes**):

$$\text{Accuracy} = 9/15 = 60\%.$$

- ▣ We can do better than 60% with learning.

# Supervised Learning Methods



# K-Nearest-Neighbors Algorithm

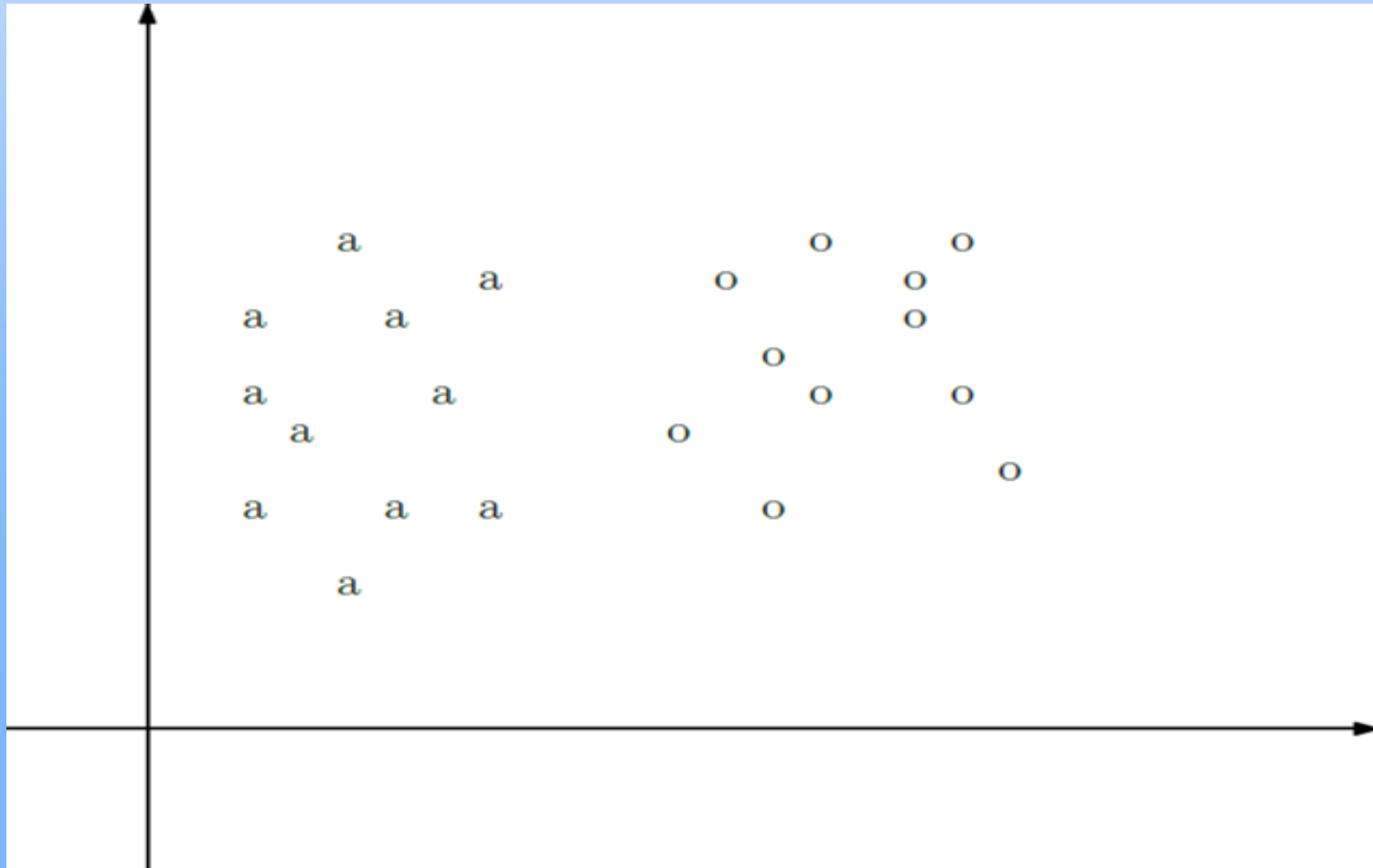
- K nearest neighbors (KNN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (distance function)
- KNN has been used in statistical estimation and pattern recognition since 1970's.

