1/ Trình bày thuật toán nhận dạng ảnh bằng Naïve Bayes

Để có thể sử dụng thuật toán Naïve Bayes ta cần làm các bước sau

**Chuẩn bị dữ liệu:**

Dữ liệu đào tạo cần được chuẩn bị với các ảnh được gán nhãn đúng. Nhãn này có thể là tên của đối tượng trong ảnh (ví dụ: "mèo", "chó").

**Chia dữ liệu đào tạo và kiểm tra:**

Tương tự như trong trường hợp phân loại văn bản, dữ liệu được chia thành hai phần: một phần để đào tạo mô hình và một phần để kiểm tra mô hình.

**Trích xuất đặc trưng:**

Đối với ảnh, đặc trưng có thể là giá trị pixel của từng điểm ảnh hoặc một số đặc trưng thống kê như histogram màu sắc, kích thước, v.v.

**Tính toán xác suất đối với từng đặc trưng:**

Mục tiêu là tính xác suất cho mỗi lớp (nhãn) dựa trên giá trị của các đặc trưng. Ví dụ, nếu sử dụng giá trị pixel, có thể tính toán xác suất của mỗi pixel dựa trên lớp và áp dụng giả định Naïve Bayes về sự độc lập.

**Giả sử độc lập:**

Giả định rằng các đặc trưng (giá trị pixel) độc lập với nhau dưới điều kiện đã biết lớp.

**Áp dụng định lý Bayes:**

Sử dụng định lý Bayes để tính xác suất đối với mỗi lớp dựa trên xác suất của các đặc trưng.

**Dự đoán:**

Chọn lớp có xác suất cao nhất làm dự đoán cho ảnh.

**Kiểm tra và đánh giá mô hình:**

Sử dụng dữ liệu kiểm tra để đánh giá hiệu suất của mô hình.

2/ Nếu có một đặc trưng thành phần không xuất hiện trong dữ liệu huấn luyện dẫn đến xác suất bằng 0. Đề xuất phương án giải quyết vấn đề này

Nếu có một đặc trưng thành phần không xuất hiện trong dữ liệu huấn luyện dẫn đến xác suất bằng 0 thì ta có thể sử dụng phương pháp Laplacian.

Trong Naïve Bayes, một cách tiếp cận phổ biến là sử dụng ước lượng Laplace smoothing (còn được gọi là "Laplacian correction" hoặc "add-one smoothing") để giảm tác động của xác suất 0. Cụ thể, công thức cập nhật Laplacian có thể được áp dụng vào ước lượng xác suất của đặc trưng dựa trên dữ liệu đào tạo.

Khác với công thức của Naïve Bayes là:

Thì khi áp dụng Laplacian smoothing ta được như sau:

* k: giá trị làm mượt (lớn hơn 0)
* X: số đặc trưng còn lại của dữ liệu (nếu đang xét A trong B mà B còn 3 dữ liệu đặc trưng thì X = 3)