***TRONG PHƯƠNG PHÁP PHÂN TÍCH THÀNH PHẦN CHÍNH-PCA***

1. ***Phân tích SDV (Singular Value Decomposition)***

* ***Định nghĩa***

*Cho là ma trận thực cỡ (ta thường viết là nó có thể phân tích dưới dạng*

*trong đó là ma trận gồm các cột trực giao là các vec tơ riêng của*

*và là ma trận gồm các cột trực giao là các vec tơ riêng của ,*

*còn là ma trận đường chéo không vuông với các phần tử trên đường chéo là các giá trị thực không âm với là các trị riêng của và r là hạng của ma trận A, (rank A = r).*

*Các căn bậc hai của các giá trị riêng của được gọi là singular value của A.*

*Mặt khác mỗi cột của U được gọi là vec tơ singular trái, và các cột của V, được gọi là vec tơ singular phải của A.*

*Chú ý :*

* *do U và V là các ma trận trực giao.*
* *là một ma trận nửa xác định dương nên các giá trị riêng của nó phải là số không âm.*
* *Ma trận A có thể viết lại dưới dạng tổng*

* ***Ví dụ 1***

*Cho ma trận , trước hết ta xác định các giá trị riêng cùa A và A với*

*A*

*Bằng cách giải phương trình đặc trưng tương ứng, tức là các phương trình*

*và*

*Ta sẽ có các nghiệm riêng của là . (các căn bậc hai của chúng tương ứng sẽ là*

*Từ đó ta thu được các vec tơ riêng của là ;*

*Suy ra các vec tơ riêng phải của A là*

*Ta thu được*

*.*

*Thực hiện tương tự như vậy cho ma trận ta sẽ thu được*

*Như vậy trong trường hợp này ta có ma trận*

*(Chú ý rằng ta sẽ có ma trận là ma trận đường chéo; và ).*

* ***Ví dụ 2***

*Cho ma trận*

*Từ đó suy ra*

*Giải phương trình đặc trưng: ta được các nghiệm riêng là*

*Giải hệ phương trình tương ứng với các nghiệm riêng*

*Ta sẽ thu được các vec tơ riêng tương ứng là:*

*Chuẩn hóa các vec tơ ta sẽ có*

*Mặt khác ta thấy rằng*

*Tìm giá trị riêng và vec tơ riêng tương ứng với ta sẽ thu được*

*Các vec tơ riêng tương ứng với sẽ là*

*Sau khi chuẩn hóa và thực hiện phép chuyển vị ta thu được*

*⇒*

*Trong trường hợp này có dạng*

*Vậy ta sẽ suy ra*

1. ***Phân tích SDV chặt cụt (Truncated Singular Value Decomposition)***

*Ta có nhận xét rằng trong ma trận D, các giá trị trên đường chéo*

*là không âm và giảm dần. Khi đó ta có thể xấp xỉ ma trận A bằng tổng k < r*

*ma trận có hạng bằng 1:*

*(1)*

*Để đánh giá về sai số trong xấp xỉ trên ta xử dụng chuẩn Frobineous xác định bởi*

*và định lý quan trọng sau.*

* ***Định lý***

*Với chuẩn Frobineous của hiệu hai ma trận ta sẽ có*

*(Ta thừa nhận không chứng minh định lý này).*

*Từ định lý trên ta thấy rằng sai số của xấp xỉ càng nhỏ nếu phần giá trị singular bị chặt cụt càng nhỏ so với phần giữ lại,như vậy đánh giá về về phần giữ lại ta sẽ chọn k là số nhỏ nhất sao cho*

*là số lớn nhất (tất nhiên số đó phải nhỏ hơn 1).*

*Khi k nhỏ, ma trận có hạng là k, là một ma trận có hạng nhỏ, vì vậy SVD chặt cụt còn được gọi là* ***phương pháp xấp xỉ hạng nhỏ*** *(Low-rank approximation).*

* ***Ví dụ 3***

*Cho 2 mẫu ngẫu nhiên (i=1,2) trong bảng giá trị sau*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| *1* | *1* | *2* | *0* | *5* | *4* | *5* | *3* |
| *3* | *2* | *3* | *3* | *4* | *5* | *5* | *4* |

*Để xác định ma trận hiệp phương sai trước hết ta xét các giá trị :*

*Ta sẽ có:*

*Ta xét mẫu được chuẩn hóa (tức là được trừ đi kỳ vọng)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| *-1,63* | *-1,63* | *-0,63* | *-2,63* | *2,38* | *1,38* | *2,38* | *0,38* |
| *-0,63* | *-1,63* | *-0,63* | *-0.63* | *0,38* | *1,38* | *1.38* | *0,38* |

*Ma trận hiệp phương sai tương ứng với hai mẫu trên là*

*Các trị riêng và vec tơ riêng sẽ xác định bởi*

*Từ đó ta thấy rằng trị riêng thứ hai là lớn nhất và lớn hơn nhiều lần so với trị riêng thứ nhất, cụ thể ta có*

*Vậy vec tơ riêng thứ hai (là cột thứ 2 trong V) có phương sai cực đại trong biểu diễn PCA mà ta cần xác định thành phần chính.*

* ***Phương pháp Power***

*Trong nhiều bài toán hiện nay (trong đó có SVD) ta phải tìm trị riêng cho một ma trận rất lớn, chẳng hạn có hơn một triệu bức ảnh pixel, việc trực tiếp tính trị riêng và vec tơ riêng là không khả thi. Hiện có một phương pháp để tính nhanh hơn, đó là phương pháp Power.*

*Phương pháp Power để tìm trị riêng và vec tơ riêng của một ma trận nửa xác định dương*

*Thuật giải với như sau*

1. *Chọn một vec tơ bất kỳ*
2. *Tính*
3. *Chuẩn hóa vec tơ*
4. *Nếu đủ nhỏ thì ta dừng. Nếu không ta lấy rồi quay lại bước 2.*
5. *chính là vec tơ riêng ứng với trị riêng lớn nhất*

* ***Ứng dụng của SVD***

*Có thể nói đến nhiều ứng dụng của SVD, nhưng ta cần nói một trong những ứng dụng quan trọng của nó là*

* *Nén ảnh (Image Compresson)*
* *Hệ thống khuyến nghị (Recommendation Systems)*

*Các ứng dụng trên ta sẽ xem xét kỹ trong môn học* ***machine learning****.*