# K-Nearest-Neighbors Algorithm

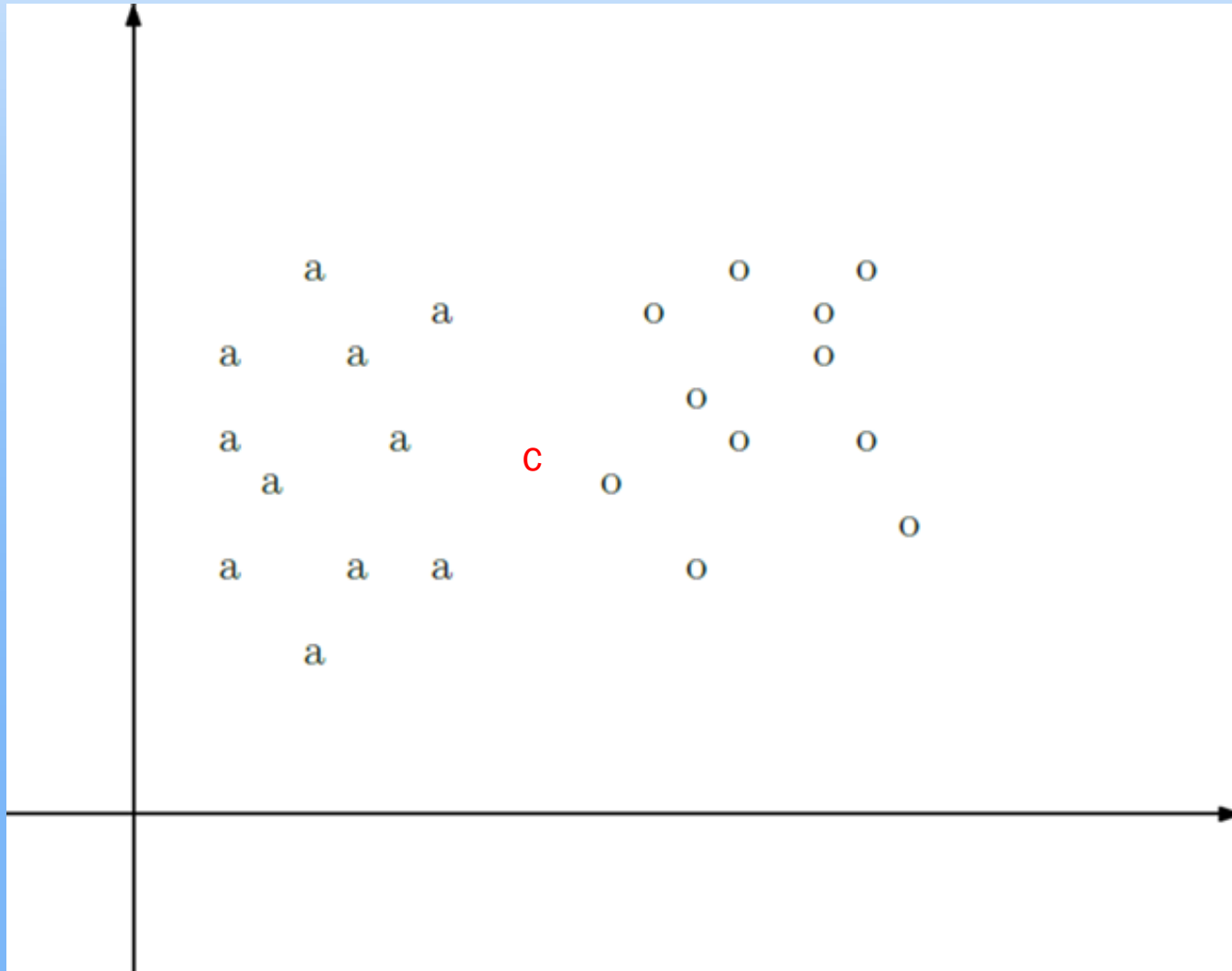
- A case is classified by a majority voting of its neighbors, with the case being assigned to the class most common among its  $K$  nearest neighbors measured by a distance function.
- If  $K=1$ , then the case is simply assigned to the class of its nearest neighbor



