

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**  
**KHOA CÔNG NGHỆ THÔNG TIN 1**

o0o



**BÀI TẬP LỚN**

**Đề tài: Phân loại ô tô sử dụng thuật toán**  
**K-means Clustering**

**Môn học: Nhập môn Trí tuệ nhân tạo**

**Nhóm lớp: 06**

**Giảng viên hướng dẫn: TS. Đào Thị Thúy Quỳnh**

**Sinh viên thực hiện:**

**Họ và tên: Nguyễn Minh Quân      MSSV: B21DCCN613**

**HÀ NỘI, 05/2024**

# LỜI CẢM ƠN

Công nghệ ngày càng phổ biến và không ai có thể phủ nhận được tầm quan trọng và những hiệu quả mà nó đem lại cho cuộc sống chúng ta. Bất kỳ trong lĩnh vực nào, sự góp mặt của trí tuệ nhân tạo sẽ giúp con người làm việc và hoàn thành tốt công việc hơn. Và gần đây, một thuật ngữ “machine learning” rất được nhiều người quan tâm. Thay vì phải code phần mềm với cách thức thủ công theo một bộ hướng dẫn cụ thể nhằm hoàn thành một nhiệm vụ đề ra thì máy sẽ tự “học hỏi” bằng cách sử dụng một lượng lớn dữ liệu cùng những thuật toán cho phép nó thực hiện các tác vụ.

Đây là một lĩnh vực khoa học tuy không mới, nhưng cho thấy lĩnh vực trí tuệ nhân tạo đang ngày càng phát triển và có thể tiến xa hơn trong tương lai. Đồng thời, thời điểm này nó được xem là một lĩnh vực “nóng” và dành rất nhiều mối quan tâm để phát triển nó một cách mạnh mẽ, bùng nổ hơn.

Hiện nay, sự quan tâm machine learning càng ngày càng tăng lên là vì nhờ machine learning giúp gia tăng dung lượng lưu trữ các loại dữ liệu sẵn, việc xử lý tính toán có chi phí thấp và hiệu quả hơn rất nhiều.

Những điều trên được hiểu là nó có thể thực hiện tự động, nhanh chóng để tạo ra những mô hình cho phép phân tích các dữ liệu có quy mô lớn hơn và phức tạp hơn đồng thời đưa ra những kết quả một cách nhanh và chính xác hơn.

Chính sự hiệu quả trong công việc và các lợi ích vượt bậc mà nó đem lại cho chúng ta khiến machine learning ngày càng được chú trọng và quan tâm nhiều hơn. Vì vậy chúng em đã chọn đề tài “ Phân loại ô tô sử dụng thuật toán K-means Clustering”.

Chúng em xin chân thành gửi lời cảm ơn tới cô Đào Thị Thúy Quỳnh đã tận tình giảng dạy, truyền đạt cho chúng em những kiến thức cũng như kinh nghiệm quý báu trong suốt quá trình học. Cô đã tận tình theo sát giúp đỡ, trực tiếp chỉ bảo, hướng dẫn trong suốt quá trình nghiên cứu và học tập của chúng em.

# Mục Lục

Danh sách hình vẽ.....	4
<b>CHƯƠNG 1. TỔNG QUAN VỀ HỌC MÁY .....</b>	<b>5</b>
<b>1.1. Khái niệm về học máy:.....</b>	<b>5</b>
<b>1.2 . Phân nhóm các thuật toán học máy .....</b>	<b>5</b>
1.2.1 Học có giám sát (Supervised Learning).....	5
1.2.2 Học phi giám sát (Unsupervised Learning).....	6
1.2.3 Học tăng cường (reinforcement learning).....	8
1.2.4 Học bán giám sát (Semi-Supervised Learning) .....	8
<b>1.3. Ứng dụng của học máy:.....</b>	<b>9</b>
<b>CHƯƠNG 2: THUẬT TOÁN K-MEANS CLUSTERING TRONG BÀI TOÁN PHÂN CỤM.....</b>	<b>11</b>
<b>2.1. Tổng quan về thuật toán K-Means Clustering.....</b>	<b>11</b>
<b>2.2. Thuật toán K-Means Clustering: .....</b>	<b>12</b>
2.2.1. Mô hình toán học: .....	12
2.2.2. Độ chính xác của thuật toán .....	13
2.2.3. Nghiệm của thuật toán K-Means Clustering .....	13
2.2.4 Tóm tắt thuật toán:.....	14
<b>CHƯƠNG 3: ỨNG DỤNG THUẬT TOÁN K-MEANS CLUSTERING .....</b>	<b>15</b>
<b>3.1. Dữ liệu hình ảnh ô tô:.....</b>	<b>15</b>
3.1.1. Phát biểu bài toán: .....	15
3.1.2. Yêu cầu: .....	16
3.1.3. Tiến hành phân cụm.....	16
<b>Tài liệu tham khảo .....</b>	<b>21</b>

## Danh sách hình vẽ

1.1	Mô hình học có giám sát.....	6
1.2	Mô hình học không giám sát.....	7
1.3	Sự khác biệt giữa 2 mô hình SL và UL .....	7
2.1	Bài toán với 3 clusters .....	11
2.2	Mô hình dữ liệu được phân cụm.....	12

# CHƯƠNG 1. TỔNG QUAN VỀ HỌC MÁY

## 1.1. Khái niệm về học máy:

Học máy là khả năng của chương trình máy tính sử dụng kinh nghiệm, quan sát, hoặc dữ liệu trong quá khứ để cải thiện công việc của mình trong tương lai thay vì chỉ thực hiện theo đúng các quy tắc đã được lập trình sẵn. Chẳng hạn, máy tính có thể học cách dự đoán dựa trên các ví dụ, hay học cách tạo ra các hành vi phù hợp dựa trên quan sát trong quá khứ. Học máy vẫn đòi hỏi sự đánh giá của con người trong việc tìm hiểu dữ liệu cơ sở và lựa chọn các kỹ thuật phù hợp để phân tích dữ liệu. Đồng thời, trước khi sử dụng, dữ liệu phải sạch, không có sai lệch và không có dữ liệu giả.

