**K-MEANS**

(Là thuật toán để Phân Cụm – chia ra những Nhóm khác nhau)

**Means**: Trung Bình

**K:** là một số nào đó. (Vd: K=2 -> Kmeans sẽ tìm ra 2 màu đặc trưng nhất của bức ảnh,…)

**Bắt buộc:** cần trang bị kiến thức về Đại Số Tuyến Tính, Toán Cao Cấp, Ma Trận, Vector …

Numpy: ma trận, vector

Trong AI không có khái niệm ĐÚNG hay SAI. Đâu đó chỉ là **Xác Suất Cao Nhất**. Và mình sẽ tính được Xác Suất đấy.

Có thể có nhiều nhiều.

Không nhất thiết phải là 1 điểm. Có thể là một vật thể nào đó trong đời sống.

Miễn là có thể quy đổi nó về dạng số, thì nó sẽ quy được về thành VECTOR.

1. Minh họa các điểm dữ liệu bằng PyGame
2. Xử lý ảnh

**VECTOR**: 2D, 3D, 4D … và nD (là cái để thể hiện lên dữ liệu)

* 2D: a(x,y) Là 1 điểm trong không gian a(1,1) b(2,1).
* 3D: b(x,y,z)
* …
* Từng con số trong VECTOR (là x, y, z, …) đó là các FEATURES (những đặc tính, đặc trưng của đối tượng)

A picture containing text, indoor

Description automatically generated

A picture containing chart

Description automatically generated

**Bài toán thực tế:** Phân loại Email (phân cụm Email thuộc nhóm nào)

Như Email về Trả lại tiền, Email Thay đổi mật khẩu, Email Hổ trợ khách hàng, Email về Vay tiền …

Khi một email mới đến, dùng K-Means phân biệt Email đó thuộc nhóm nào ở trên?

Biểu thức BITAGO:

CB = **Căn bậc 2 của Tổng ((A-B)mũ 2 + (A-C)mũ 2)**

A person sitting in front of a computer screen

Description automatically generated with low confidence

Sử dụng thư viện **Numpy Array** trong python:

a=[1,2]

* a = numpy.array(a)

Ma trận: là gộp của các VECTORS

A person sitting at a desk in front of a computer screen

Description automatically generated with low confidence

**Thuật toán K-Means: trong pygame**

A picture containing text, indoor, person, blackboard

Description automatically generated

Để dịch chuyển điểm K => tính Trung bình Cộng của các điểm cùng CỤM của nó. Để xác định lại điểm K. Sao cho nó nằm giữa CỤM là đúng nhất.

Install PYGAME for MAC: **python3 -m pip install -U pygame --user**