



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC MÁY TÍNH

LỚP CS115.P11.KHTN
MÔN TOÁN CHO KHOA HỌC MÁY TÍNH
Cholesky decomposition and generating correlated random variables
CỦ NHÂN NGÀNH KHOA HỌC MÁY TÍNH
05-11-2024

Nguyễn Trọng Tất Thành
Mã số sinh viên: 23521455

Giảng viên hướng dẫn: TS. Lương Ngọc Hoàng

TP. HỒ CHÍ MINH, NĂM 2024



Mục lục

1	Gới thiệu	3
2	Ứng dụng: Phát sinh dữ liệu ngẫu nhiên có tương quan	4
3	Demo: Mã Python cho Google Colab	5
4	Kết luận	6



1.

1 Giới thiệu

1.1 Giới thiệu về Phân tích Cholesky:

Phân tích Cholesky (Cholesky Decomposition) là một kỹ thuật trong đại số tuyến tính để phân tích một ma trận dương xác định thành tích của một ma trận tam giác dưới và ma trận chuyển vị của nó. Điều này có nghĩa là, với một ma trận dương xác định A , chúng ta có thể biểu diễn nó dưới dạng:

$$A = LL^T$$

trong đó A là ma trận dương xác định, và L là ma trận tam giác dưới. Phân tích Cholesky có nhiều ứng dụng trong thống kê và học máy, như giải hệ phương trình tuyến tính, tính toán xác suất, và tạo dữ liệu ngẫu nhiên có mối tương quan.

1.2 Mục Tiêu của Bài Tập:

- Hiểu rõ lý thuyết phân tích Cholesky và cách áp dụng nó trong việc tạo dữ liệu ngẫu nhiên có tương quan.
- Sử dụng phân tích Cholesky để sinh các biến ngẫu nhiên có phân phối chuẩn và có mối tương quan nhất định.
- Minh họa bằng mã Python để tạo ra dữ liệu ngẫu nhiên có tương quan trong Google Colab.

1.3 Cấu Trúc Bài Làm:

Bài tập sẽ bao gồm các phần chính sau:

- **Lý thuyết:** Trình bày cơ bản về phân tích Cholesky, định nghĩa và công thức.
- **Ứng dụng:** Sử dụng phân tích Cholesky để tạo dữ liệu ngẫu nhiên có tương quan.
- **Demo:** Mã Python minh họa trong Google Colab để tạo dữ liệu ngẫu nhiên có tương quan giữa hai biến.

2.

2 Ứng dụng: Phát sinh dữ liệu ngẫu nhiên có tương quan

2.1 Mô tả ứng dụng của Phân tích Cholesky:

Một trong những ứng dụng phổ biến của phân tích Cholesky là tạo ra dữ liệu ngẫu nhiên có mối tương quan. Giả sử chúng ta muốn tạo ra hai biến ngẫu nhiên X và Y có phân phối chuẩn và có một mức độ tương quan xác định.

Để thực hiện điều này, chúng ta sẽ sử dụng ma trận hiệp phương sai (covariance matrix), sau đó áp dụng phân tích Cholesky để chuyển đổi dữ liệu ngẫu nhiên ban đầu thành dữ liệu có tương quan.

2.2 Các bước thực hiện:

- *Bước 1: Xác định ma trận hiệp phương sai (Covariance Matrix)*

Giả sử ma trận hiệp phương sai của hai biến ngẫu nhiên X và Y được xác định là:

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

trong đó:

- σ_X và σ_Y là độ lệch chuẩn của X và Y ,
- ρ là hệ số tương quan giữa X và Y .
- *Bước 2: Thực hiện phân tích Cholesky trên ma trận hiệp phương sai*
Sử dụng phân tích Cholesky để tìm ma trận L sao cho $\Sigma = LL^T$.
- *Bước 3: Tạo dữ liệu ngẫu nhiên*
Sinh một vector ngẫu nhiên Z có phân phối chuẩn đa biến (mean = 0, variance = 1). Sau đó, tính $X = L \cdot Z$ để thu được dữ liệu ngẫu nhiên có mối tương quan mong muốn.