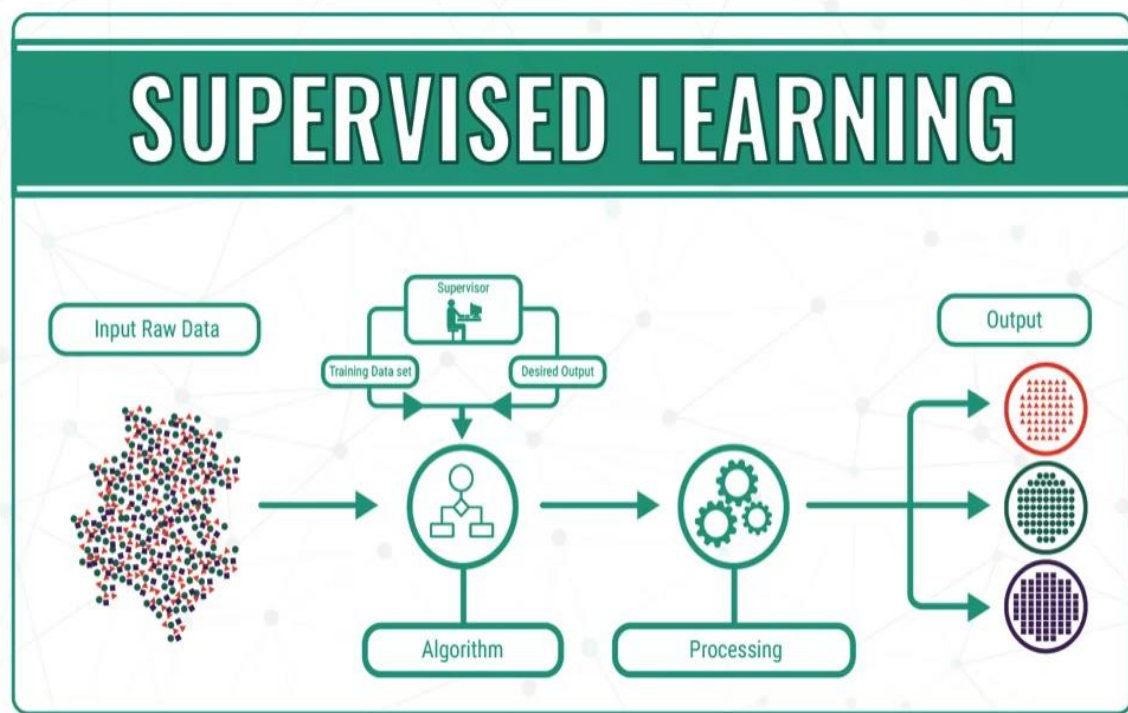
Các mô hình học máy yêu cầu lượng dữ liệu đủ lớn để "huấn luyện" và đánh giá mô hình. Trước đây, các thuật toán học máy thiếu quyền truy cập vào một lượng lớn dữ liệu cần thiết để mô hình hóa các mối quan hệ giữa các dữ liệu. Sự tăng trưởng trong dữ liệu lớn (big data) đã cung cấp các thuật toán học máy với đủ dữ liệu để cải thiện độ chính xác của mô hình và dự đoán.

## 1.2. Phân nhóm các thuật toán học máy

### 1.2.1 Học có giám sát (Supervised Learning)

Là dạng học máy trong đó cho trước tập dữ liệu huấn luyện dưới dạng các ví dụ cùng với giá trị đầu ra hay giá trị đích. Dựa trên dữ liệu huấn luyện, thuật toán học cần xây dựng mô hình hay hàm đích để dự đoán giá trị đầu ra (giá trị đích) cho các trường hợp mới.

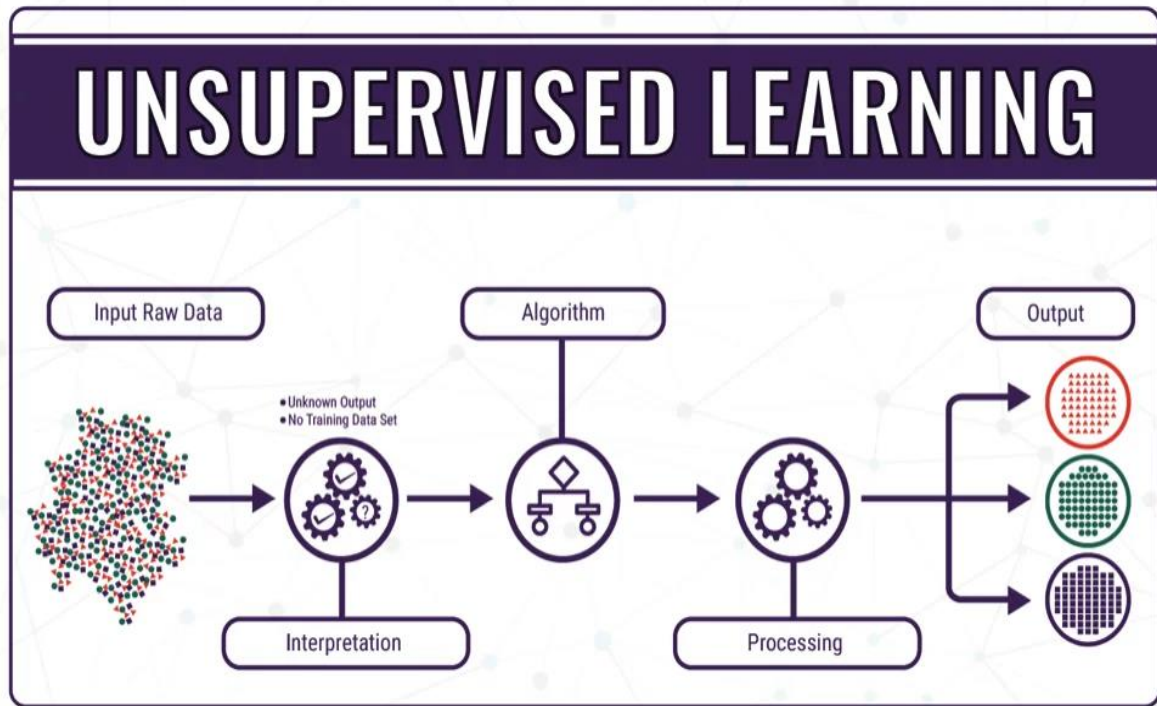
- Nếu giá trị đầu ra là rời rạc thì học có giám sát được gọi là phân loại hay phân lớp (classification).
- Nếu đầu ra nhận giá trị liên tục, tức đầu ra là số thực, thì học có giám sát được gọi là hồi quy (regression).