# K-Nearest-Neighbors



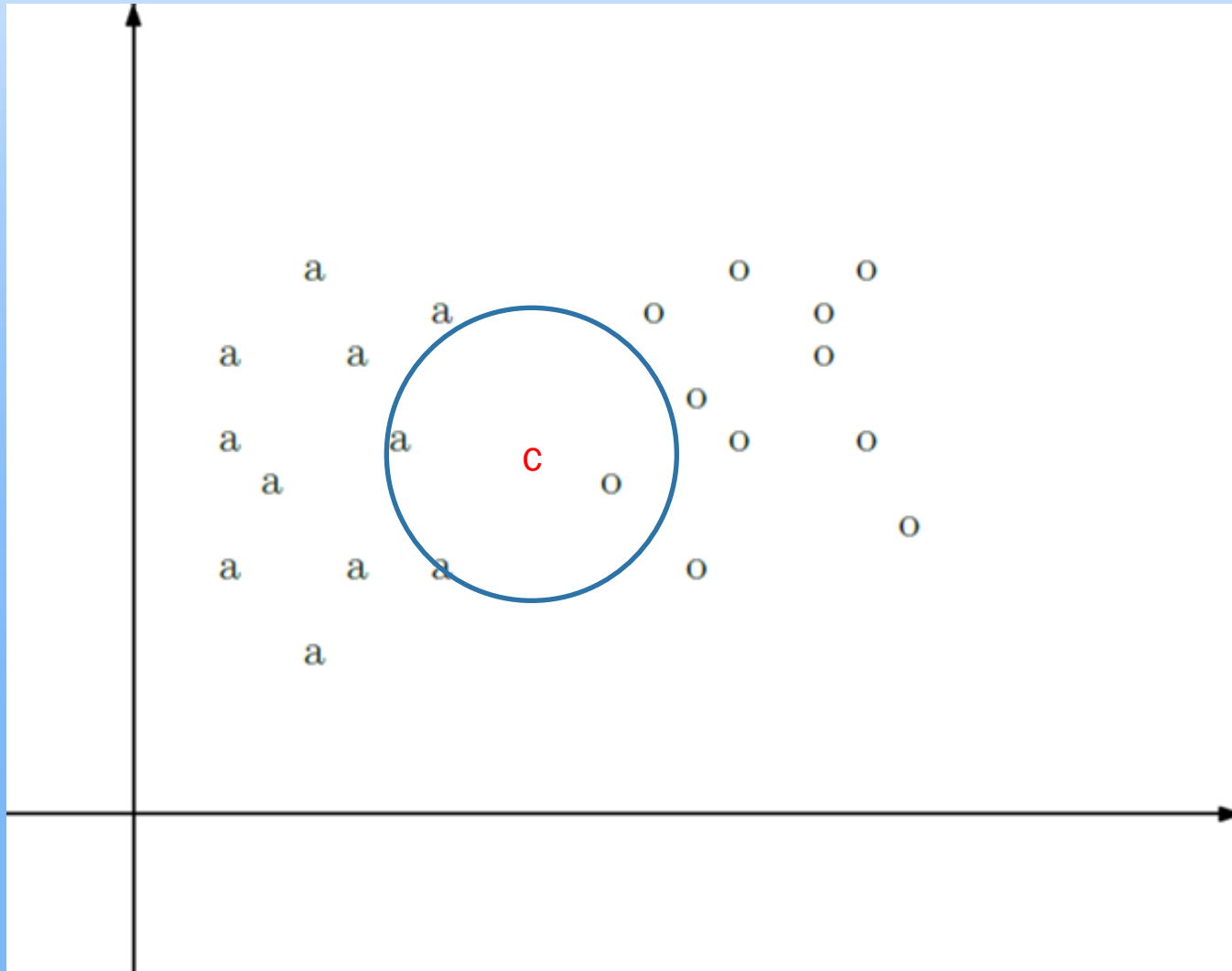
# What is the most possible label for c?