2.3 Ví dụ: Tạo dữ liệu ngẫu nhiên có tương quan giữa hai biến

Để minh họa, chúng ta sẽ tạo một ví dụ sử dụng các bước trên để phát sinh dữ liệu ngẫu nhiên có tương quan giữa hai biến, sau đó hiển thị bằng đồ thị. Mã Python cho ví dụ này sẽ được trình bày ở phần tiếp theo trong Google Colab.

3.

3 Demo: Mã Python cho Google Colab

3.1 Mục tiêu của Demo:

Mục tiêu của phần này là minh họa cách sử dụng phân tích Cholesky để tạo dữ liệu ngẫu nhiên có tương quan giữa hai biến ngẫu nhiên X và Y bằng Python. Demo này sẽ được thực hiện trên Google Colab và bạn có thể truy cập link để xem mã nguồn chi tiết và kết quả.

3.2 Quy trình phát sinh dữ liệu ngẫu nhiên có tương quan:

Các bước để tạo dữ liệu ngẫu nhiên có tương quan giữa hai biến như sau:

- *Bước 1: Thiết lập random seed*
Đặt random seed để đảm bảo tính tái lập của kết quả. Trong mã của chúng ta, seed được đặt là 23521455.
- *Bước 2: Xác định các tham số*
Đặt các tham số như trung bình và độ lệch chuẩn cho hai biến ngẫu nhiên X và Y , cùng với hệ số tương quan giữa chúng.
- *Bước 3: Tạo ma trận hiệp phương sai*
Xây dựng ma trận hiệp phương sai Σ cho hai biến ngẫu nhiên dựa trên các tham số đã xác định.
- *Bước 4: Phân tích Cholesky*
Áp dụng phân tích Cholesky để tính ma trận L , sao cho $\Sigma = LL^T$.
- *Bước 5: Sinh dữ liệu ngẫu nhiên có tương quan*
Tạo một vector ngẫu nhiên Z với phân phối chuẩn, sau đó nhân với L để có được dữ liệu ngẫu nhiên có tương quan.
- *Bước 6: Hiển thị dữ liệu*
Vẽ biểu đồ phân tán để minh họa sự tương quan giữa hai biến ngẫu nhiên.

3.3 Link tới Google Colab:

Bạn có thể truy cập file Google Colab chứa mã minh họa **tại đây**. File này chứa toàn bộ mã Python và phần minh họa cách tạo dữ liệu ngẫu nhiên có tương quan bằng phân tích Cholesky.

4 Kết luận

4.1 Tóm tắt:

Trong bài tập này, chúng ta đã tìm hiểu về phân tích Cholesky và cách áp dụng nó trong việc phát sinh dữ liệu ngẫu nhiên có tương quan. Phân tích Cholesky cho phép phân rã một ma trận dương xác định thành tích của một ma trận tam giác dưới và ma trận chuyển vị của nó. Đây là một công cụ quan trọng trong đại số tuyến tính với nhiều ứng dụng thực tiễn trong thống kê và học máy.

4.2 Các ứng dụng chính:

Thông qua việc phát sinh dữ liệu ngẫu nhiên có tương quan, chúng ta có thể ứng dụng phân tích Cholesky trong các lĩnh vực sau:

- *Mô phỏng dữ liệu tương quan*: Được sử dụng rộng rãi trong thống kê và học máy để tạo ra các mẫu dữ liệu có tương quan nhất định.
- *Giải hệ phương trình tuyến tính*: Phân tích Cholesky giúp tối ưu hóa quá trình giải hệ phương trình trong các bài toán liên quan đến ma trận dương xác định.
- *Tối ưu hóa và phương pháp Monte Carlo*: Phân tích Cholesky đóng vai trò quan trọng trong các thuật toán tối ưu hóa và các phương pháp mô phỏng Monte Carlo.

4.3 Kết luận chung:

Phân tích Cholesky là một kỹ thuật mạnh mẽ và hữu ích trong nhiều bài toán toán học và thống kê. Việc áp dụng phân tích Cholesky để tạo dữ liệu ngẫu nhiên có tương quan không chỉ là một ví dụ điển hình mà còn là một công cụ quan trọng trong các mô hình thống kê và học máy thực tiễn. Qua bài tập này, chúng ta đã có cái nhìn sâu hơn về cách triển khai và sử dụng phân tích Cholesky để xử lý dữ liệu ngẫu nhiên, đồng thời thấy được tiềm năng ứng dụng rộng rãi của nó trong các lĩnh vực khác nhau.