*Hình 1. 1. Mô hình học có giám sát*

### 1.2.2 Học phi giám sát (Unsupervised Learning)

Khác với học có giám sát, học phi giám sát sử dụng những dữ liệu chưa được gán nhãn từ trước để suy luận. Phương pháp này thường được sử dụng để tìm cấu trúc của tập dữ liệu. Tuy nhiên lại không có phương pháp đánh giá được cấu trúc tìm ra được là đúng hay sai. Ví dụ như phân cụm dữ liệu, triết xuất thành phần chính của một chất nào đó. Ứng dụng phổ biến nhất của học không giám sát là gom cụm (cluster).

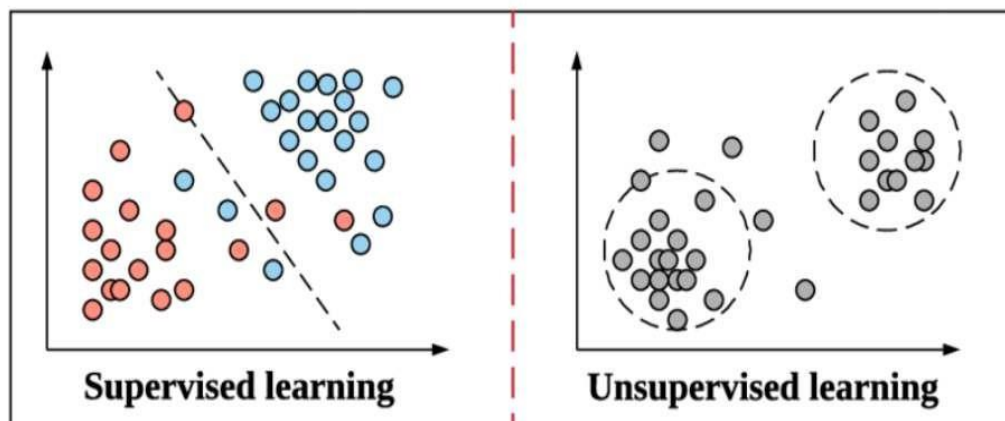


Hình 1. 2. Mô hình học không giám sát

Trong thuật toán này, chúng ta không biết được dữ liệu đầu ra hay nhãn mà chỉ có dữ liệu đầu vào. Thuật toán Học không giám sát dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó, ví dụ như phân nhóm hoặc giảm số chiều của dữ liệu để thuận tiện trong việc lưu trữ và tính toán.

Một cách toán học, Học không giám sát là khi chúng ta chỉ có dữ liệu vào  $X$  mà không biết nhãn  $Y$  tương ứng.

Sự khác nhau giữa học có giám sát và học không giám sát:



### *Hình 1.3: Sự khác biệt giữa 2 mô hình SL và UL*

Học có giám sát: Là cách huấn luyện một mô hình trong đó dữ liệu học có đầu vào và đầu ra tương ứng đầu vào đó. Mô hình được huấn luyện bằng cách giảm thiểu sai số lỗi (loss) của các dự đoán tại các vòng lặp huấn luyện. Sau quá trình huấn luyện, mô hình sẽ có khả năng đưa ra dự đoán về đầu ra với một đầu vào mới gặp (không có trong dữ liệu học).

Học không giám sát: Là cách huấn luyện một mô hình trong đó dữ liệu học chỉ bao gồm đầu vào mà không có đầu ra. Mô hình sẽ được huấn luyện cách để tìm cấu trúc hoặc mối quan hệ giữa các đầu vào. Một trong những phương pháp học không giám sát quan trọng nhất là phân cụm (clustering): Tạo các cụm khác nhau với mỗi cụm biểu diễn một đặc trưng nào đó của dữ liệu và phân các đầu vào mới vào các cụm theo các đặc trưng của đầu vào đó.

#### 1.2.3 Học tăng cường (reinforcement learning)

Đối với dạng học này, kinh nghiệm không được cho trực tiếp dưới dạng đầu vào/đầu ra cho mỗi trạng thái hoặc mỗi hành động. Thay vào đó, hệ thống nhận được một giá trị khuyến khích (reward) là kết quả cho một chuỗi hành động nào đó. Thuật toán cần học cách hành động để cực đại hóa giá trị khuyến khích. Ví dụ của học khuyến khích là học đánh cờ, trong đó hệ thống không được chỉ dẫn nước đi nào là hợp lý cho từng tình huống mà chỉ biết kết quả toàn ván cờ. Như vậy, các chỉ dẫn về nước đi được cho một cách gián tiếp và có độ trễ dưới dạng giá trị thưởng. Nước đi tốt là nước đi nằm trong một chuỗi các nước đi dẫn tới kết quả thắng toàn bộ ván cờ.

