# Homework exercise: thuật toán Viterbi cho HHM

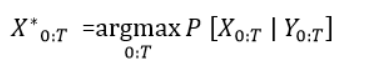
1) Trình bày thuật toán Viterbi cho Hidden Markov Model

2) Code giải thuật Viterbi trên giải bài toán POS tagging

## Thuật toán Viterbi là gì?

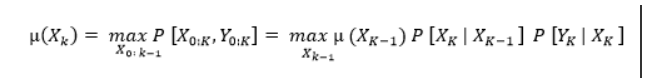
Thuật **toán Viterbi** là một **thuật toán lập trình động** để xác định ước tính xác suất tốt nhất của **chuỗi trạng thái ẩn có khả năng xảy ra nhất** dẫn đến thứ tự của các sự kiện được quan sát, đặc biệt trong bối cảnh nguồn thông tin Markov và mô hình Markov ẩn (HMM).

Thuật **toán Viterbi** nhằm mục đích **tạo ra một giả định dựa trên mô hình được đào tạo và một số dữ liệu được quan sát** . Đầu tiên nó hỏi về lựa chọn có khả năng xảy ra nhất dựa trên dữ liệu và mô hình được đào tạo. Chúng ta có thể tính toán và phân tích nó theo công thức sau:



Công thức này sẽ tìm ra các trạng thái có thể **tối đa hóa xác suất có điều kiện** của các trạng thái của dữ liệu.

Chúng ta có thể sử dụng công thức đã cho để tìm **tập trạng thái tốt nhất bằng cách sử dụng đệ quy** :



* Đây là hàm mu và nó hoạt động với công thức đầu tiên.

Kỹ thuật này được sử dụng rộng rãi trong **việc giải mã các mã xoắn** được sử dụng trong CDMA và GSM di động kỹ thuật số, modem quay số, vệ tinh và thông tin liên lạc trong không gian sâu. Hiện nay nó được sử dụng rộng rãi trong nhận dạng giọng nói, phát hiện từ khóa, ngôn ngữ học tính toán và ứng dụng tin sinh học. Ví dụ, trong chuyển giọng nói thành văn bản (nhận dạng giọng nói), sóng âm thanh được coi là chuỗi sự kiện được quan sát, trong khi một chuỗi văn bản được coi là "nguyên nhân tiềm ẩn" của tín hiệu âm thanh. Với một kích thích âm thanh, thuật toán Viterbi sẽ tìm ra chuỗi văn bản có khả năng xảy ra nhất.

## Mô hình Markov ẩn (HMM)

Mô **hình Markov ẩn** là mô hình được sử dụng trong thống kê để xác định mối **quan hệ xác suất** giữa **một chuỗi trạng thái ẩn và quan sát** . Nó có thể được sử dụng với các quan sát ẩn của bất kỳ quy trình nào, do đó nó có tên là Mô hình Markov ẩn.

Nó dự đoán các quan sát trong tương lai dựa trên quy trình ẩn chịu trách nhiệm tạo dữ liệu. Nó có hai loại biến:

* Trạng thái ẩn: Trạng thái ẩn là những biến được sử dụng để tạo ra dữ liệu ẩn nhưng được quan sát
* Quan sát: Các biến cần đo lường và quan sát là các quan sát

## Thuật toán Viterbi hoạt động như thế nào?

Mô **hình Markov ẩn trước** tiên xác định trạng thái và không gian quan sát. Sau đó, phân bố xác suất của trạng thái ban đầu được xác định. Mô hình sau đó được đào tạo. Sau khi đào tạo, **Thuật toán Viterbi** được sử dụng để **giải mã các chuỗi trạng thái ẩn có khả năng xảy ra nhất** . Kỹ thuật này cung cấp **xác suất ghi nhật ký của chuỗi trạng thái ẩn có khả năng xảy ra nhất và chính chuỗi trạng thái ẩn** đó. Sau đó, thuật toán này kiểm tra tính chính xác và hiệu suất của mô hình.