# What is the most possible label for c?

- ▣ Solution: Looking for the nearest  $K$  neighbors of  $c$ .
- ▣ Take the majority label as  $c$ 's label
- ▣ Let's suppose  $k = 3$ :

# What is the most possible label for c?



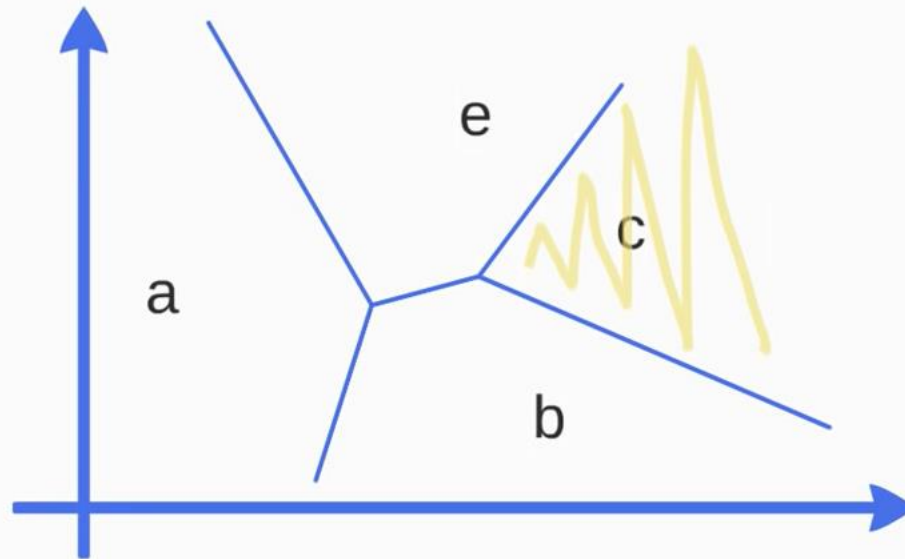
# What is the most possible label for c?

---

- The 3 nearest points to c are: a, a and o.
- Therefore, the most possible label for c is a.

# Voronoi Diagram

When  $k = 1$ , each training vector defines a region in space, defining a *Voronoi* partition of the space



$$R_i = \{x : d(x, x_i) < d(x, x_j), i \neq j\}$$

# Voronoi Diagram



# Code Link

---

<https://github.com/Kowsher/ML/blob/master/Classification/KNN/KNN-Classification>



# Next Class Work

- ▣ Linear Functions

[https://www.youtube.com/watch?v=AqIMrHOBM4g&ab\\_channel=mahadolotcom](https://www.youtube.com/watch?v=AqIMrHOBM4g&ab_channel=mahadolotcom)

- ▣ Linear Equations

[https://www.youtube.com/watch?v=8eXb-6wQUks&ab\\_channel=MyWhyU](https://www.youtube.com/watch?v=8eXb-6wQUks&ab_channel=MyWhyU)

- ▣ Odds Ratio

[https://www.youtube.com/watch?v=5zPSD\\_e\\_N04&ab\\_channel=TheNCCMT](https://www.youtube.com/watch?v=5zPSD_e_N04&ab_channel=TheNCCMT)

- ▣ Conditional Probability

[https://www.youtube.com/watch?v=ibINrxJLvIM&ab\\_channel=Dr.TreforBazett](https://www.youtube.com/watch?v=ibINrxJLvIM&ab_channel=Dr.TreforBazett)

- ▣ Sigmoid Function

[https://www.youtube.com/watch?v=NOwUCIQ7v3c&ab\\_channel=JaayBhatt](https://www.youtube.com/watch?v=NOwUCIQ7v3c&ab_channel=JaayBhatt)