#### 1.2.4 Học bán giám sát (Semi-Supervised Learning)

Các bài toán khi chúng ta có một lượng lớn dữ liệu  $X$  nhưng chỉ một phần trong chúng được gán nhãn được gọi là Semi-Supervised Learning. Những bài toán



thuộc nhóm này nằm giữa hai nhóm được nêu bên trên. Một ví dụ điển hình của nhóm này là chỉ có một phần ảnh hoặc văn bản được gán nhãn (ví dụ bức ảnh về người, động vật hoặc các văn bản khoa học, chính trị) và phần lớn các bức ảnh/văn bản khác chưa được gán nhãn được thu thập từ internet.

Thực tế cho thấy rất nhiều các bài toán Machine Learning thuộc vào nhóm này vì việc thu thập dữ liệu có nhãn tốn rất nhiều thời gian và có chi phí cao. Rất nhiều loại dữ liệu thậm chí cần phải có chuyên gia mới gán nhãn được (ảnh y học chẳng hạn). Ngược lại, dữ liệu chưa có nhãn có thể được thu thập với chi phí thấp từ internet.

### **1.3. Ứng dụng của học máy:**

Sau đây là một số ví dụ ứng dụng cụ thể của học máy:

- Nhận dạng ký tự: phân loại hình chụp ký tự thành các loại, mỗi loại ứng với một ký tự tương ứng.

- Phát hiện và nhận dạng mặt người: phát hiện vùng có chứa mặt người trong ảnh, xác định đó là mặt người nào trong số những người đã có ảnh trước đó, tức là phân chia ảnh thành những loại tương ứng với những người khác nhau.

- Lọc thư rác, phân loại văn bản: dựa trên nội dung thư điện tử, chia thư thành loại “thư rác” hay “thư bình thường”; hoặc phân chia tin tức thành các thể loại khác nhau như “xã hội”, “kinh tế”, “thể thao”.v.v.

- Dịch tự động: dựa trên dữ liệu huấn luyện dưới dạng các văn bản song ngữ, hệ thống dịch tự động học cách dịch từ ngôn ngữ này sang ngôn ngữ khác. Hệ thống dịch tự động tiêu biểu dạng này là Google Translate.

- Chẩn đoán y tế: học cách dự đoán người bệnh có mắc hay không mắc một số bệnh nào đó dựa trên triệu chứng quan sát được.

- Phân loại khách hàng và dự đoán sở thích: sắp xếp khách hàng vào một số loại, từ đây dự đoán sở thích tiêu dùng của khách hàng.

- Dự đoán chỉ số thị trường: căn cứ giá trị một số tham số hiện thời và trong lịch sử, đưa ra dự đoán, chẳng hạn dự đoán giá chứng khoán, giá vàng.v.v.

- Các hệ khuyến nghị, hay hệ tư vấn lựa chọn: cung cấp một danh sách ngắn các loại hàng hóa, phim, video, tin tức v.v. mà người dùng nhiều khả năng

quan tâm. Ví dụ ứng dụng loại này là phần khuyến nghị trên Youtube hay trên trang mua bán trực tuyến Amazon.

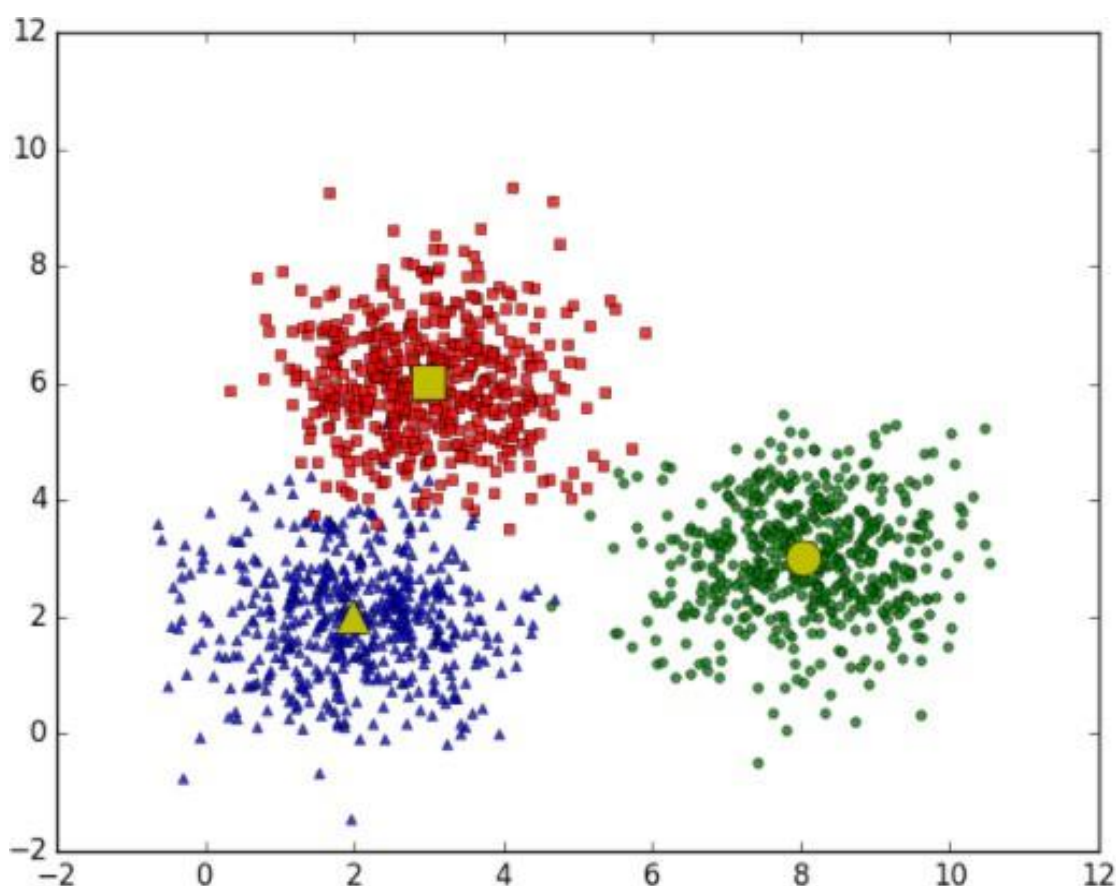
- Ứng dụng lái xe tự động: dựa trên các mẫu học chứa thông tin về các tình huống trên đường, hệ thống học máy cho phép tự ra quyết định điều khiển xe, và do vậy không cần người lái. Hiện Google đã có kế hoạch thương mại hóa xe ô tô tự động lái như vậy.

