**2.1-2.3: Long**

**2.4-2.6: Thành**

**2.1** Find the centroid and medoid (most centrally located pattern) for the following set of patterns:

Centroid được tính bằng cách tính trung bình tất cả các chiều của các điểm dữ liệu. Đặt tọa độ của centroid cần tìm là , trong đó:

Vậy centroid của tập dữ liệu là

Để tìm metroid của tập dữ liệu, ta cần tìm điểm có tổng khoảng cách tính từ nó đến các điểm dữ liệu khác là nhỏ nhất. Ta có thể dùng công thức Euclid hoặc Manhattan để tính khoảng cách giữa 2 điểm. Để đơn giản, ta sử dụng công thức Manhattan để tính khoảng cách:

Ví dụ, khoảng cách Manhattan giữa 2 điểm và là .

Ta lập bảng sau để tính theo công thức khoảng cách Manhattan.

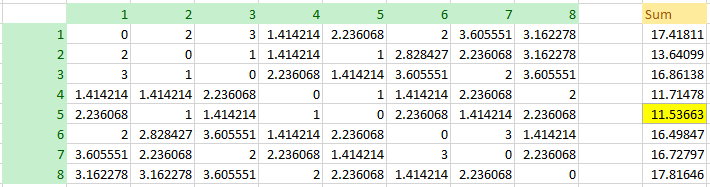
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total Cost |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 0 | 2 | 3 | 2 | 3 | 2 | 5 | 4 | 21 |
| 2 | 2 | 0 | 1 | 2 | 1 | 4 | 3 | 4 | 17 |
| 3 | 3 | 1 | 0 | 3 | 2 | 5 | 2 | 5 | 21 |
| 4 | 2 | 2 | 3 | 0 | 1 | 2 | 3 | 2 | 15 |
| 5 | 3 | 1 | 2 | 1 | 0 | 3 | 2 | 3 | 15 |
| 6 | 2 | 4 | 5 | 2 | 3 | 0 | 3 | 2 | 21 |
| 7 | 5 | 3 | 2 | 3 | 2 | 3 | 0 | 3 | 21 |
| 8 | 4 | 4 | 5 | 2 | 3 | 2 | 3 | 0 | 23 |

Như vậy ta có thể chọn điểm 4 hoặc 5 2, 3) làm metroid do hai điểm này có tổng khoảng cách từ nó tới các điểm khác ( là nhỏ nhất.

\* Nếu áp dụng công thức Euclid để tính khoảng cách: , ta có bảng sau:

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total Cost |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | - | 2.00 | 3.00 | 1.41 | 2.24 | 2.00 | 3.61 | 3.16 | 17.42 |
| 2 | 2.00 | - | 1.00 | 1.41 | 1.00 | 2.83 | 2.24 | 3.16 | 13.64 |
| 3 | 3.00 | 1.00 | - | 2.24 | 1.41 | 3.61 | 2.00 | 3.61 | 16.86 |
| 4 | 1.41 | 1.41 | 2.24 | - | 1.00 | 1.41 | 2.24 | 2.00 | 11.71 |
| 5 | 2.24 | 1.00 | 1.41 | 1.00 | - | 2.24 | 1.41 | 2.24 | 11.54 |
| 6 | 2.00 | 2.83 | 3.61 | 1.41 | 2.24 | - | 3.00 | 1.41 | 16.50 |
| 7 | 3.61 | 2.24 | 2.00 | 2.24 | 1.41 | 3.00 | - | 2.24 | 16.73 |
| 8 | 3.16 | 3.16 | 3.61 | 2.00 | 2.24 | 1.41 | 2.24 | - | 17.82 |

Ở đây ta chọn metroid là điểm 5 do tổng khoảng cách từ điểm này tới các điểm khác trong tập dữ liệu là nhỏ nhất.

**Lâm:** Kết quả giống anh Long.

**2.2** Find the edit distance between the two strings and .

Khoảng cách edit distance (Levenshtein distance) là số bước thao tác nhỏ nhất để biến đổi chuỗi ký tự cho trước sang chuỗi ký tự mục tiêu. Các thao tác biến đổi được cho phép bao gồm: insertion, replacement hoặc deletion. Để biến đổi chuỗi về ta đi qua các thao tác sau:

Như vậy khoảng cách

**Lâm:** Kết quả giống anh Long. Mọi người có thể tìm giải thuật tính Edit Distance trên mạng.

**2.3**. Given the following distance measures. Let state which one is metric and which is non-metric.

Euclidean distance, edit-distance, Manhattan distance, Dynamic Time Warping distance.

Một metric distance cần thỏa mãn các tính chất sau:

* Positive reflexivity:
* Symetric:
* Triangular inequality:

Nếu không thỏa mãn dù chỉ 1 trong 3 tính chất trên thì độ đo được gọi là non-metric distance. Như vậy:

* Metric distance: Euclidean distance, Manhattan distance
* Non-metric distance: Edit distance, Dynamic Time Wraping distance

**Lâm:** Kết quả giống anh Long.

**2.4** A dataset consists of the following patterns:

(1, 1, 1), (2, 2, 1), (1.5, 0.5, 1), (1, 3, 1), (4, 4, 2), (5, 5, 2), (4, 5, 2), (4, 6, 2)

Where each pattern consists of the x coordinate, the y coordinate and the class label. Find the direction of the w vector associated with Fisher’s Linear Discriminant.

