BÁO CÁO TIẾN ĐỘ TUẦN 5

Người báo cáo: Trần Đức Thọ.

Bài 1: Em đã push code lên gitlab.

Bài 2: RNN

RNN là gì?

RNN được gọi là hồi quy (Recurrent) bởi lẽ chúng thực hiện cùng một tác vụ cho tất cả các phần tử của một chuỗi với đầu ra phụ thuộc vào cả các phép tính trước đó. Nói cách khác, RNN có khả năng nhớ các thông tin được tính toán trước đó. Trên lý thuyết, RNN có thể sử dụng được thông tin của một văn bản rất dài, tuy nhiên thực tế thì nó chỉ có thể nhớ được một vài bước trước đó (ta cùng bàn cụ thể vấn đề này sau) mà thôi. Về cơ bản một mạng RNN có dạng như sau:

Diagram

Description automatically generated with medium confidence

RNN được dùng khi nào?

Các mô hình RNN hầu như được sử dụng trong lĩnh vực xử lí ngôn ngữ tự nhiên và ghi nhận tiếng nói

RNN được triển khai như thế nào

Đầu tiên ta duỗi 1 mạng RNN là có dạng như sau

A picture containing text, sky

Description automatically generated

Tại mỗi thời điểm t tiếp diễn chúng ta có 1 input đầu vào:

Với mỗi mỗ hình RNN đầu vào chúng ta có các hàm chuyển tiếp

Diagram

Description automatically generated

Công thức lan truyền ngược được biểu diễn như sau

Text, letter

Description automatically generated

Bài 4: PCA

+ PCA là gì?

PCA là một phương pháp giảm các chiều dữ liệu sang số chiều dữ liệu ít hơn.

+ Khi nào sử dụng PCA?

Khi chúng ta có một tập dữ liệu với thứ nguyên quá lớn.

+ Tại sao sử dụng PCA?

Khi chúng ta có một tập dữ liệu với thứ nguyên quá lớn, chúng ta nên giảm số chiều của nó theo PCA sẽ được tập dữ liệu phù hợp với mô hình của chúng ta.

+ Dùng PCA ở đâu?

Chúng ta nên dùng PCA ở các bài toán có số chiều lớn.

+ Phương pháp triển khai PCA:

1.Tính vector kỳ vọng của toàn bộ dữ liệu . (N là tổng số các điểm dữ liệu )

Schematic

Description automatically generated with medium confidence

2. Trừ mỗi điểm dữ liệu đi vector kỳ vọng của toàn bộ dữ liệu

Text

Description automatically generated with medium confidence

3. Tính ma trận hiệp phương sai

A picture containing chart

Description automatically generated

4. Tính các trị riêng và vector riêng có norm bằng 1 của ma trận này, sắp xếp chúng theo thứ tự giảm dần của trị riêng.

5. Chọn K vector riêng ứng với K trị riêng lớn nhất để xây dựng ma trận UK có các cột tạo thành một hệ trực giao. K vectors này, còn được gọi là các thành phần chính, tạo thành một không gian con *gần* với phân bố của dữ liệu ban đầu đã chuẩn hoá.

6. Chiếu dữ liệu ban đầu đã chuẩn hoá X^ xuống không gian con tìm được.

7. Dữ liệu mới chính là toạ độ của các điểm dữ liệu trên không gian mới:

Text

Description automatically generated

Dữ liệu ban đầu có thể tính được xấp xỉ theo dữ liệu mới như sau:

Logo

Description automatically generated

Lưu ý : Khi nào nên dùng PCA: khi các điểm dữ liệu rời rạc với số chiều lớn.

Không nên dùng PCA khi:

* Số chiều dữ liệu bé (1,2)
* Mật độ dữ liệu không có điểm tổng quát.