## CHƯƠNG 2: THUẬT TOÁN K-MEANS CLUSTERING

### TRONG BÀI TOÁN PHÂN CỤM

#### 2.1. Tổng quan về thuật toán K-Means Clustering

Với thuật toán K-Means Clustering, chúng ta không biết nhãn (label) của từng điểm dữ liệu. Mục đích là làm thế nào để phân dữ liệu thành các cụm (cluster) khác nhau sao cho dữ liệu trong cùng một cụm có tính chất giống nhau. Ý tưởng đơn giản nhất về cluster (cụm) là tập hợp các điểm ở gần nhau trong một không gian nào đó (không gian này có thể có rất nhiều chiều trong trường hợp thông tin về một điểm dữ liệu là rất lớn). Hình bên dưới là một ví dụ về 3 cụm dữ liệu (viết gọn là cluster).



Hình 2.1: Bài toán với 3 clusters

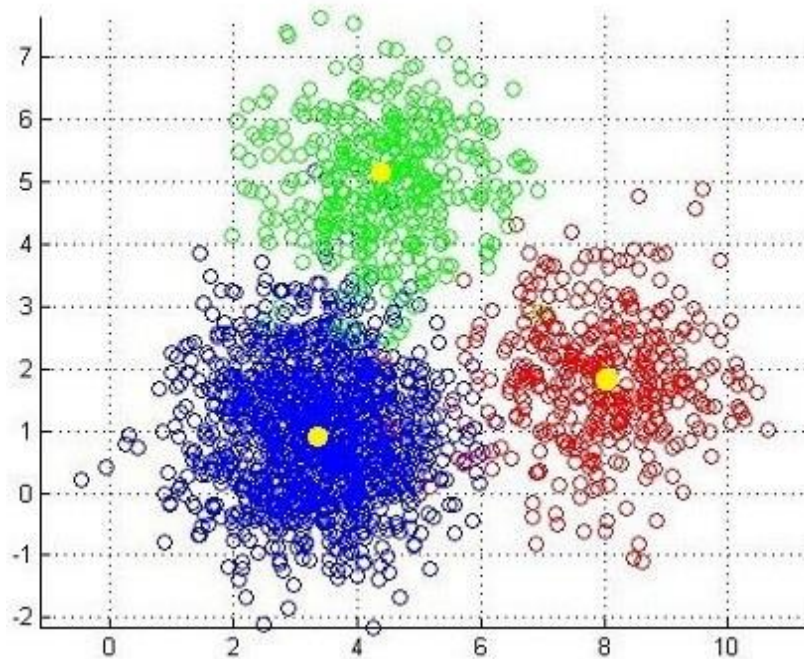
Giả sử mỗi cluster có một điểm đại diện (*center*) màu vàng. Và những

điểm xung quanh mỗi center thuộc vào cùng nhóm với center đó. Một cách đơn giản nhất, xét một điểm bất kỳ, ta xét xem điểm đó gần với center nào nhất thì nó thuộc về cùng nhóm với center đó.

## 2.2. Thuật toán K-Means Clustering:

### 2.2.1. Mô hình toán học:

Ta gọi điểm tại vị trí trung bình của tất cả các điểm dữ liệu trong một cụm là **trung tâm cụm**. Như vậy, nếu có K cụm thì sẽ có K trung tâm cụm và mỗi trung tâm cụm sẽ nằm gần các điểm dữ liệu trong cụm tương ứng hơn các trung tâm cụm khác. Trong hình dưới đây,  $K = 3$  và ta có 3 trung tâm cụm là các điểm màu vàng.



Hình 2.2: Mô hình dữ liệu được phân cụm

Để phân cụm dữ liệu bằng K-Means Clustering, trước hết ta chọn K là số cụm để phân chia và chọn ngẫu nhiên K trong số m dữ liệu ban đầu làm trung tâm cụm  $\mu_1, \mu_2, \dots, \mu_K$ . Sau đó, với điểm dữ liệu  $x(i)$  ta sẽ gán nó cho cụm  $c(i)$

là cụm có trung tâm cụm gần nó nhất.

$$c^{(i)} = \operatorname{argmin}_k \|x^{(i)} - \mu_k\|^2$$

Khi tất cả các điểm dữ liệu đã được gán về các cụm, bước tiếp theo là tính toán lại vị trí các trung tâm cụm bằng trung bình tọa độ các điểm dữ liệu trong cụm đó.

$$\mu_k = \frac{1}{n} (x^{(k_1)} + x^{(k_2)} + \dots + x^{(k_n)})$$

Với  $k_1, k_2, \dots, k_n$  là chỉ số các dữ liệu thuộc cụm thứ  $k$ . Các bước trên được lặp lại cho tới khi vị trí các trung tâm cụm không đổi sau một bước lặp nào đó.