*Một bộ dữ liệu bao gồm các mẫu sau:*

*(1, 1, 1), (2, 2, 1), (1,5, 0,5, 1), (1, 3, 1), (4, 4, 2), (5, 5, 2), (4 , 5, 2), (4, 6, 2)*

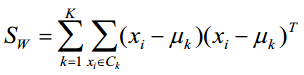
*Trong đó mỗi mẫu bao gồm tọa độ x, tọa độ y và nhãn lớp. Tìm hướng của vectơ w được liên kết với Phân biệt tuyến tính của Fisher.*

Tính mean lớp 1 và 2:

**μ1** = ((1 + 2 + 1.5 + 1) / 4, (1 + 2 + 0.5 + 3) /4) = (1.375, 1.625)

**μ2** = ((4 + 5 + 4 + 4) /4, (4 +5+5+6) /4) = (4.25 ,5)

Công thức tính **SW** (Ghi chú: lấy tọa độ của 1 điểm nằm trong 1 class trừ đi tọa độ của mean của class đó và sau đó nhân cho ma trận chuyển vị (transpose) tương ứng, và lấy tổng tất cả các điểm của các class)



Lấy ví dụ cho điểm x1 = (1,1) có mean là μ1 = (1.375, 1.625)

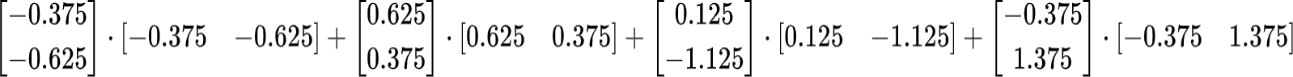
x1 - μ1 = (1, 1) - (1.375, 1.625) = (-0.375, -0.625)

= (-0.375, -0.625) \* (-0.375, -0.625)T = ((0.140625, 0.234375), (0.234375, 0.390625))

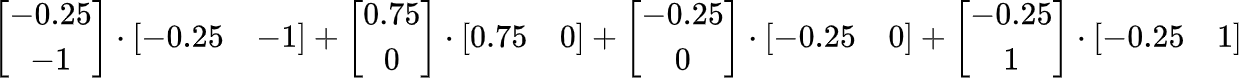
Tính toán tương tự cho các điểm còn lại.

**Sw = S1 + S2**

**=**



+



= ((0.83, 0.42), (0.42, 2.83)) + ((0.75, 0), (-0.25, 2))

= ((1.58, 0.42), (0.17, 5.83))

=> **SW-1** = ((4.02, -0.7), (-0.7, 0.21)) (Inverse matrix of **SW**)

**W = SW-1 (**μ**1 -** μ**2)**

= ((4.02, -0.7), (-0.7, 0.21)) \* (-2.875, -3.375) = (-9.195, 1.303)

**2.5** If there are 10 features and it is necessary to reduce the number of features to 6 so that the best set of six features is chosen, what is the number of feature sub-sets to be evaluated to find the optimal set of six features in the exhaustive search?

C(m, d) = d!/[m!(d – m)!]

C(6, 10) = 10!/(6!\*4!)=210

**Lâm:** Kết quả giống. Số cách chọn ra 6 đặc trưng từ 10 đặc trưng ban đầu mà không quan tâm đến thứ tự = C(6, 10).

**2.6** Give some differences between two feature abstraction methods LDA (Fisher Linear Discriminant) and PCA.

**Answer 1:**

LDA (Fisher Linear Discriminant) and PCA are both feature abstraction methods used in machine learning and data analysis. However, they differ in their goals, assumptions, and methods. Here are some key differences between LDA and PCA:

**Goal:**

* LDA: LDA is a supervised dimensionality reduction method, which means that it seeks to find a projection of the data that maximizes the separation between classes, while minimizing the variance within each class. The goal of LDA is to find the best features to separate the classes.
* PCA: PCA is an unsupervised dimensionality reduction method, which means that it seeks to find a projection of the data that captures the maximum amount of variance in the data. The goal of PCA is to find the most important features that capture the variability of the data.

**Assumptions:**

* LDA: LDA assumes that the data is normally distributed and that the covariance matrix is the same for all classes. LDA assumes that the classes have equal covariance matrices, and the feature values are normally distributed.
* PCA: PCA assumes that the data is linearly related, and there are no assumptions regarding the distribution of the data.

**Method:**

* LDA: LDA calculates the between-class and within-class scatter matrices to find the projection that maximizes the separation between classes while minimizing the variance within each class. It tries to find a linear combination of features that best separates the classes in the data.
* PCA: PCA calculates the covariance matrix and finds the eigenvectors of the covariance matrix. It then projects the data onto the eigenvectors with the highest eigenvalues. It tries to find the most important directions of variance in the data.

**Applicability:**

* LDA: LDA is generally used for classification problems where the goal is to separate data into different classes. It is often used in face recognition or biometric authentication systems.
* PCA: PCA is generally used for data visualization, data compression, or noise reduction. It is often used in image compression, data compression, and data visualization.

**In summary**:

* LDA and PCA are both feature abstraction methods used to reduce the dimensionality of data. However, LDA is a supervised method that tries to find a projection that maximizes the separation between classes, while PCA is an unsupervised method that tries to find the most important directions of variance in the data.

**Answer 2:**

LDA and PCA are both linear transformation techniques for dimensionality reduction. However, they have some key differences12:

* LDA is supervised while PCA is unsupervised. LDA uses class labels while PCA ignores them.
* LDA maximizes the separability between classes while PCA maximizes the variance in the data set.
* LDA forms a new set of components that are linear discriminants while PCA forms a new set of components that are principal components.

1: What is the difference between LDA and PCA for dimensionality reduction? (sebastianraschka.com) <https://sebastianraschka.com/faq/docs/lda-vs-pca.html>

2: LDA vs. PCA – Towards AI <https://towardsai.net/p/data-science/lda-vs-pca>