### 2.2.2. Độ chính xác của thuật toán:

Hàm mất mát của thuật toán K-Means Clustering đặc trưng cho độ chính xác của nó sẽ càng lớn khi khoảng cách từ mỗi điểm dữ liệu tới trung tâm cụm càng lớn.

$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2$$

### 2.2.3. Nghiệm của thuật toán K-Means Clustering:

Trong các bước của thuật toán, thực chất bước gán các điểm dữ liệu về trung tâm cụm gần nhất và bước thay đổi trung tâm cụm về vị trí trung bình của các điểm dữ liệu trong cụm đều nhằm mục đích giảm hàm mất mát. Thuật toán kết thúc khi vị trí các trung tâm cụm không đổi sau một bước lặp nào đó. Khi đó hàm mất mát đạt giá trị nhỏ nhất. Khi  $K$  càng nhỏ so với  $m$ , thuật toán càng dễ đi

đến kết quả chưa phải tối ưu. Điều này phụ thuộc vào cách chọn K trung tâm cụm ban đầu.

Để khắc phục điều này, ta cần lặp lại thuật toán nhiều lần và chọn phương án có giá trị hàm mất mát nhỏ nhất.

#### 2.2.4 Tóm tắt thuật toán:

**Đầu vào:** Dữ liệu X và số lượng cluster cần tìm K.

**Đầu ra:** Các center M và label vector cho từng điểm dữ liệu Y.

1. Chọn K điểm bất kỳ làm các center ban đầu.
2. Phân mỗi điểm dữ liệu vào cluster có center gần nó nhất.
3. Nếu việc gán dữ liệu vào từng cluster ở bước 2 không thay đổi so với vòng lặp trước nó thì ta dừng thuật toán.
4. Cập nhật center cho từng cluster bằng cách lấy trung bình cộng của tất cả các điểm dữ liệu đã được gán vào cluster đó sau bước 2.
5. Quay lại bước 2.

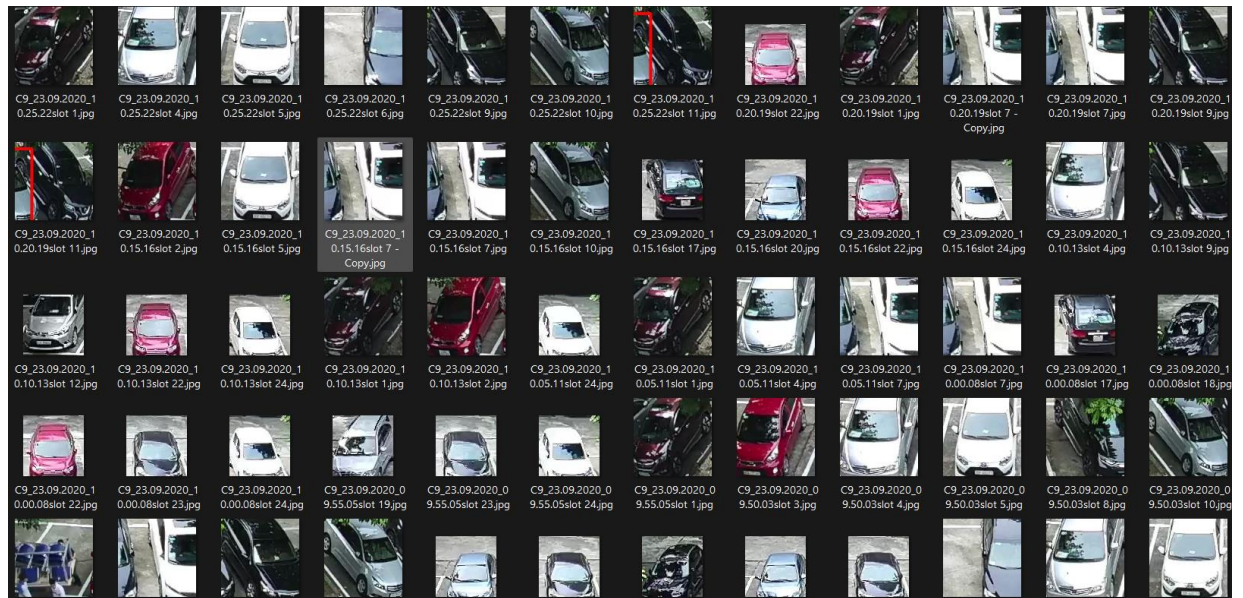
## CHƯƠNG 3: ỨNG DỤNG THUẬT TOÁN K-MEANS CLUSTERING

### 3.1. Dữ liệu hình ảnh ô tô:

#### 3.1.1. Phát biểu bài toán:

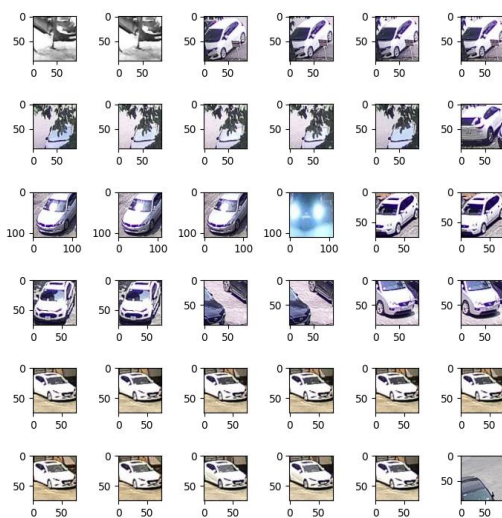
Bài toán phân loại ô tô đưa ra các tập dữ liệu là các hình ảnh của từng loại xe để phân loại màu sắc, sắc độ của chúng thành các cụm để dễ quản lý, tìm kiếm.

- Giá trị Input: Hình ảnh của các mẫu ô tô trộn lẫn với nhau

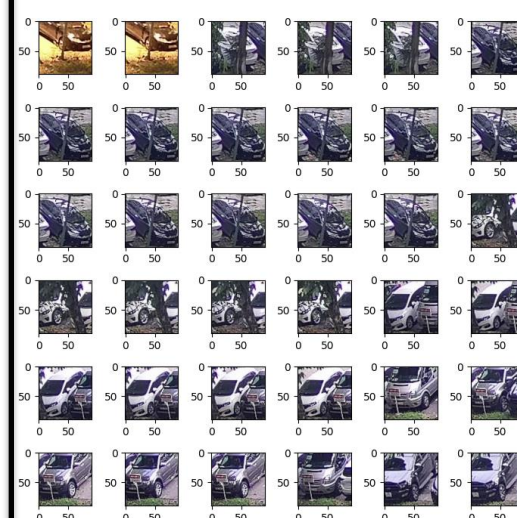


Hình 3.1. Ví dụ về mẫu input

- Giá trị Output: Các cụm chúng được phân vào

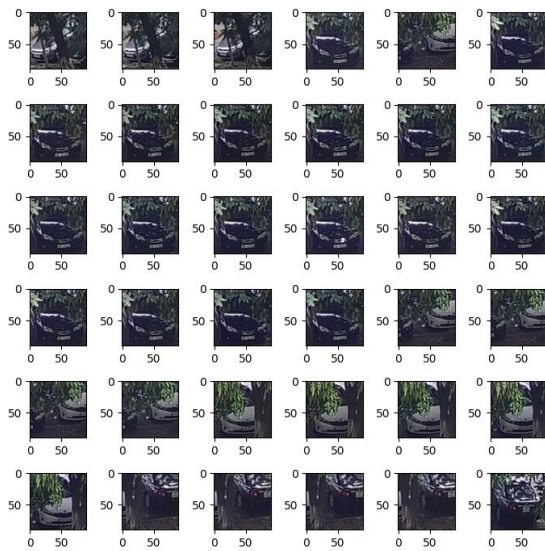


Hình 3.2. Ví dụ Cụm 1

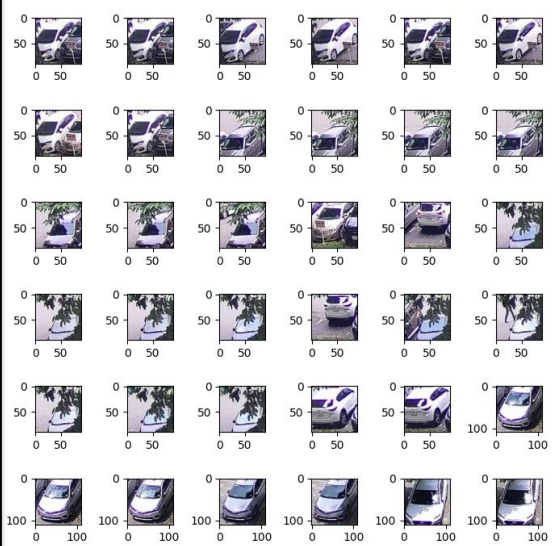


Hình 3.3. Ví dụ Cụm 2

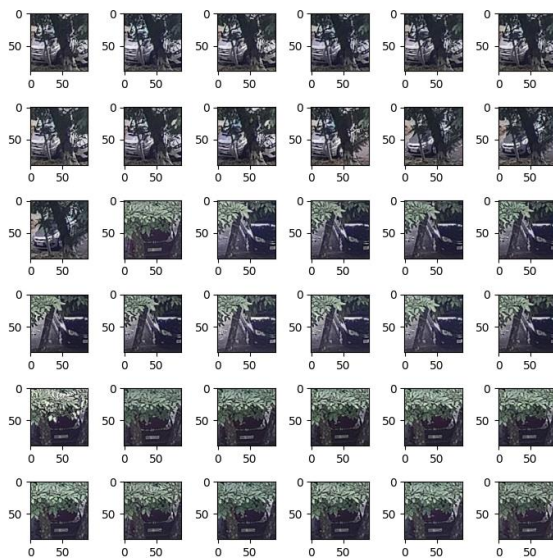




Hình 3.4. Ví dụ Cụm 3



Hình 3.5. Ví dụ Cụm 4



Hình 3.6. Ví dụ Cụm 5

### 3.1.2. Yêu cầu:

- Lấy dữ liệu hình ảnh ô tô
- Xử lý, làm sạch dữ liệu
- Tiến hành phân cụm
- Hiển thị các cụm được chia ra

### 3.1.3. Tiến hành phân cụm

- Import Thư viện:

```

2 import cv2
3 from sklearn.cluster import KMeans
4 import os
5 import numpy as np
6 import pickle
7 import matplotlib.pyplot as plt
8 import matplotlib.image as mpimg

```



- Đường dẫn data:

```
11 c_data_path = "../data/Cars"
```

- Xử lý ảnh: Do các ảnh có kích thước khác nhau nên chúng ta cần phải đưa về cùng một dạng (R, G, B) với các thuộc tính có giá trị từ 0 đến 1 để có thể xử lý.

```
14 def get_feature(img):
15     intensity = img.sum(axis=1)
16     intensity = intensity.sum(axis=0) / (255 * img.shape[0] * img.shape[1])
17     return intensity
```

- Đọc dữ liệu và gán vào 2 list X, Y: Với biến X lưu dữ liệu về hình ảnh, Y lưu tên file ( để sau khi phân cụm, ta có thể biết được file nào thuộc cụm nào ).

```
20 def load_data(data_path=c_data_path):
21     X = []          # List lưu dữ liệu hình ảnh
22     L = []          # List lưu tên file
23     # Đọc file sau đó gán X vào L
24     for file in os.listdir(data_path):
25         c_x = get_feature(cv2.imread(os.path.join(data_path, file)))
26         X.append(c_x)
27         L.append(file)
28     # Tạo 2 mảng
29     X = np.array(X)
30     L = np.array(L)
31     # Lưu dữ liệu vào 2 file binary
32     with open('data.pickle', 'wb') as handle:
33         pickle.dump(X, handle, protocol=pickle.HIGHEST_PROTOCOL)
34     with open('label.pickle', 'wb') as handle:
35         pickle.dump(L, handle, protocol=pickle.HIGHEST_PROTOCOL)
36
37     return X, L
38
```

- Để tìm số cụm cho thuật toán K-means, ta sẽ sử dụng phương pháp Elbow. Phương pháp Elbow là một cách giúp ta lựa chọn được số lượng các cụm phù hợp dựa vào đồ thị trực quan hoá bằng cách nhìn vào sự suy giảm của *hàm biến dạng (wcss)* và lựa chọn ra điểm *khủy tay (elbow point)* là điểm mà ở đó tốc

độ suy giảm của *hàm biến dạng* sẽ thay đổi nhiều nhất. Tức là kể từ sau vị trí này thì gia tăng thêm số lượng cụm cũng không giúp *hàm biến dạng* giảm đáng kể..

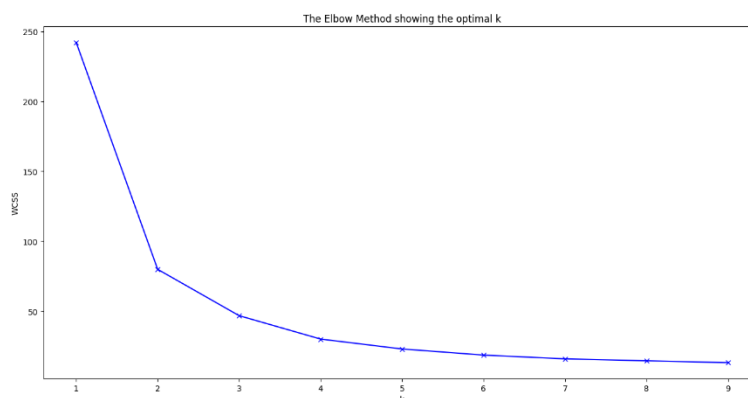
- Trước tiên ta sẽ khởi tạo 1 list wcss để lưu tổng bình phương khoảng cách của các mẫu đến tâm cụm gần nhất của chúng. Sau đó, dùng vòng lặp để chạy thuật toán K-means từ 1 cluster đến 10 clusters. Sau đó lưu giá trị wcss vào list wcss.

```
wcss = [] #wcss: đo lường sự sai lệch đến điểm centerpoints
# Khởi tạo thuật toán kmean cluster với số cluster từ 1->10
K = range(1,10)
for k in K:
    kmeanModel = KMeans(n_clusters=k)
    kmeanModel.fit(X)
    wcss.append(kmeanModel.inertia_)
```

- Vẽ biểu đồ Elbow

```
51 plt.figure(figsize=(16,8))
52 plt.plot(K, wcss, 'bx-')
53 plt.xlabel('k')
54 plt.ylabel('WCSS')
55 plt.title('The Elbow Method showing the optimal k')
56 plt.show()
```

- Từ biểu đồ Elbow ta chọn được  $k = 5$ .



- Thực hiện thuật toán K-means với 5 cluster rồi training cho biến kmeans:

```
59 kmeans = KMeans(n_clusters=5).fit(X)
```

- In ra tên file và các cluster tương ứng:

```
61 for i in range(len(kmeans.labels_)):  
62     print(kmeans.labels_[i], " - ", L[i])
```

- In ra tọa độ các tâm:

```
65 print(kmeans.cluster_centers_)
```

```
[[0.55992974 0.52965331 0.53855601]  
 [0.33954306 0.3266321 0.32542533]  
 [0.4492003 0.42724419 0.43415966]  
 [0.68249894 0.65549548 0.65291713]  
 [0.2212639 0.21285234 0.20536671]]
```

# Kết Luận

Trong quá trình thực hiện báo cáo của môn nhập môn học máy, chúng em đã được biết thêm về các chương trình ứng dụng, các thuật toán và nắm bắt được phần nào kiến thức về học máy. Từ đó chúng em cố gắng áp dụng các kiến thức đã học vào làm báo cáo để hoàn thiện đề tài của mình. Trong báo cáo của chúng em đã sử dụng thuật toán phân cụm K-Means Clustering để dự đoán phân loại các mẫu ô tô. Dù đã rất cố gắng để bài báo cáo hoàn thiện tốt nhất, song vì thời gian có hạn và tầm hiểu biết còn rất hạn hẹp nên bài báo cáo của chúng em không tránh khỏi còn nhiều thiếu sót. Chúng em rất hy vọng có sự góp ý của thầy cô, để báo cáo của chúng em được hoàn thiện hơn nữa ! Chúng em xin chân thành cảm ơn !!

# Tài liệu tham khảo

- (1) [Machine Learning cơ bản \(machinelearningcoban.com\)](http://machinelearningcoban.com)
- (2) Từ Minh Phương. Giáo trình Nhập môn Trí tuệ nhân tạo. Học viện Công nghệ Bưu chính Viễn Thông.
- (3) [Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning \(crayondata.ai\)](http://crayondata.ai)